

KARTA KURSU (realizowanego w specjalności)

Data Science

(nazwa specjalności)

Nazwa	Podstawy Data Science
Nazwa w j. ang.	<i>Introduction to Data Science</i>

Koordinator	dr Roman Czapla	Zespół dydaktyczny
		mgr Katarzyna Marczak mgr Patryk Mazurek
Punktacja ECTS*	st. stacjonarne: 3 st. niestacjonarne: 3	

Opis kursu (cele kształcenia)

Celem kursu jest wprowadzenie studentów w podstawowe zagadnienia analizy danych z wykorzystaniem języka Python i narzędzi open-source. Studenci poznają etapy pracy w Data Science, od pozyskiwania danych, przez ich przygotowanie i analizę, aż po podstawy eksploracyjnej analizy danych i wizualizacji. Kurs obejmuje również elementy statystyki, algebry liniowej, operacji na plikach oraz pracy z bazami danych. Szczególny nacisk położony jest na praktyczne umiejętności pracy z bibliotekami NumPy, Pandas i narzędziami SQL.

Zajęcia przygotowują do dalszego rozwijania kompetencji w zakresie data science i uczenia maszynowego, kładąc solidne podstawy w zakresie przetwarzania, czyszczenia i analizy danych.

Warunki wstępne

Wiedza	Student zna podstawowe pojęcia z zakresu programowania, w szczególności w języku Python, oraz rozumie działanie podstawowych struktur danych i instrukcji sterujących.
Umiejętności	Student potrafi pisać proste programy w Pythonie, korzystać z funkcji wbudowanych i bibliotek zewnętrznych, a także samodzielnie uruchamiać skrypty i pracować w środowisku interaktywnym (np. Jupyter Notebook).
Kursy	Wymagane zaliczenie kursu: Podstawy Programowania w języku Python

Efekty uczenia się

	Efekt uczenia się dla kursu	Odniesienie do efektów kierunkowych
Wiedza	Po zakończeniu kursu: W01: Student zna podstawowe narzędzia środowiska analizy danych w Pythonie, takie jak Anaconda, Jupyter Notebook oraz biblioteki NumPy, Pandas i SQLite. W02: Student zna struktury danych wykorzystywane w analizie danych (np. tablice NumPy, DataFrame Pandas) i ich podstawowe operacje. W03: Student zna sposoby pozyskiwania danych z różnych źródeł (pliki tekstowe, JSON, API, bazy danych) oraz metody ich przetwarzania. W04: Student zna podstawowe pojęcia statystyki opisowej, algebry liniowej oraz ich zastosowania w analizie danych. W05: Student zna różnice między SQL a Pandas w kontekście pracy z danymi tabelarycznymi. W06: Student zna podstawowe etapy analizy danych oraz rozumie cykl pracy w Data Science (CRISP-DM).	
Umiejętności	Po zakończeniu kursu: U01: Student potrafi zaimportować dane, przekształcać je i analizować z wykorzystaniem bibliotek NumPy i Pandas. U02: Student potrafi przeprowadzić czyszczenie danych, uzupełnianie braków, kodowanie zmiennych i normalizację. U03: Student potrafi obliczyć i zinterpretować podstawowe miary statystyczne oraz wykonać proste analizy grupowe. U04: Student potrafi wykorzystać podstawowe zapytania SQL i połączyć dane z bazą SQLite w środowisku Python. U05: Student potrafi analizować i przekształcać szeregi danych, w tym dane czasowe i wielowymiarowe. U06: Student potrafi zintegrować dane pochodzące z różnych źródeł (pliki, API, bazy danych) i przygotować je do analizy w jednym środowisku.	
Kompetencje społeczne	Po zakończeniu kursu: K01: Student potrafi samodzielnie rozwiązywać podstawowe problemy związane z przygotowaniem i analizą danych. K02: Student potrafi współpracować w zespole przy realizacji zadań analitycznych. K03: Student rozumie znaczenie rzetelności, przejrzystości i poprawności analizy danych.	

Studia stacjonarne

Organizacja							
Forma zajęć	Wykład (W)	Ćwiczenia w grupach					
		A	K	L	S	P	E
Liczba godzin	25			30			1

Studia niestacjonarne

Organizacja							
Forma zajęć	Wykład (W)	Ćwiczenia w grupach					
		A	K	L	S	P	E
Liczba godzin	15			20			1

Opis metod prowadzenia zajęć

Zajęcia mają formę wykładów oraz ćwiczeń laboratoryjnych. W trakcie wykładów omawiane są podstawowe koncepcje analizy danych, statystyki, algebry liniowej oraz pracy z danymi w środowisku Python. Ćwiczenia koncentrują się na praktycznym wykorzystaniu poznanych narzędzi i technik: studenci pracują z bibliotekami NumPy i Pandas, przetwarzają dane z różnych źródeł, przygotowują je do analizy i wykonują operacje grupujące, statystyczne i transformacyjne.

Część zajęć poświęcona jest również pracy z bazami danych i zapytaniem SQL. Studenci realizują projekt zaliczeniowy, w którym samodzielnie planują i wykonują prostą analizę danych zgodnie z poznany cyklem analitycznym

Formy sprawdzania efektów uczenia się

	E – learning	Gry dydaktyczne	Ćwiczenia w szkole	Zajęcia terenowe	Praca laboratoryjna	Projekt indywidualny	Projekt grupowy	Udział w dyskusji	Referat	Praca pisemna (esej)	Egzamin ustny	Egzamin pisemny	Zadania problemowe
W01													
W02													
U01													
U02													
K01													

Kryteria oceny	Osiągnięcie efektów kształcenia podanych powyżej uprawnia studentów do uzyskania oceny nie wyższej niż dostateczna.
	Warunkiem zaliczenia przedmiotu jest: • zaliczenie dwóch testów/kolokwiów (obejmujących zagadnienia teoretyczne i praktyczne), • zrealizowanie i oddanie projektu zaliczeniowego, • aktywny udział w zajęciach i wykonywanie bieżących zadań.

	Ocena końcowa jest wyliczana na podstawie średniej z dwóch kolokwii oraz oceny projektu. Do zaliczenia wymagane jest uzyskanie pozytywnej oceny z każdego z elementów. Zaliczenie na ocenę dobrą lub bardzo dobrą otrzymuje student, który spełnia warunki oceny dostatecznej, a ponadto: • wykazuje inicjatywę w rozwiązywaniu zadań i proponuje własne podejścia, • potrafi zastosować poznane narzędzia w bardziej złożonych przypadkach analizy danych, • tworzy projekt przemyślany, dobrze udokumentowany i zgodny z dobrymi praktykami, • wykazuje zrozumienie związków między etapami analizy danych i potrafi je świadomie planować. Obecność na wykładach jest warunkiem zaliczenia tej części kursu.
--	--

Uwagi	
-------	--

Treści merytoryczne (wykaz tematów)

1. Wprowadzenie do Data Science: omówienie roli i zastosowań Data Science, etapy cyklu analizy danych (CRISP-DM), przygotowanie środowiska pracy (Anaconda, Jupyter Notebook). 2. Podstawy pracy z danymi w Pythonie: wprowadzenie do bibliotek NumPy i Pandas; operacje na tablicach i strukturach danych, selekcja, filtrowanie, agregacje i przekształcenia. 3. Wprowadzenie do statystyki opisowej: obliczanie miar tendencji centralnej i zmienności, wizualizacja rozkładów, analiza zależności między zmiennymi. 4. Elementy algebry liniowej w analizie danych: operacje macierzowe, rozwiązywanie układów równań liniowych, wartości własne i ich zastosowanie (np. PCA). 5. Odczyt i zapis danych: wczytywanie i zapis danych z/do plików CSV, JSON, komunikacja z API, podstawy pracy z bazą danych SQLite i zapytań SQL. 6. Pandas vs SQL: porównanie możliwości obu narzędzi w zakresie selekcji, filtrowania, łączenia danych i operacji grupujących. 7. Czyszczenie i przygotowanie danych: identyfikacja i uzupełnianie braków danych, standaryzacja, normalizacja, kodowanie zmiennych kategorycznych, tworzenie nowych cech. 8. Agregacja i operacje na grupach: grupowanie danych, obliczenia zbiorcze, analiza dużych zbiorów danych z podziałem na grupy. 9. Wprowadzenie do szeregów czasowych: praca z danymi czasowymi, identyfikacja trendów i sezonowości, podstawowe techniki prognozowania. 10. Eksploracyjna analiza danych (EDA): wykrywanie wartości odstających i anomalii, analiza zależności między zmiennymi, wizualizacja danych (histogramy, wykresy pudełkowe, macierz korelacji).

Wykaz literatury podstawowej

1.

Wykaz literatury uzupełniającej

1.

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

Liczba godzin w kontakcie z prowadzącymi	Wykład	25
	Konwersatorium (ćwiczenia, laboratorium itd.)	30
	Pozostałe godziny kontaktu studenta z prowadzącym	
Liczba godzin pracy studenta bez kontaktu z prowadzącymi	Lektura w ramach przygotowania do zajęć	
	Realizacja zadań domowych (problemowych) po zapoznaniu się z niezbędną literaturą przedmiotu	
	Przygotowanie projektu lub prezentacji na podany temat (praca indywidualna lub w grupie)	
	Przygotowanie do egzaminu/zaliczenia	
Ogółem bilans czasu pracy		
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	15
	Konwersatorium (ćwiczenia, laboratorium itd.)	20
	Pozostałe godziny kontaktu studenta z prowadzącym	
Liczba godzin pracy studenta bez kontaktu z prowadzącymi	Lektura w ramach przygotowania do zajęć	
	Realizacja zadań domowych (problemowych) po zapoznaniu się z niezbędną literaturą przedmiotu	
	Przygotowanie projektu lub prezentacji na podany temat (praca indywidualna lub w grupie)	
	Przygotowanie do egzaminu/zaliczenia	
Ogółem bilans czasu pracy		
Liczba punktów ECTS w zależności od przyjętego przelicznika		3