

Nazwa przedmiotu	Analiza danych tekstowych i ekstrakcja danych
Kod przedmiotu	<i>do wypełnienia przez dziekanat</i>
Profil studiów	ogólnoakademicki
Stopień studiów	pierwszy
Kierunek studiów	Ekonomia
Specjalność (jeśli dotyczy)	Ekonomia Międzynarodowa
Forma studiów	studia stacjonarne i niestacjonarne wieczorowe
Rok studiów	III
Semestr	zimowy
Liczba godzin	30
Forma/typ zajęć	konwersatorium
Punkty ECTS	3
Rodzaj przedmiotu (przedmiot obowiązkowy / kierunkowy do wyboru / fakultatywny / seminarium)	obowiązkowy
Sposób realizacji przedmiotu (stacjonarny / zdalny / kurs internetowy / kurs realizowany metodą blended learning)	stacjonarny
Język wykładowy	polski
Wymagania formalne	brak
Założenia wstępne	Zakłada się, że student posiada podstawową wiedzę z zakresu matematyki i statystyki, rozumie elementarne pojęcia algebry oraz potrafi logicznie analizować problemy. Oczekiwana jest również orientacja w podstawach języka programowania Python.
Skrócony opis przedmiotu	Konwersatorium oferuje kompleksowe wprowadzenie do pracy z danymi za pomocą języka Python. Omawiane zagadnienia obejmują pozyskiwanie danych, ich analizę i wizualizację przy użyciu popularnych bibliotek, takich jak numpy, pandas, matplotlib, seaborn i plotly. Uczestnicy uczą się podstawowych operacji na danych, tworzenia baz danych z wykorzystaniem BigQuery oraz zdobywają umiejętności webscrapingu z użyciem biblioteki requests i selenium. Kurs obejmuje również elementy analizy danych tekstowych i ich praktyczne zastosowania. Zajęcia mają wymiar praktyczny: studenci nauczą się wykorzystywać narzędzia, rozwijając umiejętności w pracy z danymi tekstowymi w badaniach naukowych.
Treści kształcenia dla przedmiotu	Przedmiotem konwersatorium jest zapoznanie studentów z praktycznymi aspektami pracy z danymi z wykorzystaniem języka Python. Omawiany materiał obejmuje cały proces analizy danych: od ich pozyskiwania, przez przetwarzanie i eksplorację, po wizualizację i interpretację wyników. Zajęcia koncentrują się na wykorzystaniu popularnych bibliotek analitycznych oraz narzędzi wspierających pracę z danymi strukturalnymi i tekstowymi. Celem zajęć jest rozwinięcie umiejętności samodzielnej pracy z danymi w środowisku Python, w tym budowy prostych baz danych, automatycznego pozyskiwania informacji z zasobów internetowych oraz realizacji podstawowych analiz danych tekstowych na potrzeby badań naukowych. W kolejnych blokach omawiane są następujące treści:

	• Wprowadzenie do analizy danych w Pythonie. Rola Pythona w analizie danych i badaniach naukowych. Przegląd bibliotek analitycznych. • Praca z danymi numerycznymi i tabelarycznymi. Biblioteki numpy i pandas. Struktury danych (tablice, ramki danych). Podstawowe operacje na danych. • Eksploracyjna analiza danych i wizualizacja. Wprowadzenie do eksploracyjnej analizy danych (EDA). Tworzenie wykresów i wizualizacji z wykorzystaniem bibliotek matplotlib, seaborn i plotly. • Pozyskiwanie danych i web scraping. Źródła danych i formaty plików. Pobieranie danych z API i stron internetowych. Web scraping z wykorzystaniem bibliotek requests i selenium. Aspekty etyczne pozyskiwania danych. • Bazy danych i praca z BigQuery. Podstawy relacyjnych baz danych. Tworzenie i obsługa baz danych w środowisku BigQuery. Zapytania SQL i integracja baz danych z analizą w Pythonie. • Podstawy analizy danych tekstowych w Pythonie. Wstęp do przetwarzania danych tekstowych. Czyszczenie i wstępna obróbka tekstu. Proste metody analizy tekstu i ich zastosowania w badaniach naukowych.
Przedmiotowe efekty uczenia się	Po ukończeniu przedmiotu, student: W ZAKRESIE WIEDZY: • zna i rozumie podstawy pracy z danymi w języku Python, w tym rolę Pythona w analizie danych i badaniach naukowych; • zna i rozumie zasady przetwarzania, eksploracji i wizualizacji danych z wykorzystaniem bibliotek takich jak numpy, pandas, matplotlib, seaborn i plotly; • zna i rozumie podstawowe metody pozyskiwania danych, w tym web scraping oraz pracę z bazami danych, a także możliwości i ograniczenia ich zastosowań; W ZAKRESIE UMIEJĘTNOŚCI: • potrafi samodzielnie przetwarzać i analizować dane numeryczne, tabelaryczne i tekstowe z wykorzystaniem języka Python i odpowiednich bibliotek; • potrafi pozyskiwać dane z różnych źródeł, w tym stron internetowych i baz danych, oraz integrować je w procesie analizy; • potrafi tworzyć czytelne wizualizacje danych oraz interpretować wyniki analiz na potrzeby badań naukowych; W ZAKRESIE KOMPETENCJI: • jest gotów do samodzielnego rozwijania kompetencji programistycznych i analitycznych w obszarze pracy z danymi; • wykazuje świadomość etyczną i odpowiedzialność w pozyskiwaniu, przetwarzaniu i analizie danych, w tym danych tekstowych pochodzących z zasobów internetowych.
Powiązanie efektów przedmiotowych z efektami kierunkowymi/specjalnościowymi (oznaczonymi kodami z programu nauczania)	specjalność Ekonomia Międzynarodowa: S_W05 S_U04 S_K02
Nakład pracy studenta	Szacunkowy nakład pracy studenta: 3 ECTS x 25h = 75h (K) - godziny kontaktowe (S) - godziny pracy samodzielnej konwersatorium (zajęcia): 30h (K) 0h (S) konsultacje: 5h (K) 0h (S) przygotowanie do ćwiczeń: 0h (K) 10h (S) przygotowanie projektu zaliczeniowego: 0h (K) 20h (S)

	praca z materiałami dodatkowymi: 0h (K) 10h (S) Razem: 35h (K) + 40h (S) = 75h
Metody i kryteria oceniania	Uzyskanie zaliczenia przedmiotu wymaga: - obecności na ćwiczeniach (dopuszczalne są dwie nieobecności) i ich zaliczenia; podstawą zaliczenia ćwiczeń jest aktywny udział w pracy zespołowej podczas realizacji projektu zaliczeniowego oraz aktywność w czasie zajęć, polegająca w szczególności na udziale w dyskusji czy wykonywaniu zadań przy użyciu narzędzi informatycznych; - wykonania projektu zaliczeniowego w grupach, obejmującego pozyskanie, przetworzenie, analizę i wizualizację danych (w tym danych tekstowych) z wykorzystaniem poznanych bibliotek i narzędzi programistycznych; - uzyskania co najmniej 51% punktów możliwych do zdobycia w ramach projektu zaliczeniowego, na którego pozytywną ocenę składają się: poprawność metodologiczna zastosowanych rozwiązań, adekwatność doboru narzędzi i algorytmów, jakość interpretacji wyników oraz prezentacja efektów pracy zespołu; - Skala ocen: [0%-51%) – ndst [51%-60%) – dst [60%-70%) – dst + [70%-80%) – db [80%-90%) – db+ [90%-100%) – bdb.
Literatura	Literatura obowiązkowa (wybrane rozdziały): - Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta, Harshit Surana, Przetwarzanie języka naturalnego w praktyce. Przewodnik po budowie rzeczywistych systemów NLP, Helion 2023 - Piotr Wróblewski, Machine learning i natural language processing w programowaniu. Podręcznik z ćwiczeniami w Pythonie, Helion 2024 - Alex J. Gutman, Jordan Goldmeier, Analityk danych. Przewodnik po data science, statystyce i uczeniu maszynowym, Helion 2023 - Jon Krohn, Grant Beyleveld, Aglaé Bassens, Uczenie głębokie i sztuczna inteligencja. Interaktywny przewodnik ilustrowany, Helion 2021 Literatura uzupełniająca (wybrane rozdziały): - Daniel Jurafsky, James H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models, 3rd edition 2025, [https://web.stanford.edu/~jurafsky/slp3/] - Steven Bird, Ewan Klein, Edward Loper, Natural Language Processing with Python, O'Reilly Media 2019, [https://www.nltk.org/book/]
Metody dydaktyczne	W ramach zajęć wykorzystywane są następujące metody dydaktyczne: - mini-wykład interaktywny - ćwiczenia rachunkowe i analityczne - analiza danych w programach komputerowych - webquest
Oprogramowanie	MS Office, zintegrowane środowisko programistyczne (RStudio / Positron / Jupyter Notebook / PyCharm / Google Colab)