

Karta przedmiotu

Nazwa i kod przedmiotu	Przygotowanie danych , PG_00157248						
Kierunek studiów	Informatyka i ekonometria (O)						
Data rozpoczęcia studiów	październik 2024 r.		Rok akademicki realizacji przedmiotu		2024/2025		
Poziom kształcenia	II stopnia		Grupa zajęć		Grupa zajęć obowiązkowych z zakresu kierunku studiów Grupa zajęć powiązanych z prowadzonymi badaniami naukowymi w dziedzinie nauki związanej z kierunkiem - profil ogólnoakademicki		
Forma studiów	niestacjonarne		Sposób realizacji		na uczelni		
Rok studiów	1		Język wykładowy		polski		
Semestr studiów	2		Liczba punktów ECTS		3.0		
Profil kształcenia	ogólnoakademicki		Forma zaliczenia		zaliczenie		
Jednostka prowadząca	Rektor -> Wydział Zarządzania -> Katedra Statystyki						
Imię i nazwisko wykładowcy (wykładowców)	Odpowiedzialny za przedmiot		dr Arkadiusz Kozłowski				
	Prowadzący zajęcia z przedmiotu		dr hab. Krzysztof Najman				
Formy zajęć	Forma zajęć	Wykład	Ćwiczenia	Laboratorium	Projekt	Seminarium	RAZEM
	Liczba godzin zajęć	0.0	15.0	0.0	0.0	0.0	15
	W tym liczba godzin zajęć na odległość: 0.0						
Aktywność studenta i liczba godzin pracy	Aktywność studenta	Udział w zajęciach dydaktycznych, objętych planem studiów		Udział w konsultacjach		Praca własna studenta	RAZEM
	Liczba godzin pracy studenta	15		15.0		45.0	75
Cel przedmiotu	Zapoznanie studentów z teoretycznymi i praktycznymi aspektami etapu przygotowania danych. Zapoznanie studenów ze źródłami błędów występujących na etapie przygotowania danych.						

Efekty uczenia się przedmiotu	Efekt kierunkowy	Efekt z przedmiotu	Sposób weryfikacji i oceny efektu
	[liEMU2_W04] Ma pogłębioną wiedzę o zaawansowanych metodach matematycznych, statystycznych, ekonometrycznych oraz informatycznych umożliwiających pozyskiwanie, przetwarzanie i analizę danych odzwierciedlających funkcjonowanie i wzrost gospodarki narodowej i jej składowych oraz zjawisk i procesów zachodzących w ich otoczeniu.	Student zna i rozumie istotę przygotowania danych do analizy.	[SW4] test/egzamin - ustny lub pisemny
	[liEMU2_K06] Ma świadomość konieczności etycznego, zrównoważonego i społecznie odpowiedzialnego zachowywania się w życiu zawodowym i społecznym. Inicjuje i organizuje działalność na rzecz środowiska społecznego i interesu publicznego.	Potrafi zaprojektować proces analizy danych wskazując potencjalne problemy i metody ich rozwiązania.	[SK4] test/egzamin - ustny lub pisemny
	[liEMU2_K02] Potrafi swobodnie komunikować się z otoczeniem na specjalistyczne tematy z zakresu informatyki i ekonometrii w miejscu pracy i poza nim, przekazywać swoją wiedzę oraz dzielić się swoimi umiejętnościami za pomocą różnych środków przekazu. Kulturalnie uczestniczy w dyskusji, nie boi się zadawać pytań i umie wyrażać konstruktywną krytykę.	Student zna i rozumie źródła problemów procesu przygotowania danych.	[SK4] test/egzamin - ustny lub pisemny
	[liEMU2_K01] Rozumie potrzebę ustawicznego uzupełniania i pogłębiania nabytej wiedzy; inspiruje i organizuje proces uczenia się innych osób.	Student potrafi krytycznie wyjaśnić użyteczność procesu przygotowania danych.	[SK4] test/egzamin - ustny lub pisemny
	[liEMU2_U05] Potrafi na poziomie zaawansowanym dokonać opisu statystycznego podstawowych i szczegółowych kategorii ekonomicznych, sformułować oraz zweryfikować hipotezy odnośnie do ich kształtowania się.	Potrafi wskazać i krytycznie ocenić analizy potrzebne w procesie przygotowania danych.	[SU4] test/egzamin - ustny lub pisemny
	[liEMU2_W05] Ma pogłębioną wiedzę o źródłach danych społeczno-ekonomicznych, ich bazach oraz sposobie ich tworzenia.	Student zna i rozumie istotę przygotowania danych do analizy.	[SW4] test/egzamin - ustny lub pisemny
	[liEMU2_U07] Potrafi budować zaawansowane modele formalne złożonych zjawisk i procesów ekonomicznych, szacować je, przeprowadzać ich weryfikację oraz stosować do modelowania, prognozowania i optymalizacji zasobów instytucji ekonomicznych o zróżnicowanym stopniu skomplikowania, ich struktury oraz przebiegu procesów w nich zachodzących.	Student zna i rozumie źródła problemów procesu przygotowania danych.	[SU4] test/egzamin - ustny lub pisemny
	[liEMU2_K03] Potrafi swobodnie komunikować się z otoczeniem w miejscu pracy i poza nim, przekazywać swoją wiedzę oraz dzielić się swoimi umiejętnościami za pomocą różnych środków przekazu.	Potrafi zaprojektować proces analizy danych wskazując potencjalne problemy i metody ich rozwiązania.	[SK4] test/egzamin - ustny lub pisemny

Treści przedmiotu	Jakość danych statystycznych: istota, definicja, cechy zbioru danych o wysokiej jakości: odpowiedniość dla potrzeb użytkowników, terminowość i punktualność, dostępność i przejrzystość, porównywalność, spójność, dokładność. Podstawowe źródła błędów w danych statystycznych: klasyfikacja błędów losowych i nielosowych, metody ich minimalizacji i korygowania. Etapy procesu przygotowania danych dla analiz Big Data: rejestracja, kodowanie, kontrola, wyrównywanie momentów/okresów czasu, tworzenie jednostek/cech zagregowanych (mikro agregaty), imputacja braków danych, transformacje cech i przypadków, wybór (redukcja) cech i jednostek do analizy. Istota, cel, definicja, podstawowe techniki i ich własności: kodowania danych: skale pomiarowe, osłabianie i wzmacnianie skal, dychotomizacja, wyrównywania momentów/okresów czasu: chronologiczne średnie r-me, tworzenie jednostek/cech zagregowanych (mikroagregaty), imputacja braków danych: modele powstawania braków (MAR, MCAR, MNAR), wybrane metody imputacji dedukcyjnej, hot-deck, średnimi, wielokrotnej, transformacje cech i przypadków, wpływ procedur transformacyjnych na własności statystyczne zbioru danych: standaryzacja, unitaryzacja, unitaryzacja zerowa, przekształcenia ilorazowe, wybór (redukcja) jednostek: procedury losowania jednostek, proces wyboru cech: pełny zbiór cech, generowanie podzbiorów cech, ewaluacja, testowanie warunków stopu, walidacja, końcowy zbiór cech, metody wyboru cech: pełne, heurystyczne, losowe, ważenie cech, zastępowanie pojedynczych cech sztucznymi (w tym analiza korelacyjna, Analiza Głównych Składowych). Analiza danych nietypowych (outliers): istota, definicja, podstawowe techniki wykrywania i testowania danych nietypowych, konsekwencje dla analizy danych, przykłady zastosowań. Anonimizacja danych: istota, cel, aspekty prawne, (prawo polskie, EU i międzynarodowe, opinie GODO), podstawowe techniki anonimizacji danych. Łączenie i integracja zbiorów danych: istota, cel, podstawowe techniki integracji danych: deterministyczne i stochastyczne, parowanie danych.		
Wymagania wstępne i dodatkowe	analiza wielowymiarowa, metoda reprezentacyjna		
Sposoby i kryteria oceniania osiąganych efektów uczenia się	Sposób oceniania (składowe)	Próg zaliczeniowy	Składowa oceny końcowej
	kolokwium	51.0%	100.0%
Zalecana lista lektur	Podstawowa lista lektur	Barnett V., Lewis T., Outliers in Statistical Data. John Wiley, 1994. Balicki A., Migdał Najman K., Najman K., Normalizacja odpowiedzi w badaniach opinii a problem grupowania, Prace i Materiały Wydziału Zarządzania Uniwersytetu Gdańskiego, nr 3, 2008, s. 53-62. Migdał Najman K., Najman K., Data Analysis, Machine Learning and Applications, Applying the Kohonen Self-organizing Map Networks to Selecting Variables, Studies in Classification, Data Analysis and Knowledge Organization, Presisach C., Burkhardt H., Schmidt-Thieme L., Decker R., Springer Verlag Berlin Heidelberg, 2008, s. 45-54. Walesiak M., Problemy selekcji i ważenia zmiennych w zagadnieniu klasyfikacji, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, 1076, Taksonomia, 12, 2005, s. 106-118.	
	Uzupełniająca lista lektur	Kordos J., Jakość danych statystycznych, PWE, 1988. Szymkowiak M., Imputacja i kalibracja - nowe możliwości estymacji w badaniach statystycznych z brakami odpowiedzi, Zeszyty Naukowe Uniwersytetu Ekonomicznego w Poznaniu, 2009 nr 116, s. 90-105.	
	Adresy eZasobów	Adresy na platformie eNauczanie:	
	Przykładowe zagadnienia/ przykładowe pytania/ realizowane zadania		
Praktyki zawodowe w ramach przedmiotu	Nie dotyczy		

Dokument wygenerowany elektronicznie. Nie wymaga pieczęci ani podpisu.