



Karta przedmiotu

Nazwa i kod przedmiotu	Advanced data mining, PG_00045380						
Kierunek studiów	Inżynieria danych						
Data rozpoczęcia studiów	październik 2019 r.			Rok akademicki realizacji przedmiotu		2022/2023	
Poziom kształcenia	I stopnia - inżynierskie		Grupa zajęć		Grupa zajęć fakultatywnych Grupa zajęć powiązanych z prowadzonymi badaniami naukowymi w dziedzinie nauki związanej z kierunkiem - profil ogólnoakademicki		
Forma studiów	stacjonarne		Sposób realizacji		na uczelni		
Rok studiów	4		Język wykładowy		angielski		
Semestr studiów	7		Liczba punktów ECTS		4.0		
Profil kształcenia	ogólnoakademicki		Forma zaliczenia		zaliczenie		
Jednostka prowadząca	Wydział Elektroniki -> Telekomunikacji i Informatyki -> Katedra Inżynierii Oprogramowania						
Imię i nazwisko wykładowcy (wykładowców)	Odpowiedzialny za przedmiot		dr Paweł Weichbroth				
	Prowadzący zajęcia z przedmiotu		dr Paweł Weichbroth				
Formy zajęć i metody nauczania	Forma zajęć	Wykład	Ćwiczenia	Laboratorium	Projekt	Seminarium	RAZEM
	Liczba godzin zajęć	0.0	0.0	30.0	0.0	0.0	30
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	30		0.0		70.0	100
Cel przedmiotu	Celem przedmiotu jest wprowadzenie studentów w tematykę eksploracji oraz wizualizacji dużych zbiorów danych z użyciem skalowalnego klastra obliczeniowego przy wykorzystaniu nowoczesnych języków funkcyjnych oraz pakietów statystycznych.						
Efekty uczenia się przedmiotu	Efekt kierunkowy		Efekt z przedmiotu		Sposób weryfikacji i oceny efektu		
	[K6_W07] zna metody informatycznego przetwarzania, przechowywania, ekstrakcji danych zapisanych w różnych modelach w tym: relacyjnym, grafowym i dokumentowym		Student potrafi oczyszczać duże zbiory danych, agregować, ekstrahować oraz zapisywać w rozproszonym systemie plików, w modelach kolumnowym oraz grafowym		[SW1] Ocena wiedzy faktograficznej		
	[K6_U01] programuje w językach programowania proceduralnego, obiektowego, funkcjonalnego i w logice, koduje programy na poziomie instrukcji procesora, uruchamia i testuje programy		Student potrafi uruchomić rozproszony algorytm w języku funkcyjnym Scala w środowisku Spark. Student potrafi wizualizować dane używając platformy R.		[SU4] Ocena umiejętności korzystania z metod i narzędzi [SU2] Ocena umiejętności analizy informacji		
	[K6_W08] zna modele i strukturę procesu eksploracji danych i ich wielowymiarowe analizy oraz potrafi ocenić wyniki takich analiz		Student potrafi tworzyć systemy rekomendacyjne, klasyfikatory AUC oraz drzewa decyzyjne, grupować dane, wykrywać anomalie. Student potrafi wizualizować wynik przetwarzania.		[SW1] Ocena wiedzy faktograficznej		

Treści przedmiotu	1. Wprowadzenie do języka Scala 2. Wprowadzanie do języka R 3. Języki funkcyjne Scala oraz R w kontekście platformy Spark 4. Przygotowywanie danych 5. Model rekomendacyjny metodą najmniejszych kwadratów 6. Ocena jakości rekomendacji 7. Model drzewa decyzyjnego 8. Dobór hiperparametrów modelu 9. Prognozowanie 10. Wizualizacja danych w R 11. Wykrywanie anomalii (grupowanie metodą K-means)		
Wymagania wstępne i dodatkowe	Znajomość programowania w Java, wskazana wiedza na temat programowania w językach funkcyjnych oraz podstaw statystyki.		
Sposoby i kryteria oceniania osiągniętych efektów uczenia się	Sposób oceniania (składowe)	Próg zaliczeniowy	Składowa oceny końcowej
	ćwiczenia praktyczne	60.0%	100.0%
Zalecana lista lektur	Podstawowa lista lektur	https://www.r-project.org/ http://www.scala-lang.org/ http://spark.apache.org/	
	Uzupełniająca lista lektur	S.Ryza, U.Laserson, S.Owen & J.Wills, Advanced Analytics with Spark (Spark. Zaawansowana analiza danych), O'Reilly (Helion) Karau, H., Konwinski, A., Wendell, P., Zaharia, M.: Learning Spark: Lightning-Fast Big Data Analysis. O'Reilly, 2015	
Przykładowe zagadnienia/ przykładowe pytania/ realizowane zadania	W trakcie zajęć student stworzy oraz oceni model rekomendacji muzyki.		
Praktyki zawodowe w ramach przedmiotu	Nie dotyczy		