

KARTA OPISU MODUŁU KSZTAŁCENIA		
Nazwa modułu/przedmiotu Integracja i eksploracja danych		Kod 1010332521010335195
Kierunek studiów Informatyka	Profil kształcenia (ogólnoakademicki, praktyczny) (brak)	Rok / Semestr 1 / 2
Ścieżka obieralności/specjalność -	Przedmiot oferowany w języku: polski	Kurs (obligatoryjny/obieralny) obligatoryjny
Stopień studiów: II stopień	Forma studiów (stacjonarna/niestacjonarna) stacjonarna	
Godziny Wykłady: 30 Ćwiczenia: - Laboratoria: 15 Projekty/seminaria: 15	Liczba punktów 5	
Status przedmiotu w programie studiów (podstawowy, kierunkowy, inny) (brak)		(ogólnouczelniany, z innego kierunku) (brak)
Obszar(y) kształcenia i dziedzina(y) nauki i sztuki		Podział ECTS (liczba i %)
Odpowiedzialny za przedmiot / wykładowca:		
dr hab. Tadeusz Pankowski email: tadeusz.pankowski@put.poznan.pl tel. 607-033-007 Wydział Elektryczny ul. Piotrowo 3A 60-965 Poznań		
Wymagania wstępne w zakresie wiedzy, umiejętności, kompetencji społecznych:		
1	Wiedza:	Ma uporządkowaną i podbudowaną teoretycznie wiedzę w zakresie baz danych.
2	Umiejętności:	Potrafi pozyskiwać informacje z literatury, baz danych i innych źródeł; potrafi integrować uzyskane informacje, dokonywać ich interpretacji, a także wyciągać wnioski oraz formułować i uzasadniać opinie.
3	Kompetencje społeczne	Ma kompetencje społeczne wynikające z programu szkoły średniej.
Cel przedmiotu:		
Poznanie metod i wybranych narzędzi przeznaczonych do integracji, analizy i eksploracji danych.		
Efekty kształcenia i odniesienie do kierunkowych efektów kształcenia		
Wiedza:		
1. Ma uporządkowaną i podbudowaną teoretycznie wiedzę w zakresie baz danych, hurtowni danych, analizy i eksploracji danych. - [K_W08]		
Umiejętności:		
1. Potrafi pozyskiwać informacje z literatury, baz danych i innych źródeł; potrafi integrować uzyskane informacje, dokonywać ich interpretacji, a także wyciągać wnioski oraz formułować i uzasadniać opinie. - [K_U01]		
2. Potrafi zaprojektować oraz zrealizować prosty system integracji danych; potrafi analizować i eksplorować dane - [K_U12]		
3. Potrafi - przy formułowaniu i rozwiązywaniu informatycznych zadań inżynierskich - dostrzegać ich aspekty pozatechniczne, ekonomiczne i prawne. - [K_U21]		
Kompetencje społeczne:		
1. Ma świadomość odpowiedzialności za pracę własną oraz gotowość podporządkowania się zasadom pracy w zespole i ponoszenia odpowiedzialności za wspólnie realizowane zadania. - [K_K04]		
Sposoby sprawdzenia efektów kształcenia		
Egzamin pisemny, zaliczenie zajęć laboratoryjnych i projektowych na podstawie zrealizowanych zadań i projektów.		
Treści programowe		

<p>Wykłady</p> <p>Znaczenie eksploracji danych (data mining). Eksploracja danych na tle rozwoju nauki i technologii przetwarzania danych. Proces odkrywania wiedzy z danych. Metody integracji danych, modelowania danych wielowymiarowych, budowy hurtowni danych. Język MDX i jego znaczenie w przetwarzaniu danych wielowymiarowych. Drzewa decyzyjne jako metoda klasyfikacji obiektów. Wybór zbioru treningowego, problem analizy i predykcji za pomocą drzew decyzyjnych, dziedziny zastosowań. Budowa drzewa decyzyjnego: podstawy teorii informacji (entropia układu); analiza informacyjna ? miara ilości informacji. Klasyfikacja metodą Bayesa. Odkrywanie reguł asocjacyjnych. Algorytm a priori ? odkrywanie zbiorów częstych. Grupowanie obiektów, analiza skupień (klastrowanie obiektów). Klasyfikacja metod grupowania. Grupowanie hierarchiczne. Miary odległości między klastrami (grupami). Algorytm k-średnich. Algorytm k-medoidów. Norma kosinusowa i jej znaczenie w systemach wyszukiwania informacji. Boolowskie systemy wyszukiwania informacji. Eksploracja danych tekstowych. Wykorzystanie metod statystycznych do opisu i analizy danych.</p>		
<p>Zajęcia laboratoryjne</p> <p>Poznanie narzędzi integracji i eksploracji danych i zaznajamianie się z ich praktycznym wykorzystaniem. SSIS (SQL Server Integration Services - z tworzeniem pakietów przepływu włącznie); SSAS (SQL Server Analysis Services - język MDX, kostki, wyodrębnianie miar, wymiarów, agregacje); język DMX; SSRS (SQL Server Reporting Services - integracją z Excelem 2007. Wykorzystanie wspomnianych narzędzi do eksploracji danych: budowa drzew decyzyjnych, klasyfikacja obiektów metodą Baysa, odkrywania reguł asocjacyjnych, eksploracja tekstów (text mining).</p>		
<p>Zajęcia projektowe</p> <p>Celem zajęć projektowych jest pogłębienie wiedzy w zakresie stosowania metod statystycznych do opisu i analizy danych. Wykonywane są projekty w zakresie: statystyki opisowej, badania zgodności rozkładów, estymacji parametrów, wnioskowania statystycznego, analizy wariancji, analizy korelacji i regresji. Wspomniane zadania i projekty realizowane są z wykorzystaniem systemu Statistica i Excel 2010.</p>		
<p>Literatura podstawowa:</p> <ol style="list-style-type: none"> 1. Sturm J., Hurtownie danych. Microsoft SQL Server. Przewodnik Techniczny, Microsoft Press/APN PROMISE, Warszawa. 2. Han J., Kamber M., Data Mining. Concepts and Techniques, Morgan Kaufmann, 2001. 3. C. Seidman, Zgłębienie i analiza danych w Microsoft SQL Server 2000. Przewodnik Techniczny, Microsoft Press/APN PROMISE, Warszawa, 2002. 4. M. Rabiej, Statystyka z programem Statistica, Helion, Gliwice, 2012. 		
<p>Literatura uzupełniająca:</p> <ol style="list-style-type: none"> 1. D. Mendrala, M. Szeliga, Server SQL 2008. Usługi biznesowe, Helion, Gliwice, 2009. 2. R. Baeza-Yates, B. Ribeiro-Neto, Modern Information Retrieval, ACM Press, 1999. 		
<p>Bilans nakładu pracy przeciętnego studenta</p>		
Czynność		Czas (godz.)
1. Uczestnictwo w wykładach		30
2. Udział w zajęciach laboratoryjnych		15
3. Udział w zajęciach projektowych		15
4. Przygotowanie projektów i ich realizacja		30
5. Studiowanie notatek z zajęć oraz zalecanej literatury przedmiotu		35
<p>Obciążenie pracą studenta</p>		
forma aktywności	godzin	ECTS
Łączny nakład pracy	125	5
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	30	2
Zajęcia o charakterze praktycznym	30	3