

<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>Poznawanie narzędzi integracji i eksploracji danych i zaznajamianie się z ich praktycznym wykorzystaniem. SSIS (SQL Server Integration Services - z tworzeniem pakietów przepływu włączanie); 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>		
<p>Literatura uzupełniająca:</p>		
<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