

Nazwa przedmiotu: **Drażnienie danych**

Forma: Wykład Godzin: 15 Semestr: 3 Rok: II Forma zaliczenie: zaliczenie na ocenę ECTS 2

Wydział: ZIF Kierunek: Informatyka i Ekonometria Tryb: stacjonarny Rodzaj: magisterskie

Specjalność: **Metody i systemy wspomaganie decyzji II**

Tytuł, stopień	Imię i nazwisko lub nazwa	e-mail	telefon	Pok/budd
prof. dr hab	Jerzy Korczak	jerzy.korczak@ue.wroc.pl	713680389	615/Z

1. Wymagania wstępne

Podstawy systemów informacyjnych, Technologie informacyjne, Statystyka w biznesie

2. Program przedmiotu

Ćwiczenia laboratoryjne podzielone są na pięć bloków tematycznych:

1. Proces eksploracji baz danych. Przegląd oprogramowania data mining. Opis obiektów i obserwacji. Przygotowanie danych. Filtrowanie i „czyszczenie” danych. Filtrowanie. Redukcja przestrzeni. Wybór cech charakterystycznych. Metody dyskretyzacji i normalizacji danych. Problem brakujących danych. Analiza przykładowych baz danych ekonomiczno-finansowych.
2. Sformułowanie problemu klasyfikacji. Drzewa indukcyjne – algorytm C4.5. Aspekt kosztu atrybutu. Naiwny klasyfikator Bayes'a. Ekstrakcja reguł. Walidacja. Macierz konfuzji. Współczynniki pomiaru jakości klasyfikacji. Diagram ROC. Zastosowania w zarządzaniu: ocena wniosków kredytowych, przewidywanie bankructwa firm, segmentacja rynku, CRM. Wykrywanie danych nietypowych, anomalii (outliers).
3. Reguły asocjacji. Sformułowanie problemu. Algorytmy odkrywania binarnych reguł asocjacyjnych – Wielopoziomowe i wielowymiarowe reguły asocjacyjne. Wyszukiwanie wzorców sekwencji – Zastosowania w zarządzaniu: analiza koszyka zakupów, handel elektroniczny, analiza transakcji klientów.
4. Metody grupowania. Miary podobieństwa. Grupowanie hierarchiczne. Grupowanie iteracyjno- optymalizacyjne. Problem skalowalności. Algorytmy grupowania - algorytm k-Means. Zastosowania: grupowanie produktów, segmentacja rynku.
5. Data mining w sieciach społecznych. Charakterystyki sieci. Analiza korespondencji elektronicznej. Badanie struktur i hierarchii. Eksploracja opinii. Wizualizacja wiedzy.

3. Metodyka zajęć

Poszczególne klasy problemów prezentowane są wg schematu: od podstaw teoretycznych, poprzez analizę algorytmów i oprogramowania, do zadań praktycznych wspomagających zrozumienie, zapamiętanie i umiejętność wykorzystania podanej wiedzy.

W realizacji zadań praktycznych i projektów wykorzystane będą narzędzia i platformy programowe eksploracji danych.

4. Cel dydaktyczny przedmiotu

Celem wykładu jest zapoznanie studentów z podstawowymi pojęciami, metodami i algorytmami eksploracji danych. Oprócz podstaw teoretycznych, studenci zapoznają się z licznymi przykładami zastosowań w ekonomii i finansach oraz narzędziami programowymi.

5. Wykaz literatury podstawowej (do 4 pozycji)

1. Hand D.J., Mannila H., Smyth P., *Principles of Data Mining*, The MIT Press 2001.
2. Maimon O., Rokach L., *Data Mining and Knowledge Discovery Handbook*, Springer, 2005.
3. Witten J., Eibe F., *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann, 2005.
4. Zdrako M., Larose D., *Eksploracja zasobów internetowych*, Wyd. Nauk. PWN, 2009.

6. Wykaz literatury uzupełniającej

1. Cichosz P., *Systemy uczące się*, WNT, 2000.
2. Korczak J. (ed), *Data Mining and Business Intelligence*, Wyd. Uniw. Ekonomicznego we Wrocławiu., 2011.
3. Mitchel T., *Machine Learning*, McGraw Hill, 1997.
4. Weiss, S., Indurkha, N., *Predictive Data Mining - A Practical Guide*, Morgan Kaufmann Publishers Inc, 1998.