

Eksploracja danych - opis przedmiotu

Informacje ogólne	
Nazwa przedmiotu	Eksploracja danych
Kod przedmiotu	04.2-WE-BEP-ED
Wydział	Wydział Informatyki, Elektrotechniki i Automatyki
Kierunek	Biznes elektroniczny
Profil	praktyczny
Rodzaj studiów	pierwszego stopnia z tyt. inżyniera
Semestr rozpoczęcia	semestr zimowy 2016/2017

Informacje o przedmiocie	
Semestr	2
Liczba punktów ECTS do zdobycia	4
Typ przedmiotu	obowiązkowy
Język nauczania	polski
Sylabus opracował	dr hab. inż. Marek Kowal, prof. UZ

Formy zajęć					
Forma zajęć	Liczba godzin w semestrze (stacjonarne)	Liczba godzin w tygodniu (stacjonarne)	Liczba godzin w semestrze (niestacjonarne)	Liczba godzin w tygodniu (niestacjonarne)	Forma zaliczenia
Wykład	15	1	9	0,6	Zaliczenie na ocenę
Laboratorium	30	2	18	1,2	Zaliczenie na ocenę

Cel przedmiotu

Prezentacja oprogramowania wykorzystywanego do eksploracji danych. Zapoznanie słuchaczy z metodami czyszczenia danych. Prezentacja metod odkrywania asocjacji i sekwencji. Prezentacja metod klasteryzacji danych. Prezentacja metod klasyfikacji danych. Wykształcenie praktycznych umiejętności obsługi wybranych systemów eksploracji danych. Wykształcenie umiejętności w zakresie stosowania metod eksploracji danych w e-biznesie (segmentacji klientów, ocenie ryzyka kredytowego, strategii cross-sellingu, wykrywaniu nadużyć, ukierunkowanej reklamie).

Wymagania wstępne

Zakres tematyczny

Przegląd i charakterystyka oprogramowania wykorzystywanego do eksploracji danych. Metody dostępu do baz danych z poziomu oprogramowani do eksploracji danych. Charakterystyka typowych danych biznesowych podlegających eksploracji. Problem eksploracji dużych wolumenów danych. Przykłady komercyjnych technologii eksploracji danych (IBM Watson Explorer, IBM DATA Mining, Microsoft, SAP, Oracle)

Przygotowanie danych do eksploracji. Odkrywanie danych nadmiarowych i odstających. Uzupełnianie danych brakujących. Próbkowanie danych.

Definicja problemu odkrywania wzorców asocjacji. Rodzaje reguł asocjacyjnych. Miary opisujące statystyczną ważności i siłę reguł asocjacji. Problem analizy koszyka zakupów. Złożoność obliczeniowa problemu odkrywania reguł asocjacji. Omówienie algorytmu Apriori i algorytmu Frequent Pattern Growth. Ćwiczenia praktyczne z wykorzystania poznanych metod odkrywania asocjacji.

Definicja problemu odkrywania wzorców sekwencji. Rodzaje reguł sekwencji. Miary opisujące statystyczną ważności i siłę wzorców sekwencji. Złożoność obliczeniowa problemu odkrywania wzorców sekwencji. Omówienie algorytmów: Generalized Sequential Pattern, Prefix Span oraz klasteryzacji sekwencji. Ćwiczenia praktyczne z wykorzystania poznanych metod odkrywania wzorców sekwencji.

Definicja zagadnienia klasteryzacji danych. Metody grupowania hierarchicznego. Metody grupowania iteracyjno-optymalizacyjnego. Miary odległości stosowane w klasteryzacji. Wyznaczanie charakterystyk klastrow. Metody szacowania liczby klastrow. Ćwiczenia praktyczne z wykorzystania poznanych metod klasteryzacji danych.

Definicja zagadnienia klasyfikacji danych. Przygotowanie danych treningowych i testowych. Miary jakości klasyfikacji. Klasyfikacja danych za pomocą drzew klasyfikacyjnych, algorytmu k najbliższych sąsiadów, naiwnego klasyfikatora Bayesa, sieci neuronowych, algorytm wektorów nośnych. Ćwiczenia praktyczne z wykorzystania poznanych metod klasyfikacji danych.

Zastosowanie metod eksploracji danych: w segmentacji klientów, ocenie ryzyka kredytowego, ukierunkowanej reklamie, strategii cross-sellingu i up-sellingu, wykrywaniu nadużyć.

Metody kształcenia

Wykład - wykład konwencjonalny z wykorzystaniem wideoprojektora.
Laboratorium - ćwiczenia praktyczne w laboratorium komputerowym.

Efekty uczenia się i metody weryfikacji osiągnięcia efektów uczenia się

Opis efektu	Symbole efektów	Metody weryfikacji	Forma zajęć
Zna technologie i narzędzia informatyczne stosowane w eksploracji danych	• K_W04 • K_W07	• test z pytaniami zamkniętymi i otwartymi	• Wykład
Zna typowe zastosowania metod eksploracji danych w e-biznesie	• K_W06 • K_W18	• test z pytaniami zamkniętymi i otwartymi	• Wykład
Zna metody eksploracji danych	• K_W02 • K_W17	• test z pytaniami zamkniętymi i otwartymi	• Wykład
Potrafi przygotować dane do eksploracji wykonując ich czyszczenie	• K_U16 • K_U20	• obserwacja i ocena aktywności na zajęciach • obserwacje i ocena umiejętności praktycznych studenta • sprawdzian z progami punktowymi	• Laboratorium
Potrafi posługiwać się algorytmami odkrywania asocjacji i sekwencji	• K_U02 • K_U16 • K_U18 • K_U20	• obserwacja i ocena aktywności na zajęciach • obserwacje i ocena umiejętności praktycznych studenta • sprawdzian z progami punktowymi	• Laboratorium
Potrafi dobrać metodę klasteryzacji danych, przeprowadzić proces grupowania danych i wyznaczyć charakterystyki klastrów	• K_U02 • K_U16 • K_U18 • K_U20	• obserwacja i ocena aktywności na zajęciach • obserwacje i ocena umiejętności praktycznych studenta • sprawdzian z progami punktowymi	• Laboratorium
Potrafi dobrać metodę klasyfikacji danych, zbudować klasyfikator i przeprowadzić testy dokładności klasyfikacji	• K_U02 • K_U16 • K_U18 • K_U20	• obserwacja i ocena aktywności na zajęciach • obserwacje i ocena umiejętności praktycznych studenta • sprawdzian z progami punktowymi	• Laboratorium
Potrafi zastosować metody eksploracji danych w innowacyjnych rozwiązaniach dla e-biznesu	• K_K03 • K_K04 • K_K08 • K_K09	• aktywność w trakcie zajęć • wykonanie sprawozdań laboratoryjnych	• Laboratorium
Rozumie potrzebę ciągłego uzupełniania i poszerzania wiedzy w zakresie intensywnie rozwijających się metod eksploracji danych	• K_K01 • K_K11	• aktywność w trakcie zajęć • wykonanie sprawozdań laboratoryjnych	• Laboratorium

Warunki zaliczenia

Wykład - warunkiem zaliczenia jest uzyskanie pozytywnych ocen z testów przeprowadzonych, co najmniej raz w semestrze.

Laboratorium - warunkiem zaliczenia jest uzyskanie pozytywnych ocen ze wszystkich ćwiczeń laboratoryjnych oraz kolokwiiów i sprawdzianów przewidzianych w ramach laboratorium.

Składowe oceny końcowej = wykład: 50% + laboratorium: 50%

Literatura podstawowa

1. Hand, D., Mannila, H., Smyth, P., Eksploracja danych, Warszawa, WNT, 2005.
2. Larose, D. T., Odkrywanie wiedzy z danych, Warszawa, PWN, 2006.
3. Larose, D. T., Metody i modele eksploracji danych, Warszawa, PWN, 2008.
4. Kwiatkowska, A. M., Systemy wspomaganie decyzji: jak korzystać z wiedzy i informacji w praktyce, Warszawa, PWN, 2007.
5. Han, J., Kamber, M., Data Mining: Concepts and Techniques, Morgan Kaufmann, 2011.

Literatura uzupełniająca

1. Słupor, K., Automatyczna klasyfikacja obiektów, EXIT, 2005.
2. Berry, M. J. A., Linoff, G. S., Mastering Data Mining: The Art and Science of Customer Relationship Management, Wiley, 1999.

Uwagi

Zmodyfikowane przez dr hab. inż. Marcin Mrugalski, prof. UZ (ostatnia modyfikacja: 29-09-2016 21:13)

Wygenerowano automatycznie z systemu SylabUZ