



Przedmiot: Eksploracja danych				
Forma zajęć: wykład	Semestr: 6	Rok: 3	Wymiar godzin: 15	Punkty ECTS: 3
Forma zaliczenia: egzamin	Typ przedmiotu: obowiązkowy		Język nauczania: polski	
Kierunek: Ekonomia		Tryb: stacjonarne		Rodzaj: licencjackie
Specjalność: analityk rynku				
Katedra: Ekonomii Matematycznej				
Stopień naukowy wykładowcy: dr		Imię i nazwisko wykładowcy: Tomasz Galewski		

Wymagania wstępne (przedmioty wprowadzające):

Statystyka

Program przedmiotu:

- Wprowadzenie do eksploracji danych. Cross-Industry Standard Process- Data Mining (CRISP-DM). Pomiar i dane. Przekształcanie danych. Czyszczenie danych. Obsługa brakujących danych. Identyfikacja błędnych klasyfikacji. Identyfikacja punktów oddalonych. Normalizacja i standaryzacja.
- Postępowanie ze skorelowanymi zmiennymi. Badanie zmiennych jakościowych. Badanie zmiennych numerycznych. Badanie relacji wielowymiarowych. Dyskretyzacja.
- Metody eksploracji danych
 - Algorytm k-najbliższych sąsiadów. Metody nadzorowane i nienadzorowane. Zadanie klasyfikacji.
 - Drzewa decyzyjne. Drzewa klasyfikacyjne i regresyjne. Reguły decyzyjne.
 - Sieci neuronowe. Szacowanie i przewidywanie za pomocą sieci neuronowych. Prosty przykład sieci neuronowej. Współczynnik korekcji (uczenia się). Analiza czułości.
 - Grupowanie hierarchiczne. Metody grupowania. Metoda pojedynczego połączenia. Metoda całkowitego połączenia. Algorytm k-średnich.
 - Sieci Kohonena. Sieci samoorganizujące się. Przykład uczenia sieci Kohonena. Zastosowanie sieci Kohonena do grupowania.
 - Reguły asocjacyjne. Analiza podobieństwa koszyków sklepowych. Wskaźniki wsparcia i ufności.
 - Eksploracja tekstu. Wyszukiwanie dokumentów. Reprezentacje tekstu. Ukryte indeksowanie semantyczne.
 - Eksploracja sieci WWW. Podstawowe algorytmy rankingu stron, eksploracja logów serwera..

Metody dydaktyczne:

Wykład składać się będzie z prezentacji poszczególnych algorytmów data mining

Cele przedmiotu:

Studenci nabędą umiejętności wydobywania informacji z bardzo dużych zbiorów danych. Posłużą to do odnajdowania nieznanych dotąd wzorców i korelacji.

Warunki zaliczenia:

Egzamin - test z pytaniami otwartymi

Literatura podstawowa (do 4 pozycji):

- Odkrywanie wiedzy z danych : wprowadzenie do eksploracji danych / Daniel T. Larose ; z jęz. ang. przeł. Anna Wilbik. - Warszawa : Wydawnictwo Naukowe PWN, 2006.
- Systemy uczące się : rozpoznawanie wzorców, analiza skupień i redukcja wymiarowości / Mirosław Krzyśko [et al.]. - Warszawa : Wydawnictwa Naukowo-Techniczne, 2008.
- Systemy uczące się / Paweł Cihończak - Wzrost - Warszawa : WNT, 2007

Literatura uzupełniająca (do 4 pozycji):