

KARTA OPISU MODUŁU KSZTAŁCENIA		
Nazwa modułu/przedmiotu Eksploracja i analiza danych		Kod 1010512311010513907
Kierunek studiów Informatyka	Profil kształcenia (ogólnoakademicki, praktyczny) ogólnoakademicki	Rok / Semestr 1 / 1
Ścieżka obieralności/specjalność Software Engineering (Inżynieria)	Przedmiot oferowany w języku: angielski	Kurs (obligatoryjny/obieralny) obligatoryjny
Stopień studiów: II stopień	Forma studiów (stacjonarna/niestacjonarna) stacjonarna	
Godziny Wykłady: 30 Ćwiczenia: - Laboratoria: 30 Projekty/seminaria: -		Liczba punktów 5
Status przedmiotu w programie studiów (podstawowy, kierunkowy, inny) kierunkowy		(ogólnouczelniany, z innego kierunku) z danego kierunku
Obszar(y) kształcenia i dziedzina(y) nauki i sztuki nauki techniczne		Podział ECTS (liczba i %) 5 100%
Odpowiedzialny za przedmiot / wykładowca:		
<p>dr hab. inż. Jerzy Stefanowski, prof. PP email: Jerzy.Stefanowski@cs.put.poznan.pl tel. 61 6652933 Instytut Informatyki PP ul. Piotrowo 2, 60-965 Poznań</p>		
Wymagania wstępne w zakresie wiedzy, umiejętności, kompetencji społecznych:		
1	Wiedza:	Efekty kształcenia ze studiów I stopnia prezentowane są w serwisie internetowym wydziału www.fc.put.poznan.pl Student rozpoczynający ten przedmiot powinien posiadać z rachunku prawdopodobieństwa (twierdzenie Bayesa, podstawowe rozkłady zmiennych), statystyki matematycznej (zasady testowania hipotez).
2	Umiejętności:	Student rozpoczynający ten przedmiot powinien posiadać z rachunku prawdopodobieństwa (twierdzenie Bayesa, podstawowe rozkłady zmiennych), statystyki matematycznej (zasady testowania hipotez). Powinien posiadać umiejętność rozwiązywania podstawowych problemów ze statystyki oraz rachunku prawdopodobieństwa (wykorzystanie testów statystycznych) oraz umiejętność pozyskiwania informacji ze wskazanych źródeł. Powinien również rozumieć konieczność poszerzania swoich kompetencji / mieć gotowość do podjęcia współpracy w ramach zespołu.
3	Kompetencje społeczne	Powinien również rozumieć konieczność poszerzania swoich kompetencji / mieć gotowość do podjęcia współpracy w ramach zespołu. Ponadto w zakresie kompetencji społecznych student musi prezentować takie postawy jak uczciwość, odpowiedzialność, wytrwałość, ciekawość poznawcza, kreatywność, kultura osobista, szacunek dla innych ludzi.
Cel przedmiotu:		
<ol style="list-style-type: none"> Przekazanie studentom podstawowej wiedzy na temat wybranych metod statystycznej analizy wielowymiarowej, testów nieparametrycznej oraz wybranych metod klasyfikacji z zakresu eksploracji danych. Rozwijanie u studentów umiejętności rozwiązywania problemów z zakresu zarządzania projektami, ekonomii, analizy zachowań ludzkich wymagających stosowania narzędzi statystycznych oraz algorytmów eksploracji danych. Nabywanie powyższych umiejętności poprzez rozwiązywanie na ćwiczeniach praktycznych zadań. Kształtowanie u studentów umiejętności przeprowadzanie eksperymentów z danymi dot. powyższych zadań przy wykorzystaniu różnorodnych gotowych narzędzi programistycznych. 		
Efekty kształcenia i odniesienie do kierunkowych efektów kształcenia		
Wiedza:		

<p>1. Otrzyma uporządkowaną, podbudowaną teoretycznie wiedzę ogólną w zakresie statystycznej analizy danych oraz metod eksploracji danych - [K2st_W2]</p> <p>2. Pozyska szczegółową wiedzę związaną z wybranymi zagadnieniami, takimi jak: predykcja z wykorzystaniem analizy regresji oraz wybranych testów nieparametrycznych dla analizy danych ankietowych, metody konstruowania klasyfikatorów, a także grupowania danych. - [K2st_W3]</p> <p>3. Ma wiedzę o trendach rozwojowych i najistotniejszych nowych osiągnięciach w dziedzinach metod klasyfikacji danych oraz wykorzystania ich w zastosowaniach, w szczególności z zarządzania w odniesieniu do inżynierii oprogramowania - [K2st_W4]</p> <p>4. Zna podstawy przeprowadzanie badań eksperymentalnych z zakresu analizy danych oraz zasad przetwarzanie danych pozyskanych w ten sposób. - [K2st_W6]</p>
<p>Umiejętności:</p> <p>1. Potrafi pozyskiwać informacje nt. analizy i eksploracji danych z literatury, baz danych oraz innych źródeł (w języku ojczystym i angielskim), integrować je, dokonywać ich interpretacji i krytycznej oceny, wyciągać wnioski oraz formułować i wyczerpująco uzasadniać opinie - [K2st_U1, K2st_U4, K2st_U3]</p> <p>2. Potrafi określić kierunki dalszego uczenia się i zrealizować proces samokształcenia ? w szczególności w zakresie wykorzystania algorytmów analizy i eksploracji danych. - [K2st_U6 K2st_U3, K2st_U10]</p> <p>3. Potrafi wykorzystać do formułowania i rozwiązywania zadań inżynierskich i prostych problemów badawczych testy statystyczne, metody klasyfikacji nadzorowanej, predykcji zmiennej liczbowej oraz grupowania danych - [K2st_U6]</p> <p>4. Potrafi ? przy formułowaniu i rozwiązywaniu zadań inżynierskich ? integrować wiedzę z różnych obszarów Informatyki powiązanych z analizą i eksploracją danych. - [K2st_U5]</p> <p>5. Potrafi wykonywać proste eksperymenty badawcze a także formułować i testować hipotezy związane z zadanymi problemami. - [K2st_U4]</p> <p>6. Potrafi ocenić zalety i ograniczenia wybranych algorytmów i ich implementacji w zależności od charakterystyki zadania. - [K2st_U8, K2st_U9]</p>
<p>Kompetencje społeczne:</p> <p>1. Rozumie, że przy systemów informatycznych nabyta wiedza i umiejętności wymagają dalszego kształcenia się z uwagi na dynamiczny rozwój dziedziny - [K2st_K1]</p> <p>2. Zna przykłady procesu eksploracji i analizy danych - rozumie ich zalety i ograniczenia - [K2st_K2]</p> <p>3. Potrafi odpowiednio określić priorytety służące realizacji zadania w którym należy wykorzystać metody analizy danych - [Kst2_K2]</p>

Sposoby sprawdzenia efektów kształcenia

Ogólna ocena obejmuje

a) w zakresie wykładów:

? na podstawie odpowiedzi na pytania dotyczące materiału omówionego na poprzednich wykładach oraz ćwiczonych praktycznie w ramach laboratoriów ? realizowana poprzez egzamin pisemny

b) w zakresie laboratoriów:

? na podstawie oceny bieżącego postępu realizacji zadań ? ćwiczenia przygotowane w postaci instrukcji udostępnionych studentom, przygotowanych samodzielnych raportów z rozwiązania powyższych zadań

Ocena podsumowująca:

a) w zakresie wykładów weryfikowanie założonych efektów kształcenia realizowane jest przez:

? ocenę wiedzy i umiejętności wykazanych na egzaminie pisemnym o charakterze prostych mini-zadań, pytań otwartych albo w formie testu wielokrotnego wyboru ? egzamin może liczyć od 7 do 11 takich pytań w zależności od ich formy,

? omówienie wyników egzaminu,

b) w zakresie laboratoriów weryfikowanie założonych efektów kształcenia realizowane jest przez:

? ocenę przygotowania studenta do poszczególnych sesji zajęć laboratoryjnych (sprawdzian "wejściowy")

? ocenianie ciągle, na każdych zajęciach (odpowiedzi ustne) ? premiowanie przyrostu umiejętności posługiwania się poznanymi zasadami i metodami oraz narzędziami programowymi,

? ocenę sprawozdania przygotowywanego częściowo w trakcie zajęć, a częściowo po ich zakończeniu,

? ocenę wiedzy i umiejętności związanych z realizacją zadań projektowych / laboratoryjnych poprzez 2 kolokwia w semestrze,

Uzyskiwanie punktów dodatkowych za aktywność podczas zajęć, a szczególnie za:

? omówienia dodatkowych aspektów zagadnienia,

? uwagi związane z udoskonaleniem materiałów dydaktycznych.

Treści programowe

Program wykładu obejmuje następujące zagadnienia:

Eksploracja danych a statystyczna analiza danych - różnice. Zastosowanie metod statystycznych do oceny jakości danych. Analiza regresji (modele wielowymiarowe, metody weryfikacji, regresja nieliniowa, selekcja zmiennych), wielowarstwowe sieci neuronowe. Ich zastosowania do predykcji na przykładzie zagadnień zarządzania i ekonomii. Metody tworzenia systemów klasyfikujących (drzewa decyzyjne, inne wybrane metody). Analiza skupień (algorytmy iteracyjno- optymalizacyjne oraz hierarchiczne). Wybrane testy nieparametryczne w analizie danych jakościowych. Metody prowadzenie badań ankietowych

oraz badania statystycznego. Zastosowania w problemach inżynierii oprogramowania. Metody wizualizacji i prezentacji danych. Przegląd dostępnego oprogramowania do analizy eksploracji danych.
Zajęcia laboratoryjne prowadzone są w formie piętnastu 2-godzinnych ćwiczeń, odbywających się w laboratorium. Ćwiczenia realizowane są indywidualnie lub w przypadku b. dużej liczny studentów przez 2-osobowe zespoły.

Program laboratorium obejmuje głównie następujące zagadnienia:

Ocena wiarygodności modeli regresji wielowymiarowej. Dobór funkcji regresji nieliniowej w zależności od charakterystyki zadania. Konstruowanie warstwowej sieci neuronowych z neuronami nieliniowymi. Zasady strojenia parametrów algorytmu wstecznej propagacji błędów. Rozwiązywanie zdania z wykorzystaniem drzew decyzyjnych (ID3-C4.5). Wstępne przetwarzanie danych. Metody dyskretyzacji atrybutów liczbowych. Algorytmy k-średnich i AHC. Analiza danych ankietowych z wykorzystaniem testów nieparametrycznych. Zapoznanie się z oprogramowaniem komercyjnych jak i narzędziami z kategorii wolnego oprogramowania

Materiały dydaktyczne - dostępne na serwerze WWW

1. wykład: prezentacja multimedialna, prezentacja ilustrowana przykładami podawanymi na tablicy, rozwiązywanie prostych zadań, demonstracja użycia wybranego oprogramowania
2. ćwiczenia laboratoryjne: rozwiązywanie zadań, ćwiczenia praktyczne, wykonywanie eksperymentów, dyskusja, studium przypadków, demonstracja

Literatura podstawowa:

1. Discovering Knowledge in Data. An Introduction to Data Mining. (tłumaczenie polskie Odkrywanie wiedzy z danych. Wprowadzenie do eksploracji danych, PWN, 2006)., D.Larose, Wiley, 2005
2. Statystyka w zarządzaniu. A. Aczel, PWN, Warszawa, 2002.
3. Statystyczne systemy uczące się. J.Koronacki, J.Ćwik, EXIT, Warszawa 2008.

Literatura uzupełniająca:

1. Principles of Data Mining. (tłumaczenie polskie Eksploracja danych, WNT, 2005)., D.Hand, H.Mannila, P.Smyth, MIT Press, Cambridge US, 2001
2. Business Intelligence ? Systemy wspomaganie decyzji biznesowych. J.Surma, PWN, Warszawa, 2009.

Bilans nakładu pracy przeciętnego studenta

Czynność	Czas (godz.)
1. Udział w zajęciach laboratoryjnych / ćwiczeniach : 15 x 2 godz.,	30
2. Przygotowanie do ćwiczeń laboratoryjnych	10
3. Dokończenie (w ramach pracy własnej) sprawozdań z ćwiczeń laboratoryjnych	10
4. Udział w konsultacjach związanych z realizacją procesu kształcenia, w szczególności ćwiczeń laboratoryjnych	2 10
5. przygotowanie do sprawdzianów / kolokwium	30
6. Udział w wykładach	15
7. Zapoznanie się ze wskazaną literaturą / materiałami dydaktycznymi	20
8. Przygotowanie do egzaminu i obecność na nim	

Obciążenie pracą studenta

forma aktywności	godzin	ECTS
Łączny nakład pracy	127	5
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	62	2
Zajęcia o charakterze praktycznym	60	2