

KARTA OPISU MODUŁU KSZTAŁCENIA		
Nazwa modułu/przedmiotu Zaawansowana eksploracja danych		Kod 1010512321010513981
Kierunek studiów Informatyka	Profil kształcenia (ogólnoakademicki, praktyczny) ogólnoakademicki	Rok / Semestr 1 / 2
Ścieżka obieralności/specjalność Systemy informatyczne w zarządzaniu	Przedmiot oferowany w języku: polski	Kurs (obligatoryjny/obieralny) obieralny
Stopień studiów: II stopień	Forma studiów (stacjonarna/niestacjonarna) stacjonarna	
Godziny Wykłady: 30 Ćwiczenia: - Laboratoria: 30 Projekty/seminaria: -		Liczba punktów 4
Status przedmiotu w programie studiów (podstawowy, kierunkowy, inny) (ogólnouczelniany, z innego kierunku) kierunkowy z danego kierunku		
Obszar(y) kształcenia i dziedzina(y) nauki i sztuki nauki techniczne		Podział ECTS (liczba i %) 4 100%
Odpowiedzialny za przedmiot / wykładowca: prof. dr hab. inż. Tadeusz Morzy email: Tadeusz.Morzy@put.poznan.pl tel. 61 6652906 Instytut Informatyki ul. Piotrowo 2, 60-965 Poznań		
Wymagania wstępne w zakresie wiedzy, umiejętności, kompetencji społecznych:		
1	Wiedza:	Efekty kształcenia ze studiów I stopnia zdefiniowane w Uchwale Senatu PP, a szczególnie efekty K_W1-2, K_W4, K_W6-15, K_U1-2, K_U4, K_U7-8, K_U14-20, K_U22-23, K_U26, K_K1-9, weryfikowane w procesie rekrutacji na studia 2 stopnia ? efekty te prezentowane są w serwisie internetowym wydziału www.fc.put.poznan.pl Student rozpoczynający ten przedmiot powinien posiadać podstawową wiedzę w zakresie podstawowych pojęć, technik i algorytmów eksploracji danych. Ponadto przydatna jest wiedza ze statystyki matematycznej i rachunku prawdopodobieństwa i podstawowa wiedza w zakresie teorii grafów.
2	Umiejętności:	Student powinien posiadać umiejętność rozwiązywania podstawowych problemów z eksploracji danych oraz umiejętność pozyskiwania informacji ze wskazanych źródeł.
3	Kompetencje społeczne	Powinien również rozumieć konieczność poszerzania swoich kompetencji i 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 zaawansowanych algorytmów eksploracji bardziej złożonych reprezentacji danych oraz realizacji różnych etapów procesu odkrywania wiedzy z danych (w tym przetwarzania wstępnego danych i oceny wyników eksploracji). Rozwijanie u studentów umiejętności rozwiązywania problemów w powyższej dziedzinie (poprzez tzw. case studies odnoszące się do różnorodnych reprezentacji danych i zadań uczenia maszynowego). Nabycie powyższych umiejętności poprzez rozwiązywanie na ćwiczeniach laboratoryjnych praktycznych zadań klasyfikacji nadzorowanej, nienadzorowanej, predykcji z danych zależnych od czasu oraz eksploracji danych społecznościowych. Przekazanie wiedzy na temat technik i algorytmów umożliwiających odkrywanie wiedzy i wzorców, ze szczególnym naciskiem na dane relacyjne i tekstowe. Kształtowanie u studentów umiejętności przeprowadzania powtarzalnych eksperymentów z danymi dot. powyższych zadań przy wykorzystaniu różnorodnych języków programowania (R, python). 		
Efekty kształcenia i odniesienie do kierunkowych efektów kształcenia		
Wiedza:		

<p>1. ma szczegółową wiedzę w zakresie wybranych działów matematyki, w szczególności w obszarze statystyki, rachunku prawdopodobieństwa, analizy matematycznej (regresja) - [K_W3]</p> <p>2. ma uporządkowaną, podbudowaną teoretycznie wiedzę ogólną w zakresie eksploracji danych - także wobec złożonych typów i modeli reprezentacji danych - [K_W4]</p> <p>3. ma podbudowaną teoretycznie szczegółową wiedzę związaną z wybranymi zagadnieniami, takimi jak: metody przetwarzania wstępnego danych, budowa modeli predykcji zmiennej liczbowej (regresja, sieci neuronowe), metody wyboru i oceny klasyfikatorów oraz predykcji, ocena algorytmów skupień, eksploracja strumieni danych, metody wizualnej eksploracji danych, przetwarzanie języka naturalnego, - [K_W5]</p> <p>4. 3. ma podbudowaną teoretycznie szczegółową wiedzę związaną z wybranymi zagadnieniami, takimi jak: rozpraszania eksploracji danych na wiele maszyn obliczeniowych, wybranych klasyfikatorów dla strumieni danych i adaptujących się do zmiennych środowisk - [K_W5]</p> <p>5. ma wiedzę o trendach rozwojowych i najistotniejszych nowych osiągnięciach w eksploracji danych, zwłaszcza w odniesieniu do złożonych reprezentacji danych - [K_W6]</p> <p>6. zna podstawowe metody, techniki i narzędzia stosowane przy rozwiązywaniu złożonych zadań z zastosowaniem algorytmów eksploracji danych, zna zagadnienia przetwarzania języka naturalnego i danych relacyjnych, potrafi wdrożyć opracowywane modele do projektowanych systemów informatycznych. - [K_W8]</p>
<p>Umiejętności:</p> <p>1. potrafi pozyskiwać informacje nt. 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 opinie - [K_U1]</p> <p>2. potrafi określić kierunki dalszego uczenia się i zrealizować proces samokształcenia - w szczególności w zakresie wykorzystania algorytmów eksploracji danych, poprzez lekturę literatury naukowej i popularnonaukowej - [K_U5]</p> <p>3. potrafi wykorzystać do formułowania i rozwiązywania zadań inżynierskich i prostych problemów badawczych metody klasyfikacji nadzorowanej, predykcji zmiennej liczbowej, grupowania danych oraz przetwarzania języka naturalnego - [K_U9]</p> <p>4. potrafi - przy formułowaniu i rozwiązywaniu zadań inżynierskich - integrować wiedzę z różnych obszarów informatyki związanych z pozyskiwaniem danych z różnych źródeł, ich przetwarzaniem wstępnym oraz eksploracją, oceną uzyskanych wzorców i zastosowaniem znalezionych wzorców - [K_U10]</p> <p>5. potrafi wykonywać proste eksperymenty badawcze, wykorzystując do tego języki programowania R i Python, a także komunikować wyniki eksperymentów za pomocą powtarzalnych raportów (np. knitr i jupyter notebook) - [K_U12]</p> <p>6. potrafi ocenić zalety i ograniczenia wybranych algorytmów eksploracji danych i ich implementacji w zależności od charakterystyki zadania. - [K_U13]</p>
<p>Kompetencje społeczne:</p> <p>1. rozumie, że przy tworzeniu inteligentnych systemów wykorzystujących techniki eksploracji danych i podczas adaptacji tych technik do środowiska nabyta wiedza i umiejętności szczególnie szybko się zmieniają i wymagają od osoby dalszego kształcenia się z uwagi na dynamiczny rozwój dziedziny - [K_K1]</p> <p>2. zna przykłady rozwiązań systemów eksploracji danych i rozumie ich ograniczenia, potrafi podać przykłady dobrych i złych praktyk w eksploracji danych - [K_K4]</p> <p>3. potrafi odpowiednio określić priorytety służące realizacji zadania, w którym należy wykorzystać zaawansowaną eksplorację danych - [K_K6]</p>

Sposoby sprawdzenia efektów kształcenia

Ocena formująca:

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 - ćwiczeń oraz projektów (związanych z rozwiązaniem tzw. case studies) przygotowane w postaci instrukcji udostępnionych studentom, przygotowanych samodzielnych raportów z rozwiązania powyższych zadań, wykonywanie ćwiczeń przy użyciu różnych środowisk i narzędzi informatycznych na podstawie pisemnych instrukcji

Ocena podsumowująca:

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

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

- omówienie wyników kolokwium,

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

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

- kartkówki przed niektórymi zajęciami dydaktycznymi,

- ocenę sprawozdania z realizacji case study; ocena ta obejmuje także umiejętność pracy w zespole - jeśli ćwiczenie było wykonywane w zespołach 2 osobowych,

- ocenę z tworzenia prostej aplikacji internetowej korzystającej z algorytmów eksploracji danych

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

- omówienia dodatkowych aspektów zagadnienia,

- uwagi związane z udoskonaleniem materiałów dydaktycznych,

- udział w międzynarodowych konkursach algorytmicznych, ze szczególnym naciskiem na pracę zespołową

Treści programowe

Program wykładu obejmuje następujące zagadnienia:

Charakterystyka procesu odkrywania wiedzy w bazach danych. Główne metody przetwarzania wstępnego danych (w szczególności wykrywanie sytuacji konfliktowych podczas połączenia różnych źródeł, oczyszczenie danych z błędów, uwzględnianie niezdefiniowanych wartości atrybutów), redukcja wymiarowości danych (selekcja cech, tworzenie nowych cech, metody projekcji do przestrzeni niskowymiarowych), transformacje oraz metody dyskretyzacji. Modele predykcyjne dla zmiennych liczbowych wielowymiarowa i nieliniowa analiza regresji (także wykorzystanie sieci neuronowych jako alternatywy); Analiza szeregów czasowych. Zagadnienia metodologiczne oceny wiedzy (dla modeli klasyfikacyjnych i predykcyjnych; miary oceny oraz metody eksperymentalnego oszacowania tych miar). Eksploracja danych silnie niezrównoważonych. Ocena wyników analizy skupień. Przetwarzanie złożonych reprezentacji danych oraz danych dostępnych przyrostowo lub w postaci tzw. strumieni danych. Konstruowanie klasyfikatorów w środowisku danych strumieniowych, gdzie definicja pojęć ulega zmianom wraz z upływem czasu. Elementy wizualizacji danych oraz metody wizualnej eksploracji złożonych typów danych oraz danych wysoce wielowymiarowych. Związki eksploracji danych z komercyjnymi systemami zarządzania przedsiębiorstwami oraz Business Intelligence.

Przebieg ćwiczeń laboratoryjnych obejmuje naukę eksploracji danych z wykorzystaniem języków R i Python. Ponadto dwa studia przypadków, mają na celu ukierunkować studentów na aspekty praktyczne realizacji różnych etapów w odkrywaniu wiedzy z wybranych zbiorów danych. W trakcie ćwiczeń laboratoryjnych studenci przechodzą kurs języka R i poznają pakiety R przydatne do rozwiązywania problemów regresji, klasyfikacji, analizy skupień, wizualizacji i wstępnego przetwarzania danych. Studenci wykonują również w R aplikację internetową pozwalającą stworzyć prototyp produktu wykorzystującego algorytmy eksploracji danych. Studenci zapoznają się także z bibliotekami uczenia maszynowego dla języka programowania Python, ze szczególnym naciskiem na biblioteki wspomagające przetwarzanie języka naturalnego (nlTK, gensim). Wybrane laboratoria są także poświęcone technikom wizualnej oceny i eksploracji danych oraz tworzenia powtarzalnych eksperymentów za pomocą bibliotek caret, knitr (R) oraz scikit-learn, jupyter notebook (Python). Studenci poznają również sposoby przetwarzania większych zbiorów danych poprzez rozpraszanie obliczeń (AWS, Apache Spark).

Metody dydaktyczne:

1. wykład: prezentacja multimedialna, prezentacja ilustrowana przykładami podawanymi na tablicy, rozwiązywanie prostych zadań, demonstracja użycia wybranego oprogramowania

2. ćwiczenia laboratoryjne: rozwiązywanie zadań, wykonywanie eksperymentów, dyskusja, studium przypadków, gry obliczeniowe i konkursy programistyczne, kartkówki

Literatura podstawowa:

1. Document Engineering, Robert J. Glushko, Tim McGrath, MIT Press, Cambridge, 2005

2. Adobe Flex LiveDocs - <http://livedocs.adobe.com/flex/>

Literatura uzupełniająca:		
1. Yii 1.1 Application Development Cookbook, Alexander Makarov, Packt Publishing, Birmingham, 2011		
2. Agile Web Application Development with Yii 1.1 and PHP5, Jeffrey Winesett, Packt Publishing, Birmingham, 2010		
3. jQuery Documentation - http://api.jquery.com		
Bilans nakładu pracy przeciętnego studenta		
Czynność	Czas (godz.)	
1. Udział w zajęciach laboratoryjnych	30	
2. Udział w wykładach	30	
3. Udział w konsultacjach związanych z realizacją ćwiczeń laboratoryjnych (częściowo realizowane drogą elektroniczną)	5	
4. Przygotowanie do ćwiczeń laboratoryjnych	5	
5. Przygotowanie do egzaminu i obecność na kolokwium zaliczeniowym: 8 godz. + 2 godz.	10	
6. Zapoznanie się ze wskazaną literaturą (50 stron)	5	
7. Pisanie i testowanie programów przygotowywanych na konkursy algorytmiczne	10	
8. Wypełnienie quizów i realizacja zadań samodzielnych do poszczególnych zajęć laboratoryjnych	15	
Obciążenie pracą studenta		
forma aktywności	godzin	ECTS
Łączny nakład pracy	110	4
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	67	3
Zajęcia o charakterze praktycznym	60	2