

KARTA OPISU MODUŁU KSZTAŁCENIA		
Nazwa modułu/przedmiotu Odkrywanie wiedzy		Kod 1010512321010510320
Kierunek studiów Informatyka	Profil kształcenia (ogólnoakademicki, praktyczny) ogólnoakademicki	Rok / Semestr 1 / 2
Ścieżka obieralności/specjalność Technologie wytwarzania oprogramowania	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 5
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 nauki techniczne		Podział ECTS (liczba i %) 5 100% 5 100%
Odpowiedzialny za przedmiot / wykładowca: dr inż. Marcin Szelaąg email: marcin.szelaag@cs.put.poznan.pl tel. 61 665 3023 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 w sprawie zatwierdzenia kierunkowych efektów kształcenia dla studiów prowadzonych na Politechnice Poznańskiej nr 42 z dnia 24 kwietnia 2017 roku, a szczególnie efekty K1st_W1-8, weryfikowane w procesie rekrutacji na studia 2 stopnia - efekty te prezentowane są w serwisie internetowym wydziału www.fc.put.poznan.pl
2	Umiejętności:	Efekty kształcenia ze studiów I stopnia zdefiniowane w Uchwale Senatu w sprawie zatwierdzenia kierunkowych efektów kształcenia dla studiów prowadzonych na Politechnice Poznańskiej nr 42 z dnia 24 kwietnia 2017 roku, a szczególnie efekty K1st_U2-14, weryfikowane w procesie rekrutacji na studia 2 stopnia - efekty te prezentowane są w serwisie internetowym wydziału www.fc.put.poznan.pl
3	Kompetencje społeczne	Efekty kształcenia ze studiów I stopnia zdefiniowane w Uchwale Senatu w sprawie zatwierdzenia kierunkowych efektów kształcenia dla studiów prowadzonych na Politechnice Poznańskiej nr 42 z dnia 24 kwietnia 2017 roku - efekty te prezentowane są w serwisie internetowym wydziału www.fc.put.poznan.pl
Cel przedmiotu: 1. Przekazanie studentom wiedzy w zakresie: wybranych technik przetwarzania danych pod kątem ich dalszej analizy, wybranych metod analizy skupień (grupowania), algorytmów i technik tworzenia regułowych modeli wiedzy, metod odkrywania wiedzy z danych niekompletnych (zawierających tzw. wartości brakujące), systemów wnioskowania na podstawie podobieństwa do wcześniej napotkanych przykładów (ang. case-based reasoning), modelowania systemów z wykorzystaniem zbiorów rozmytych (ang. fuzzy sets), funkcji oferowanych przez wybrane nowoczesne systemy odkrywania wiedzy, programowania i uczenia inteligentnych programów grających, w szczególności silników szachowych (ang. chess engines). 2. Rozwijanie u studentów umiejętności rozwiązywania problemów praktycznych związanych z odkrywaniem ukrytych zależności w danych i modelowaniem pozyskanej wiedzy. Odkrywanie wiedzy w rzeczywistych zbiorach danych z wykorzystaniem poznanych technik i narzędzi. Tworzenie uczących się programów grających, w szczególności silników szachowych. 3. Kształtowanie u studentów umiejętności pracy zespołowej.		
Efekty kształcenia i odniesienie do kierunkowych efektów kształcenia		
Wiedza:		

<ol style="list-style-type: none">1. zna metody reprezentacji gry, oceny pozycji i uczenia się stosowane w programowaniu silników szachowych - [K2st_W1]2. ma zaawansowaną wiedzę szczegółową dotyczącą wybranych numerycznych i konceptualnych algorytmów analizy skupień (grupowania) - [K2st_W3]3. ma zaawansowaną wiedzę szczegółową dotyczącą algorytmu indukcji reguł decyzyjnych VCDomLEM - [K2st_W3]4. ma zaawansowaną wiedzę szczegółową dotyczącą zastosowania DRSA do uczenia się na podstawie podobieństwa - metoda DRSA-CBR - [K2st_W3]5. ma zaawansowaną wiedzę szczegółową dotyczącą systemów wnioskowania rozmytego Mamdani i Takagi-Sugeno - [K2st_W3]6. ma szczegółową wiedzę o kolejnych etapach zgodnego ze standardem CRISP-DM procesu odkrywania wiedzy z danych prowadzącego do stworzenia systemu wspomaganego decyzji na podstawie pozyskanej wiedzy - [K2st_W5]7. zna techniki radzenia sobie z problemem wartości brakujących w odkrywaniu wiedzy z danych w problemach klasyfikacji - [K2st_W6]8. zna różne algorytmy eksploatacji wartościowanej relacji binarnej na zbiorze obiektów prowadzące do utworzenia rankingu tych obiektów - [K2st_W6]
Umiejętności:
<ol style="list-style-type: none">1. potrafi w procesie odkrywania wiedzy z danych integrować wiedzę ze statystyki, wspomaganego decyzji i uczenia maszynowego, uwzględniając przy tym dostępną wiedzę dziedzinową - [K2st_U5]2. potrafi ocenić przydatność dostępnych środowisk analizy danych i odkrywania wiedzy do rozwiązywania konkretnych problemów praktycznych - [K2st_U6]3. potrafi dobrać odpowiednie przekształcenia zbioru danych zwiększające efektywność klasyfikatora uczonego na tym zbiorze danych - [K2st_U10]4. potrafi dobrać algorytm grupowania odpowiedni do specyfiki zbioru danych i celów badawczych - [K2st_U10]5. potrafi - zgodnie zadaną specyfikacją, uwzględniając aspekty pozatechniczne - zaprojektować silnik szachowy wykorzystujący wiedzę odkrytą z danych uczących oraz zrealizować ten projekt używając właściwych metod i narzędzi - [K2st_U11]6. potrafi określić kierunki dalszego uczenia się i zrealizować proces samokształcenia - [K2st_U16]
Kompetencje społeczne:
<ol style="list-style-type: none">1. rozumie, że metody i narzędzia odkrywania wiedzy są z czasem udoskonalane i zastępowane przez nowsze rozwiązania - [K2st_K1]2. rozumie znaczenie wykorzystywania najnowszej wiedzy z zakresu odkrywania wiedzy w rozwiązywaniu problemów praktycznych - [K2st_K2]

Sposoby sprawdzenia efektów kształcenia

Efekty kształcenia przedstawione wyżej weryfikowane są w następujący sposób:

Ocena formująca:

- a) w zakresie wykładów:
 - na podstawie odpowiedzi na pytania dotyczące materiału omówionego na poprzednich wykładach,
- b) w zakresie laboratoriów / ćwiczeń:
 - na podstawie oceny bieżącego postępu realizacji zadań.

Ocena podsumowująca:

- a) w zakresie wykładów weryfikowanie założonych efektów kształcenia realizowane jest przez:
 - ocenę wiedzy i umiejętności wykazanych na kolokwium zaliczeniowym podczas ostatniego wykładu, w formie testu dydaktycznego z zadaniami wielokrotnego wyboru, zadaniami obliczeniowymi oraz z zadaniami wymagającymi uzupełnienia; test składa się z 30 pytań; maksymalna liczba punktów to 120 pkt., z czego na uzyskanie oceny 3.0 trzeba zdobyć min. 60 pkt.
 - omówienie wyników kolokwium,
- b) w zakresie laboratoriów weryfikowanie założonych efektów kształcenia realizowane jest przez:
 - ocenę umiejętności związanych z realizacją ćwiczeń laboratoryjnych oraz ewentualną 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,
 - ocenę projektu realizowanego w grupach dwuosobowych,
 - ocenę wiedzy i umiejętności związanych z realizacją zadań laboratoryjnych poprzez 1 względnie 2 kolokwia w semestrze (wg preferencji studentów).

Możliwe jest uzyskiwanie punktów dodatkowych za aktywność podczas zajęć, a szczególnie za:

- efektywność zastosowania zdobytej wiedzy podczas rozwiązywania zadanego problemu,
- uwagi związane z udoskonaleniem materiałów dydaktycznych,
- wskazywanie trudności percepcyjnych studentów umożliwiające bieżące doskonalenie procesu dydaktycznego.

Treści programowe

Program wykładu obejmuje następujące zagadnienia:

Wprowadzenie do odkrywania wiedzy - definicja odkrywania wiedzy w danych, kolejne etapy procesu odkrywania wiedzy, definicja eksploracji danych, reprezentacje odkrywanej wiedzy, zadania eksploracji danych, standard CRISP-DM, przykłady zastosowań eksploracji danych. Wybrane techniki przetwarzania danych - postać (schemat) danych, przekleństwo wymiarowości, ocena jakości danych, problem brakujących wartości w danych, metody identyfikacji punktów oddalonych, dyskretyzacja, indukcja konstruktywna i redukcja rozmiaru danych - metody selekcji atrybutów i ich przekształcania, redukcja rozmiaru zbioru danych za pomocą różnych technik próbkowania. Analiza skupień (grupowanie) - wybrane numeryczne i konceptualne metody tworzenia skupień, ocena jakości grupowania, miary odległości pomiędzy obiektami dla atrybutów różnego typu. Rozszerzenia algorytmu k-means (BFR, CURE, fuzzy c-means), algorytm k-medoids, metody grupowania: oparte na gęstości obserwacji (ang. density-based clustering), np. DBSCAN, OPTICS, oparte na modelu statystycznym (ang. distribution-based clustering) - EM clustering, wykorzystujące teorię zbiorów przybliżonych. Ograniczenia metod numerycznych analizy skupień. Algorytm indukcji reguł AQ. Algorytmy konceptualnej analizy skupień - CLUSTER/2, COBWEB. Indukcja regułowych modeli wiedzy, w szczególności omówienie algorytmu sekwencyjnego pokrywania VCDomLEM. Metody analizy danych porządkowych i nieporządkowych z brakującymi wartościami atrybutów. Indukcyjne uczenie się preferencji w problemie rankingu - sformułowanie problemu, pozyskiwanie informacji preferencyjnej z różnych źródeł, indukcja modelu preferencji, wybrane procedury eksploatacji wartościowanej relacji przewyższania i ich własności, przykłady zastosowań. Metody wnioskowania na podstawie podobieństwa do znanych przykładów (ang. case-based reasoning, CBR). Zbiory rozmyte, arytmetyka rozmyta, modele rozmyte - systemy Mamdani i Takagi-Sugeno. Przegląd nowoczesnych systemów odkrywania wiedzy, np. - WEKA, RapidMiner, KNIME, Orange. Programowanie inteligentnych silników szachowych - reprezentacje szachownicy, generowanie posunięć, przeszukiwanie drzewa gry, algorytmy oceny pozycji, uczenie się z przykładów, reprezentacja wiedzy.

Zajęcia laboratoryjne prowadzone są w formie piętnastu 2-godzinnych ćwiczeń, odbywających się w laboratorium komputerowym. Ćwiczenia realizowane są indywidualnie, z wykorzystaniem narzędzi MS Excel, WEKA, RapidMiner, KNIME, Statistica, Matlab / GNU Octave / Scilab, R, Orange, biblioteka obliczeniowa Java Rough Set (jRS) / biblioteka obliczeniowa ruleLearn, ruleRank/RUDE.

Program laboratorium obejmuje następujące zagadnienia:

Wybrane narzędzia do wstępnej analizy i wizualizacji danych. Wybrane techniki przetwarzania danych - próbkowanie, selekcja atrybutów, dyskretyzacja. Analiza skupień - wybrane algorytmy grupowania, numeryczne i konceptualne. Wykorzystanie biblioteki jRS/ruleLearn i aplikacji ruleRank/RUDE do uczenia się regułowych modeli danych porządkowych w problemie klasyfikacji i rankingu. Zastosowanie wybranych metod analizy danych do zbiorów z brakującymi wartościami. Tworzenie modeli rozmytych z użyciem pakietu Matlab. Programowanie własnych silników szachowych - mini-projekt w zespołach 2-osobowych. Odkrywanie wiedzy w rzeczywistych zbiorach danych z wykorzystaniem nowoczesnych systemów eksploracji danych. Techniki i narzędzia wizualizacji odkrytej wiedzy.

Metody dydaktyczne:

1. wykład: prezentacja multimedialna, prezentacja ilustrowana przykładami podawanymi na tablicy, dyskusja.
2. ćwiczenia laboratoryjne: słowne wprowadzenie, notatka elektroniczna, prezentacja multimedialna, zadania praktyczne typu krok po kroku (ang. tutorial), otwarte zadania praktyczne, wykonywanie eksperymentów.

Literatura podstawowa:

1. Eksploracja danych. D. Hand, H. Mannila, P. Smyth, WNT, Warszawa, 2005.
2. Odkrywanie wiedzy z danych. Wprowadzenie do eksploracji danych. Daniel T. Larose, PWN, Warszawa, 2006.
3. Statystyczne systemy uczące się. J. Koronacki, J. Ćwik, WNT, Warszawa, 2005.
4. Mining of massive datasets. J. Leskovec, A. Rajaraman, J.D. Ullman, Cambridge University Press, 2014, 2nd edition. (książka dostępna on-line: <http://www.mmnds.org/#ver21>)
5. Pattern Recognition and Machine Learning. C.M. Bishop, Springer 2006.

Literatura uzupełniająca:

1. Systemy uczące się. P. Cichosz, WNT, Warszawa, 2000.
2. The Elements of Statistical Learning. Data Mining, Inference, and Prediction. T. Hastie, R. Tibshirani, J. Friedman, Springer, New York, 2016, 2nd edition.
3. J. Błaszczyński, R. Słowiński, M. Szeląg, Sequential Covering Rule Induction Algorithm for Variable Consistency Rough Set Approaches. Information Sciences, 181, 2011, pp. 987-1002.
4. M. Szeląg, J. Błaszczyński, R. Słowiński, Rough Set Analysis of Classification Data with Missing Values. [In]: L. Polkowski et al. (Eds.): Rough Sets, International Joint Conference, IJCRS 2017, Olsztyn, Poland, July 3-7, 2017, Proceedings, Part I. Lecture Notes in Artificial Intelligence, vol. 10313, Springer, 2017, pp. 552-565.
5. M. Szeląg, S. Greco, R. Słowiński, Variable Consistency Dominance-Based Rough Set Approach to Preference Learning in Multicriteria Ranking. Information Sciences, 277, 2014, pp. 525-552.
6. M. Szeląg, S. Greco, R. Słowiński, Similarity-Based Classification with Dominance-Based Decision Rules. [In]: V. Flores et al. (Eds.): Rough Sets, International Joint Conference, IJCRS 2016, Santiago de Chile, Chile, October 7-11, 2016, Proceedings. Lecture Notes in Artificial Intelligence, vol. 9920, Springer, 2016, pp. 355-364.
7. <https://www.chessprogramming.org>.

Bilans nakładu pracy przeciętnego studenta

Czynność		Czas (godz.)
1. udział w zajęciach laboratoryjnych / ćwiczeniach		30
2. przygotowanie do ćwiczeń laboratoryjnych		7
3. realizacja mini-projektu		25
4. udział w konsultacjach związanych z realizacją procesu kształcenia		2
5. przygotowanie do 1-2 kolokwium pisemnych na laboratorium		12
6. udział w wykładach		30
7. zapoznanie się ze wskazaną literaturą / materiałami dydaktycznymi		6
8. przygotowanie do kolokwium zaliczeniowego z wykładu		12
9. omówienie wyników kolokwium zaliczeniowego z wykładu		1
Obciążenie pracą studenta		
forma aktywności	godzin	ECTS
Łączny nakład pracy	125	5
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	63	3
Zajęcia o charakterze praktycznym	62	2