

Karta (sylabus) modułu/przedmiotu
Robotyzacja procesów wytwórczych
Studia pierwszego stopnia

Przedmiot:	Techniki eksploracji danych
Rodzaj przedmiotu:	Obieralny
Kod przedmiotu:	RPW-1-S-0-6-MK54-2_0
Rok:	III
Semestr:	6
Forma studiów:	Studia stacjonarne
Rodzaj zajęć i liczba godzin w semestrze:	
Wykład:	30
Ćwiczenia:	
Laboratorium:	30
Projekt:	
Liczba punktów ECTS:	4
Sposób zaliczenia:	zaliczenie
Język wykładowy:	polski

Cel przedmiotu	
C1	Zdobycie wiedzy z zakresu podstaw technik eksploracji danych i wskazanie kontekstu zastosowań w naukach technicznych.
C2	Wykształcenie umiejętności posługiwania się aparatem pojęciowym technik stosowanych w analizie wielowymiarowych zbiorów danych oraz nabycia praktyki w opracowaniu i prezentacji wyników obserwacji.

Wymagania wstępne w zakresie wiedzy, umiejętności i innych kompetencji	
1	Podstawy statystyki matematycznej, analizy matematycznej i algebry liniowej.
2	Znajomość języka programowania i obsługi środowiska obliczeń numerycznych MATLAB

Efekty kształcenia	
	W zakresie wiedzy:
EK1	Zna metodykę przygotowania, metody analizy i wizualizacji wyników eksploracji danych
EK2	Zna cel i kontekst zastosowania technik eksploracji danych
	W zakresie umiejętności:
EK3	Potrafi dobrać metody selekcji, analizy i wizualizacji danych w kontekście rozwiązywanego problemu technicznego
EK4	Potrafi posługiwać się oprogramowaniem specjalistycznym wykorzystywanym w technikach eksploracji danych

Treści programowe przedmiotu	
Forma zajęć: wykłady	
	Treści programowe:
W1	Wprowadzenie do problematyki i zastosowań eksploracji danych (ang. <i>data mining</i>). Podstawowe pojęcia stosowane w technikach eksploracji danych. Zasadnicze etapy, metody i klasy algorytmów eksploracji danych. System eksploracji danych.
W2	Metody organizacji i reprezentacji danych. Macierze; statystyki opisowe. Graficzna reprezentacja danych; grafy wzrostu; grafy Chernoff'a; Metryki zbiorów wielowymiarowych. Repozytoria i bazy danych wykorzystywane w eksploracji danych.
W3	Wprowadzenie do statystyk wielowymiarowych. Wielowymiarowy rozkład normalny i jego właściwości. Statystyki z próby wielowymiarowej. Znaczenie macierzy wariancyjno-kowariancyjnej. Wielowymiarowe testy statystyczne wektora średnich; statystyka Hotellinga T^2 . Wielowymiarowe przedziały ufności.
W4	Etap pozyskiwania/przygotowania/czyszczenia danych, selekcji danych. Wielowymiarowe karty kontrolne; detekcja obserwacji odstających; metody generowania cech diagnostycznych; metody selekcji cech diagnostycznych;
W5	Redukcja wymiaru danych/liczby zmiennych. Metoda składowych głównych PCA. Geometria i model składowych głównych. Standaryzacja danych i interpretacja struktury macierzy wariancyjno-kowariancyjnej. Zastosowania metody składowych głównych.
W6	Metody eksploracji danych; odkrywanie/modelowanie związków przyczynowo-skutkowych; reguły asocjacyjne; odkrywanie charakterystyk danych; regresja liniowa i logistyczna; klasyfikatory Bayesa; sieci neuronowe; analiza skupień; dyskryminacja i klasyfikacja; drzewa decyzyjne;
W7	Ocena jakości modelu danych. Metody oceny jakości klasyfikacji i predykcji. Charakterystyka ROC. Zróżnicowanie kosztu błędnej klasyfikacji. Aspekty wdrożenia systemu eksploracji danych.
W8	Przykłady zastosowań technik eksploracji danych; Metody wizualizacji wyników eksploracji danych.
Forma zajęć: laboratoria	
	Treści programowe:
L1	Wprowadzenie do środowiska numerycznego wykorzystywanego w cyklu zajęć laboratoryjnych;
L2	Statystyki wielowymiarowe i metody prezentacji zbiorów wielowymiarowych - przykłady i ćwiczenia ilustrujące zagadnienia opisu zbiorów wielowymiarowych;
L3	Metody reprezentacji i sposoby pozyskiwania danych z zasobów informatycznych. Systemy baz danych i interfejsy programistyczne dedykowane technikom eksploracji danych; Przykłady i zadania ilustrujące problemy akwizycji danych;
L4	Selekcja i przygotowanie danych do analizy i eksploracji informacji; Zastosowania metod statystycznych; Metody generowania cech diagnostycznych - przykłady i zadania;
L5	Zastosowania metody składowych głównych w redukcji zmiennych i konstrukcji modelu regresji - przykłady i zadania;
L6	Zadania i problemy ilustrujące wybrane metody eksploracji danych: reguły asocjacyjne; odkrywanie charakterystyk danych; regresja liniowa i logistyczna; klasyfikatory Bayesa; sieci neuronowe; analiza skupień; dyskryminacja i klasyfikacja; drzewa decyzyjne;

Metody dydaktyczne	
1	Wykład z prezentacją multimedialną.
2	Zajęcia laboratoryjne wykorzystujące oprogramowanie specjalistyczne.

Obciążenie pracą studenta	
<i>Forma aktywności</i>	<i>Średnia liczba godzin na zrealizowanie aktywności</i>
Godziny kontaktowe z wykładowcą:	60
W tym: Udział w wykładach:	30
Udział w ćwiczeniach:	
Udział w zajęciach laboratoryjnych:	30
Udział w zajęciach projektowych:	
Praca własna studenta:	40
W tym: Samodzielne studiowanie tematyki wykładów, przygotowanie i udział w kolokwium zaliczającym wykład:	20
Przygotowanie do ćwiczeń rachunkowych	
Przygotowanie do zajęć laboratoryjnych, opracowanie sprawozdań:	20
Przygotowanie projektu:	
Łączny czas pracy studenta:	100
Sumaryczna liczba punktów ECTS dla przedmiotu:	4
Liczba punktów ECTS w ramach zajęć o charakterze praktycznym (ćwiczenia, laboratoria, projekty):	2

Literatura podstawowa	
1	Osowski S.: Metody i narzędzia eksploracji danych, Wydawnictwo BTC, 2013
2	Morzy T.: Eksploracja danych: metody i algorytmy, Wydawnictwo PWN, 2013

Literatura uzupełniająca	
1	Hand D., Mannila H., Smyth P.: Eksploracja danych, Wydawnictwo PWN, 2005

Macierz efektów kształcenia					
Efekt kształcenia	Odniesienie danego efektu kształcenia do efektów zdefiniowanych dla całego programu (PEK)	Cele przedmiotu	Treści programowe	Metody dydaktyczne	Metody oceny
EK1	RPW1A_W03++ RPW1A_W07+	C1	W1-W8	1,2	O1
EK2	RPW1A_W08++	C1	W1-W8	1,2	O1
EK3	RPW1A_U10+	C2	L1-L6	2	O2
EK4	RPW1A_U10+	C2	L1-L6	2	O2

Metody i kryteria oceny		
Symbol metody oceny	Opis metody oceny	Próg zaliczeniowy
O1	Zaliczenie treści wykładu	51%
O2	Zaliczenie laboratorium	51%

Autor programu:	dr Marcin Bogucki
Adres e-mail:	m.bogucki@pollub.pl
Jednostka organizacyjna:	Katedra Automatykacji