

**AGH**AGH UNIVERSITY OF SCIENCE
AND TECHNOLOGY

Nazwa modułu zajęć: Metody statystyczne w analizie danych i modelowaniu systemów

Rok akademicki: 2019/2020 Kod: ZSDA-3-0004-s Punkty ECTS: 6

Wydział: Szkoła Doktorska AGH

Kierunek: Szkoła Doktorska AGH Specjalność: —

Poziom studiów: Studia III stopnia Forma studiów: Stacjonarne

Język wykładowy: Polski Profil: Ogólnoakademicki (A) Semestr: 0

Strona www: <http://home.agh.edu.pl/~kulpi/>Prowadzący moduł: prof. dr hab. inż. Kulczycki Piotr (kulczycki@agh.edu.pl)**Treści programowe zapewniające uzyskanie efektów uczenia się dla modułu zajęć**

Tematyka przedmiotu obejmuje współczesne metody pozyskiwania wiedzy z danych. Szczegółowo rozważane są procedury wykrywania elementów nietypowych (odosobnionych), grupowania (klasteryzacji) i klasyfikacji. W powyższym zakresie omawiana jest jednolita metodyka oparta na statystycznej estymacji nieparametrycznej, użyta w dalszej części również do zagadnienia modelowania złożonych systemów, traktowanego jako proces wspomaganie decyzji.

Opis efektów uczenia się dla modułu zajęć

Kod MEU	Student, który zaliczył moduł zajęć zna i rozumie/potrafi/jest gotów do	Powiązania z KEU	Sposób weryfikacji i oceny efektów uczenia się osiągniętych przez studenta w ramach poszczególnych form zajęć i dla całego modułu zajęć
Wiedza: zna i rozumie			
M_W001	Poszerzenie wiedzy z zakresu statystyki matematycznej.	SDA3A_W03, SDA3A_W02, SDA3A_W01	Aktywność na zajęciach
M_W002	Istota i uwarunkowania pozyskiwania wiedzy z danych.	SDA3A_W02, SDA3A_W01	Aktywność na zajęciach
M_W003	Znajomość procedur analizy i eksploracji danych.	SDA3A_W03, SDA3A_W02, SDA3A_W01	Aktywność na zajęciach
Umiejętności: potrafi			
M_U001	Stosowanie procedur analizy i eksploracji danych.	SDA3A_U02, SDA3A_U01	Kolokwium

M_U002	Wykorzystanie procedur eksploracji danych do praktycznych zagadnień pozyskiwania wiedzy z danych.	SDA3A_U02, SDA3A_U01	Kolokwium
M_U003	Pozyskiwanie informacji ze źródeł różnego typu oraz prezentacja zagadnienia badawczego i uzyskanych wyników.	SDA3A_U03, SDA3A_U02, SDA3A_U01, SDA3A_U04	Kolokwium
Kompetencje społeczne: jest gotów do			
M_K001	Praca zespołowa.	SDA3A_K01, SDA3A_K03	Aktywność na zajęciach
M_K002	Profesjonalność i etyka.	SDA3A_K01, SDA3A_K03, SDA3A_K02	Aktywność na zajęciach
M_K003	Konieczność ustawicznego samokształcenia.	SDA3A_K01	Aktywność na zajęciach

Liczba godzin zajęć w ramach poszczególnych form zajęć

Suma	Forma zajęć dydaktycznych										
	Wykład	Ćwiczenia audytoryjne	Ćwiczenia laboratoryjne	Ćwiczenia projektowe	Konwersatorium	Zajęcia seminaryjne	Zajęcia praktyczne	Zajęcia terenowe	Zajęcia warsztatowe	Prace kontrolne i przejściowe	Lektorat
65	30	0	20	15	0	0	0	0	0	0	0

Matryca kierunkowych efektów uczenia się w odniesieniu do form zajęć i sposobu zaliczenia, które pozwalają na ich uzyskanie

Kod MEU	Student, który zaliczył moduł zajęć zna i rozumie/potrafi/jest gotów do	Forma zajęć dydaktycznych										
		Wykład	Ćwiczenia audytoryjne	Ćwiczenia laboratoryjne	Ćwiczenia projektowe	Konwersatorium	Zajęcia seminaryjne	Zajęcia praktyczne	Zajęcia terenowe	Zajęcia warsztatowe	Prace kontrolne i przejściowe	Lektorat
Wiedza: zna i rozumie												
M_W001	Poszerzenie wiedzy z zakresu statystyki matematycznej.	+	-	+	-	-	-	-	-	-	-	-
M_W002	Istota i uwarunkowania pozyskiwania wiedzy z danych.	+	-	-	+	-	-	-	-	-	-	-
M_W003	Znajomość procedur analizy i eksploracji danych.	+	-	+	+	-	-	-	-	-	-	-
Umiejętności: potrafi												
M_U001	Stosowanie procedur analizy i eksploracji danych.	+	-	+	+	-	-	-	-	-	-	-

M_U002	Wykorzystanie procedur eksploracji danych do praktycznych zagadnień pozyskiwania wiedzy z danych.	+	-	+	+	-	-	-	-	-	-	-
M_U003	Pozyskiwanie informacji ze źródeł różnego typu oraz prezentacja zagadnienia badawczego i uzyskanych wyników.	-	-	+	+	-	-	-	-	-	-	-
Kompetencje społeczne: jest gotów do												
M_K001	Praca zespołowa.	-	-	+	+	-	-	-	-	-	-	-
M_K002	Profesjonalność i etyka.	+	-	-	-	-	-	-	-	-	-	-
M_K003	Konieczność ustawicznego samokształcenia.	+	-	-	+	-	-	-	-	-	-	-

Nakład pracy studenta (bilans punktów ECTS)

Forma aktywności studenta	Obciążenie studenta
Udział w zajęciach dydaktycznych/praktyka	65 godz
Przygotowanie do zajęć	30 godz
przygotowanie projektu, prezentacji, pracy pisemnej, sprawozdania	30 godz
Samodzielne studiowanie tematyki zajęć	55 godz
Sumaryczne obciążenie pracą studenta	180 godz
Punkty ECTS za moduł	6 ECTS

Pozostałe informacje

Szczegółowe treści kształcenia w ramach poszczególnych form zajęć (szczegółowy program wykładów i pozostałych zajęć)

Wykład

Wprowadzenie

Przedmiot analizy i eksploracji danych; pozyskiwanie wiedzy z danych. Związek analizy danych ze statystyką matematyczną.

Pojęcia wstępne

Typy danych.

Miary podobieństwa.

Wstępne przetwarzanie danych: czyszczenie, obsługa brakujących atrybutów, standaryzacja/normalizacja.

Wykrywanie elementów odosobnionych (nietypowych)

Elementy oddalone.

Test istotności.

Grupowanie (klasteryzacja)

Algorytm k-najbliższych sąsiadów.

Metody hierarchiczne.
Procedury częstotliwościowe.

Klasyfikacja

Algorytm k-najbliższych sąsiadów.
Drzewa decyzyjne.
Ujęcie bayesowskie.

Metodyka oparta na statystycznej estymacji nieparametrycznej

Nieparametryczne metody estymacji statystycznej; estymatory jądrowe.
Jednolita metodyka w zakresie wykrywania elementów nietypowych, klasteryzacji i klasyfikacji.

Modelowanie matematyczne i wspomaganie decyzji

Modelowanie matematyczne systemów. Identyfikacja.
Wspomaganie decyzji.
Algorytm identyfikacji parametrycznej przy niesymetrycznej funkcji strat.

Zastosowania

Przykładowe aplikacje w zagadnieniach badań systemowych, inżynierii sterowania i marketingu.
Ewaluacja wyników.

Tematy opcjonalne

Redukcja wymiaru danych.
Duże (liczne) zbiory danych.
Strumienie danych.

Ćwiczenia laboratoryjne

Laboratorium

Zajęcia wprowadzające.
Wykrywanie elementów odosobnionych (nietypowych).
Grupowanie (klasteryzacja).
Klasyfikacja.
Estymatory jądrowe.
Procedury oparte na estymatorach jądrowych.
Identyfikacja parametryczna z niesymetryczną funkcją strat.
Procedury analizy danych w Internecie.
Temat opcjonalny: redukcja wymiaru danych.
Zajęcia zaliczeniowe.

Ćwiczenia projektowe

Projekt

Opracowanie zagadnień zgodnych z tematyką wykładu, według indywidualnych ustaleń.

Metody i techniki kształcenia:

Wykład: Treści prezentowane na wykładzie są przekazywane w formie prezentacji multimedialnej w połączeniu z klasycznym wykładem tablicowym, wzbogaconymi o pokazy odnoszące się do prezentowanych zagadnień.

Ćwiczenia laboratoryjne: W trakcie zajęć laboratoryjnych studenci samodzielnie rozwiązują zadany problem praktyczny, dobierając odpowiednie narzędzia informatyczne. Prowadzący stymuluje grupę do wnioskowania w zakresie badanego zagadnienia.

Ćwiczenia projektowe: Studenci wykonują zadany projekt samodzielnie, bez większej ingerencji prowadzącego. Końcowe wyniki są prezentowane i omawiane w szerszym gronie studentów.

Warunki i sposób zaliczenia poszczególnych form zajęć, w tym zasady zaliczeń poprawkowych, a także warunki dopuszczenia do egzaminu:

Laboratorium komputerowe: zaliczenie wszystkich kolokwiów na ocenę co najmniej 3,0. Ocena końcowa jest średnią z tych ocen.

Projekt: każda grupa projektowa powinna uzgodnić z prowadzącym temat i zakres prac nie później niż do końca 8 tygodnia zajęć, a następnie przedłożyć uzyskane wyniki badań w celu dopuszczenia do prezentacji. Zajęcia prezentacyjne odbędą się w uzgodnionych terminach po około 6 godzin i około 10 grup w poszczególnych terminach – każda grupa powinna przedstawić swoje wyniki na wybranym z nich i czynnie uczestniczyć w całych zajęciach tego terminu. Końcowa ocena obejmuje wartość merytryczną wykonanych badań oraz jakość prezentacji.

Warunek zaliczenia przedmiotu: zaliczenie projektu i laboratorium komputerowego.

Zasady udziału w poszczególnych zajęciach, ze wskazaniem, czy obecność studenta na zajęciach jest obowiązkowa:

Wykład:

- Obecność obowiązkowa: Nie

- Zasady udziału w zajęciach: Studenci uczestniczą w zajęciach poznając kolejne treści nauczania zgodnie z sylabusem przedmiotu. Uczestnicy winni na bieżąco zadawać pytania i wyjaśniać wątpliwości.

Ćwiczenia laboratoryjne:

- Obecność obowiązkowa: Tak

- Zasady udziału w zajęciach: Studenci wykonują ćwiczenia laboratoryjne zgodnie z materiałami udostępnionymi przez prowadzącego. Student jest zobowiązany do przygotowania się w przedmiocie wykonywanego ćwiczenia. Zaliczenie zajęć odbywa się na podstawie zaprezentowania rozwiązania sformułowanego problemu.

Ćwiczenia projektowe:

- Obecność obowiązkowa: Tak

- Zasady udziału w zajęciach: Studenci wykonują prace praktyczne mające na celu uzyskanie kompetencji zakładanych przez sylabus. Ocenie podlega sposób wykonania projektu oraz efekt końcowy przedłożony na zajęciach prezentacyjnych.

Sposób obliczania oceny końcowej

Ocena końcowa jest średnią oceny z laboratorium komputerowego i projektu.

Sposób i tryb wyrównywania zaległości powstałych wskutek nieobecności studenta na zajęciach:

Ćwiczenia laboratoryjne: nieobecności muszą być odrobione w formie i terminach uzgodnionych z prowadzącym. Połowa nieobecności (także usprawiedliwionych) skutkuje brakiem zaliczenia bez możliwości poprawek.

Projekt: konsultacje w godzinach podanych przez prowadzącego. Nie ma możliwości odrobienia zajęć prezentacyjnych poza uzgodnionymi terminami.

Wymagania wstępne i dodatkowe, z uwzględnieniem sekwencyjności modułów

Wiedza z zakresu matematyki na poziomie nabytym na studiach II stopnia, zwłaszcza z zakresu statystyki matematycznej.

Umiejętność użytkowania sprzętu komputerowego.

Zalecana literatura i pomoce naukowe

Kulczycki P., Estymatory jądrowe w analizie systemowej, WNT, 2005.

Krzyśko M., Wołyński W., Górecki T., Skorzybut M., Systemy uczące się; rozpoznawanie wzorców, analiza skupień i redukcja wymiarowości, WNT, 2008.

Morzy T., Eksploracja danych; Metody i algorytmy, PWN, 2013.

- Literatura uzupełniająca:

Larose D.T., Odkrywanie wiedzy z danych; wprowadzenie do eksploracji danych, PWN, 2006.

Larose D.T., Odkrywanie wiedzy z danych; metody i modele eksploracji danych, PWN, 2008.

Kulczycki P., Hryniewicz O., Kacprzyk J. (red.), Techniki informacyjne w badaniach systemowych, WNT, 2007.

Kulczycki P., Korbicz J., Kacprzyk J. (red.), Automatyka, robotyka i przetwarzanie informacji, PWN, 2019 (w druku).

Publikacje naukowe osób prowadzących zajęcia związane z tematyką modułu

del>/del> prof. dr hab. inż. Piotr Kulczycki

- Książka:

1. P. Kulczycki, „Estymatory jądrowe w analizie systemowej”, WNT, Warszawa, 2005.

- Wybrane publikacje w czasopismach z listy JCR:

1. P. Kulczycki, M. Charytanowicz, P.A. Kowalski, S. Łukasik, „The Complete Gradient Clustering Algorithm: Properties in Practical Applications”, Journal of Applied Statistics, vol. 39, nr 6, ss. 1211-1224, 2012.

2. P. Kulczycki, M. Charytanowicz, „Conditional Parameter Identification with Different Losses of Under- and Overestimation”, Applied Mathematical Modelling, vol. 37, nr 4, ss. 2166-2177, 2013.

3. P. Kulczycki, S. Łukasik, „An Algorithm for Reducing Dimension and Size of Sample for Data Exploration Procedures”, International Journal of Applied Mathematics and Computer Science, vol. 24, nr 1, ss. 133-149, 2014.

4. P. Kulczycki, P.A. Kowalski, „Bayes Classification for Nonstationary Patterns”, International Journal of Computational Methods, vol. 12, nr 2, ID 1550008 (19 stron), 2015.

5. P. Kulczycki, D. Kruszewski, „Identification of Atypical Elements by Transforming Task to Supervised Form with Fuzzy and Intuitionistic Fuzzy Evaluations”, Applied Soft Computing, vol. 60, nr 11, ss. 623-633, 2017.

del>/del> dr inż. Szymon Łukasik

- Książka:

1. S. Łukasik, „Algorytm redukcji wymiaru i liczności próby dla celów procedur eksploracyjnej analizy danych”, WPK, Kraków, 2013.

- Wybrane publikacje w czasopismach z listy JCR:

1. P. Kulczycki, M. Charytanowicz, P.A. Kowalski, S. Łukasik, „The Complete Gradient Clustering Algorithm: Properties in Practical Applications”, Journal of Applied Statistics, vol. 39, nr 6, ss. 1211-1224, 2012.

2. P. Kulczycki, S. Łukasik, „An Algorithm for Reducing Dimension and Size of Sample for Data Exploration Procedures”, International Journal of Applied Mathematics and Computer Science, vol. 24, nr 1, ss. 133-149, 2014.

3. P.A. Kowalski, S. Łukasik, „Training Neural Networks with Krill Herd Algorithm”, Neural Processing Letters, vol. 44, nr 1, ss. 5-17, 2016.

4. D. Domańska, S. Łukasik, „Handling high-dimensional data in air pollution forecasting tasks”, Ecological Informatics, vol. 34, nr 4, ss. 70-91, 2016.

5. A. Mora, T. Santos, S. Łukasik, J. Silva, A.J. Falcão, J.M. Fonseca, R.A. Ribeiro, „Land Cover Classification from Multispectral Data Using Computational Intelligence Tools: A Comparative Study”, Information, vol. 8, nr 4, paper 1. 147, 2017.

Informacje dodatkowe

Przedmiot może być zaliczony na podstawie analogicznych zajęć na innych studiach, jeżeli ich łączny wymiar godzinowy jest nie mniejszy niż 65 godziny, a ostateczna ocena nie mniejsza niż 4,0. Ostateczna decyzja należy do osoby prowadzącej moduł.