

**KARTA MODUŁU / KARTA PRZEDMIOTU**

Kod modułu	<b>Z-ID-504</b>
Nazwa modułu	<b>Eksploracja i przygotowanie danych do analiz</b>
Nazwa modułu w języku angielskim	<b>Data Exploration and Preparation to Analyses</b>
Obowiązuje od roku akademickiego	<b>2015/2016</b>

**A. USYTUOWANIE MODUŁU W SYSTEMIE STUDIÓW**

Kierunek studiów	<b>Inżynieria danych</b>
Poziom kształcenia	<b>I stopień</b>
Profil studiów	<b>Praktyczny</b>
Forma i tryb prowadzenia studiów	<b>Studia stacjonarne</b>
Specjalność	<b>Wszystkie specjalności</b>
Jednostka prowadząca moduł	<b>Katedra Informatyki i Matematyki Stosowanej</b>
Koordinator modułu	<b>Dr hab. Marzena Nowakowska</b>
Zatwierdził	

**B. OGÓLNA CHARAKTERYSTYKA PRZEDMIOTU**

Przynależność do grupy/bloku przedmiotów	<b>Kierunkowy</b>
Status modułu	<b>Obowiązkowy</b>
Język prowadzenia zajęć	<b>Polski</b>
Usytuowanie modułu w planie studiów - semestr	<b>Semestr V</b>
Usytuowanie realizacji przedmiotu w roku akademickim	<b>Semestr zimowy</b>
Wymagania wstępne	<b>Algebra liniowa, Elementy rachunku prawdopodobieństwa i statystyki, Bazy danych, Logika, Algorytmy i struktury danych, Akwizycja danych pomiarowych, Bezpieczeństwo danych, Zarządzanie bazami danych - SQL</b>
Egzamin (TAK/NIE)	<b>TAK</b>
Liczba punktów ECTS	<b>4</b>

<b>Forma prowadzenia zajęć</b>	<b>wykład w</b>	<b>ćwiczenia ć</b>	<b>laboratorium l</b>	<b>projekt p</b>	<b>inne i</b>
<b>Liczba godzin w semestrze</b>	<b>15</b>		<b>30</b>		

### C. EFEKTY KSZTAŁCENIA I METODY SPRAWDZANIA EFEKTÓW KSZTAŁCENIA

Cel modułu	Nabycie przez studenta umiejętności przetwarzania danych w celu ich przygotowania do zaawansowanych analiz. Wykształcenie świadomości ważności tego procesu pracy z danymi jako fundamentalnego etapu pośredniego między etapem rejestracji i zarządzania danymi a etapem analizy związków w danych. Wskazanie możliwości wykorzystania programów: Access, Excel, SAS® (w tym pakietu Enterprise Miner™ oraz modułu Enterprise Guide™) w realizacji zadań eksploracji i przygotowania danych do analiz.
------------	---

Symbol efektu	Efekty kształcenia	Forma prowadzenia zajęć	Odniesienie do efektów kierunkowych	Odniesienie do efektów obszarowych
W_01	Student wie jak poszukiwać i pozyskiwać dane potrzebne do analiz.	w, l	K_W05, K_W12	T1P_W03 T1P_W04 T1P_W06 T1P_W07
W_02	Student ma świadomość znaczenia jakości danych. Ma podstawową wiedzę w zakresie diagnozowania jakości danych oraz narzędzi wykorzystywanych do czyszczenia danych.	w, l	K_W05, K_W09, K_W12	T1P_W02 T1P_W03 T1P_W04 T1P_W06 T1P_W07 X1P_W04
W_03	Student ma wiedzę dotyczącą przygotowania danych do analiz poprzez ich ewidencję i wstępną eksplorację.	w, l	K_W05, K_W13	T1P_W03 T1P_W04 T1P_W06 X1P_W04
W_04	Student wie jak i jakie narzędzia programistyczne stosować do eksploracji i przygotowania danych do analiz.	w, l	K_W12, K_W13	T1P_W03 T1P_W04 T1P_W06 X1P_W04
U_01	Student potrafi pozyskać dane do analiz oraz scharakteryzować ich strukturę.	w, l	K_U04	T1P_U15
U_02	Student potrafi wykonać podstawową diagnozę jakości danych i wykonać podstawowe operacje czyszczenia tych danych.	l	K_U06, K_U16	T1P_U05 T1P_U08 T1P_U09 T1P_U16
U_03	Student potrafi wykonać eksplorację danych ilościowych i jakościowych, wykorzystując właściwe oprogramowanie.	l	K_U06, K_U14	T1P_U01 T1P_U08 T1P_U09 T1P_U13
U_04	Student potrafi zidentyfikować podstawowe problemy w danych oraz zaproponować rozwiązanie tych problemów, również z wykorzystaniem właściwego oprogramowania.	l	K_U14, K_U16	T1P_U01 T1P_U09 T1P_U15
U_05	Student umie opracować wyniki swojej pracy w formie dokumentacji lub raportu.	l	K_U02	T1P_U02 T1P_U03 X1P_U05 X1P_U08
K_01	Student ma świadomość odpowiedzialności za pracę własną oraz ponoszenia odpowiedzialności za wspólnie realizowane zadania projektowe.	w, l	K_K04	T1P_K03 T1P_K04 X1P_K02
K_02	Student umie komunikować się w zespole interdyscyplinarnym w zakresie wykraczającym poza zagadnienia czysto techniczne.	w, l	K_K05	T1P_K03 X1P_K02

## Treści kształcenia

### 1. Treści kształcenia w zakresie wykładu

Nr wykładu	Treści kształcenia	Odniesienie do efektów kształcenia dla modułu
1	Źródła danych i metody pozyskiwania danych do analiz (dane oportunistyczne, eksperymentalne). Diagnozowanie struktur danych. Klasyfikacja danych. Znaczenie jakości danych.	W_01, U_01
2	Niekompletność i niepoprawność danych. Możliwości diagnozowania jakości danych i usuwania w nich usterek – zagadnienie czyszczenia danych. Weryfikacja danych w oparciu o źródła wewnętrzne i zewnętrzne. Pozyskiwanie i tworzenie zasobów zewnętrznych do weryfikacji danych.	W_01, W_02, U_02
3	Spójność wewnętrzna i integralność referencyjna w bazach danych. Metody i narzędzia do weryfikacji i naprawy danych numerycznych (w tym danych typu data/godzina). Metody i narzędzia do weryfikacji i naprawy danych tekstowych.	W_02, U_02
4	Eksploracja danych jakościwnych (rozkłady, ocena zmiennych jakościwnych, badanie niezależności zmiennych jakościwnych). Eksploracja danych kwantytatywnych (miary statystyczne, histogramy, analiza korelacji).	W_03, W_04, U_03
5	Problemy z danymi: duże zbiory danych, liczne atrybuty, liczne rekordy, niekompletność danych, nierównomierny rozkład atrybutów. Metody rozwiązywania problemów z danymi.	W_03, W_04, U_03, U_04
6	Narzędzia do rozwiązywania problemów z danymi w systemie SAS. Wykorzystanie węzłów wspomagających przetwarzanie danych w środowisku pakietu Enterprise Miner.	W_03, W_04, U_04
7	Przekształcanie danych dla celów analiz ( <i>data manipulation</i> ): <ul style="list-style-type: none"> <li>dane ilościowe: odrzucanie wartości ekstremalnych, normalizacja, dyskretyzacja (kategoryzacja) atrybutów ciągłych,</li> <li>dane jakościowe: wartości rzadkie, agregacja wartości cechy, agregacja cech.</li> </ul>	W_03, W_04, U_03, U_04
8	Egzamin.	

### 2. Treści kształcenia w zakresie ćwiczeń

### 3. Treści kształcenia w zakresie zadań laboratoryjnych

Nr zajęć lab.	Treści kształcenia	Odniesienie do efektów kształcenia dla modułu
1	Pozyskanie danych do analiz. Charakterystyka źródła danych i sposobu ich pozyskania, opis struktury pozyskanego zbioru danych i charakterystyka jego zawartości, klasyfikacja danych w zbiorze.	W_01, W_02, U_01
2	<u>Opracowanie raportu</u> . Sprawozdanie z wykonania prac z ćwiczeń nr 1. Wnioski i zalecenia.	U_05, K_01, K_02
3	Identyfikacja danych niekompletnych, błędów w danych i możliwych źródeł ich powstania. Wskazanie zakresu poprawy danych. Wybrane metody naprawy danych. Przygotowanie źródeł zewnętrznych do wsparcia procedur czyszczenia.	W_01, W_02, U_02, U_03
4	Opracowanie i zastosowanie narzędzi do poprawy danych w oparciu o źródła zewnętrzne i dane wewnętrzne: weryfikacja i naprawa danych numerycznych (w tym danych typu data/godzina), weryfikacja i naprawa danych tekstowych. Statystyka weryfikacji i naprawy danych.	W_02, U_01, U_02, U_03
5	<u>Opracowanie raportu</u> . Sprawozdanie z wykonania prac z ćwiczeń nr 3 i 4. Wnioski i zalecenia.	U_05, K_01, K_02
6	Eksploracja danych jakościwnych z wykorzystaniem programu Excel i modułu Enterprise Guide (EG) systemu SAS: zakresy wartości i rozkłady zmiennych jakościwnych, ocena zmiennych jakościwnych, badanie niezależności zmiennych jakościwnych.	W_03, W_04, U_03

7	Eksploatacja danych kwantytatywnych z wykorzystaniem programu Excel i modułu Enterprise Guide (EG) systemu SAS: miary statystyczne zmiennych ilościowych, rozkłady zmiennych ilościowych, analiza korelacji między zmiennymi ilościowymi.	W_03, W_04, U_03
8	<u>Opracowanie raportu.</u> Sprawozdanie z wykonania prac z ćwiczeń nr 6 i 7. Wnioski.	U_05, K_01, K_02
9-10	Rozwiązywanie problemów z danymi w systemie SAS. Węzły wspomagające przetwarzanie danych w środowisku pakietu Enterprise Miner (EM). Budowanie diagramu przepływu informacji w obszarze roboczym EM do przygotowania danych do analiz.	W_03, W_04, U_03, U_04
11	<u>Opracowanie raportu.</u> Sprawozdanie z wykonania prac z ćwiczeń nr 9 i 10. Wnioski.	U_05, K_01, K_02
12	Przekształcanie danych ilościowych w środowisku systemu SAS w celu ich przygotowania do analiz (narzędzia i metody przetwarzania danych): wartości ekstremalne, normalizacji, dyskretyzacja atrybutów ciągłych.	W_03, W_04, U_03, U_04
13	Przekształcanie danych jakościowych w środowisku systemu SAS w celu ich przygotowania do analiz (narzędzia i metody przetwarzania danych): wartości rzadkie, agregacja wartości zmiennych jakościowych, agregacja zmiennych jakościowych.	W_03, W_04, U_03, U_04
14	<u>Opracowanie raportu.</u> Sprawozdanie z wykonania prac z ćwiczeń nr 12 i 13. Wnioski.	U_05, K_01, K_02
15	Kolokwium podsumowujące.	

4. Charakterystyka zadań projektowych

5. Charakterystyka zadań w ramach innych typów zajęć dydaktycznych

### Metody sprawdzania efektów kształcenia

<b>Symbol efektu</b>	<b>Metody sprawdzania efektów kształcenia</b> (sposób sprawdzenia, w tym dla umiejętności – odwołanie do konkretnych zadań projektowych, laboratoryjnych, itp.)
W_01	Zadania do samodzielnego wykonania na zajęciach.
W_02	Zadania do samodzielnego wykonania na zajęciach. Egzamin.
W_03	Zadania do samodzielnego wykonania na zajęciach. Egzamin.
W_04	Zadania do samodzielnego wykonania na zajęciach. Egzamin.
U_01	Zadania do samodzielnego wykonania na zajęciach. Praca przy komputerze oraz opracowanie raportu.
U_02	Zadania do samodzielnego wykonania na zajęciach. Praca przy komputerze oraz opracowanie raportu.
U_03	Zadania do samodzielnego wykonania na zajęciach. Praca przy komputerze oraz opracowanie raportu.
U_04	Zadania do samodzielnego wykonania na zajęciach. Praca przy komputerze oraz opracowanie raportu.
U_05	Opracowanie raportów z prac oraz dyskusja nad ich strukturą logiczną i treściami merytorycznymi.
K_01	Komentarze i dyskusja opracowań studenckich na ćwiczeniach laboratoryjnych i wykładach.
K_02	Komentarze i dyskusja opracowań studenckich na ćwiczeniach laboratoryjnych i wykładach.

**D. NAKŁAD PRACY STUDENTA**

<b>Bilans punktów ECTS</b>			
<b>Lp.</b>	<b>Rodzaj aktywności</b>	<b>Obciążenie studenta</b>	<b>Jednostka</b>
1.	Udział w wykładach	15	h
2.	Udział w ćwiczeniach		
3.	Udział w laboratoriach	30	h
4.	Udział w zajęciach projektowych		
5.	Udział w konsultacjach (2-3 razy w semestrze)	4	h
6.	Konsultacje projektowe		
7.	Udział w egzaminie	2	h
8.			
9.	<b>Liczba godzin realizowanych przy bezpośrednim udziale nauczyciela akademickiego</b>	51	h
10.	<b>Liczba punktów ECTS, którą student uzyskuje na zajęciach wymagających bezpośredniego udziału nauczyciela akademickiego</b> <i>(1 punkt ECTS=27 godzin obciążenia studenta)</i>	1,9	ECTS
11.	Samodzielne studiowanie tematyki wykładów	18	h
12.	Samodzielne przygotowanie się do ćwiczeń		
13.	Samodzielne przygotowanie się do kolokwium		
14.	Samodzielne przygotowanie się do laboratoriów	10	h
15.	Wykonanie sprawozdań	20	h
16.	Przygotowanie do kolokwium końcowego z laboratorium		
17.	Wykonanie projektu lub dokumentacji		
18.	Przygotowanie do egzaminu	9	h
19.			
20.	<b>Liczba godzin samodzielnej pracy studenta</b>	57	h
21.	<b>Liczba punktów ECTS, którą student uzyskuje w ramach samodzielnej pracy</b> <i>(1 punkt ECTS=27 godzin obciążenia studenta)</i>	2,4	ECTS
22.	<b>Sumaryczne obciążenie pracą studenta</b>	108	h
23.	<b>Punkty ECTS za moduł</b> <i>1 punkt ECTS=27 godzin obciążenia studenta</i>	4	ECTS
24.	<b>Nakład pracy związany z zajęciami o charakterze praktycznym</b> <i>Suma godzin związanych z zajęciami praktycznymi</i>	64	h
25.	<b>Liczba punktów ECTS, którą student uzyskuje w ramach zajęć o charakterze praktycznym</b> <i>1 punkt ECTS=27 godzin obciążenia studenta</i>	2,4	ECTS

**E. LITERATURA**

Wykaz literatury	<ol style="list-style-type: none"> <li>1. Cichosz P., <i>Systemy uczące się</i>, Wydawnictwa Naukowo-Techniczne, Warszawa 2000.</li> <li>2. Larose D.T., <i>Metody i modele eksploracji danych</i>, Wydawnictwo Naukowe PWN, Warszawa 2008.</li> <li>3. Hand D., Manila H., Smyth P., <i>Eksploracja danych</i>, Wydawnictwa Naukowo-Techniczne, Warszawa 2005.</li> <li>4. Olhost F.J., <i>Big Data Analytics. Turning Big Data into Big Money</i>, John Wiley &amp; Sons, Inc., Hoboken, New Jersey 2013.</li> <li>5. Panek T., <i>Statystyczne metody wielowymiarowej analizy porównawczej</i>, Szkoła Główna Handlowa w Warszawie, Warszawa 2009.</li> <li>6. Dokumentacja programu SAS dostępna on-line.</li> </ol>
Witryna WWW modułu/przedmiotu	