

**KARTA MODUŁU / KARTA PRZEDMIOTU**

Kod modułu	<b>Z-ID-608a</b>
Nazwa modułu	<b>Analiza danych niestukturalnych</b>
Nazwa modułu w języku angielskim	<b>Unstructured Data Analysis</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>Analityka danych i modelowanie</b>
Jednostka prowadząca moduł	<b>Katedra Informatyki i Matematyki Stosowanej</b>
Koordinator modułu	<b>Mgr inż. Karolina Karwowska Dr hab. Marzena Nowakowska</b>
Zatwierdził	

**B. OGÓLNA CHARAKTERYSTYKA PRZEDMIOTU**

Przynależność do grupy/bloku przedmiotów	<b>Specjalnościowy</b>
Status modułu	<b>Obowiązkowy</b>
Język prowadzenia zajęć	<b>Polski</b>
Usytuowanie modułu w planie studiów - semestr	<b>Semestr VI</b>
Usytuowanie realizacji przedmiotu w roku akademickim	<b>Semestr letni</b>
Wymagania wstępne	<b>Wprowadzenie do programowania w środowisku analitycznym, Odkrywanie związków w danych wielowymiarowych</b>
Egzamin (TAK/NIE)	<b>NIE</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>10</b>		<b>30</b>		

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

<b>Cel modułu</b>	Zwiększenie świadomości studentów na temat użyteczności dużych ilości danych tekstowych. Zapoznanie z technikami eksploracji tekstu. Nabycie umiejętności posługiwania się programem SAS® Text Miner oraz testowanie niekomercyjnych narzędzi do eksploracji tekstu.
-------------------	--

Symbol efektu	Efekty kształcenia	Forma prowadzenia zajęć	Odniesienie do efektów kierunkowych	Odniesienie do efektów obszarowych
W_01	Student zna podstawowe metody <i>Text Mining</i> , służące do strukturyzacji informacji tekstowych i automatycznej analizy dokumentów tekstowych oraz zna ich zastosowanie.	w, l	K_W05	T1P_W02 T1P_W03 T1P_W06
W_02	Student zna narzędzie SAS Text Miner oraz wybrane darmowe narzędzia do eksploracji danych.	w, l	K_W13	T1P_W03 T1P_W04 T1P_W06 X1P_W02 X1P_W04
W_03	Student ma wiedzę na temat poszczególnych metod statystycznych przydatnych w analizie danych niestrukturalnych oraz zna ich przykładowe zastosowania prowadzące do znalezienia nieznanymi zależności, wzorców i trendów pomiędzy danymi.	w, l	K_W02 K_W09	T1P_W01 T1P_W02 T1P_W04 T1P_W06 X1P_W02 X1P_W03 X1P_W04
W_04	Student rozumie konieczność przekształcania dokumentów tekstowych oraz zna różne rodzaje reprezentacji tekstu.	w, l	K_W13	T1P_W03 T1P_W04 T1P_W06 X1P_W02 X1P_W04
U_01	Student potrafi wykonać proste zadania konwersji dokumentów tekstowych do wektorów cech.	w, l	K_U05	T1P_U01 T1P_U08 T1P_U09 T1P_U12 T1P_U16
U_02	Student potrafi dokonać klasyfikacji zbioru dokumentów tekstowych z wykorzystaniem odpowiednich algorytmów i narzędzi.	l	K_U06 K_U16	T1P_U02 T1P_U05 T1P_U08 T1P_U09 T1P_U13 T1P_U16 T1P_U19
U_03	Student potrafi dokonać grupowania zbioru dokumentów tekstowych za pomocą odpowiednich algorytmów i narzędzi.	l	K_U06 K_U16	T1P_U02 T1P_U05 T1P_U08 T1P_U09 T1P_U13 T1P_U16 T1P_U19
U_04	Student potrafi samodzielnie pozyskiwać odpowiednie dane tekstowe do analizy.	l	K_U01	T1P_U01 T1P_U03 T1P_U06 X1P_U07
K_01	Student rozumie potrzebę ciągłego poszerzania wiedzy z obszaru <i>Data Mining</i> i <i>Text Mining</i> .	w, l	K_K01	T1P_K01 X1P_K01 X1P_K05

## Treści kształcenia

### 1. Treści kształcenia w zakresie wykładu

Nr wykładu	Treści kształcenia	Odniesienie do efektów kształcenia dla modułu
1	Wprowadzenie do metod analizy danych niestrukturalnych. Techniki <i>Data Mining</i> , <i>Text Mining</i> , <i>Web Mining</i> i ich zastosowanie. Wprowadzenie do SAS Text Miner.	W_01 W_02
2	Wstępna analiza danych tekstowych. Metody doskonalenia reprezentacji dokumentów (tokenizacja, stemming, stopwords).	W_03
3	Modele reprezentacji tekstu (macierz TFM, TF-IDF, miary odległości dla reprezentacji wektorowej). Transformacja danych tekstowych (redukcja wymiarów macierzy częstości).	W_04
4	Metody eksploracji danych – klasyfikacja tekstu. Naiwny klasyfikator bayesowski, algorytm k najbliższych sąsiadów.	W_03
5	Metody eksploracji danych – grupowanie dokumentów. Hierarchiczne i niehierarchiczne metody grupowania. Grupowanie dokumentów tekstowych za pomocą algorytmów aglomeracyjnych i podziałowych oraz algorytmu k-średnich. Zapoznanie ze środowiskiem Rapid Miner.	W_03

### 2. Treści kształcenia w zakresie ćwiczeń

### 3. Treści kształcenia w zadaniach laboratoryjnych

Nr zajęć lab.	Treści kształcenia	Odniesienie do efektów kształcenia dla modułu
1	Identyfikacja źródeł oraz pozyskiwanie danych tekstowych do analiz. Zapoznanie ze środowiskiem SAS Text Miner. Wczytywanie danych tekstowych w różnych formatach i przetwarzanie zbioru dokumentów w jedną bazę danych SAS – input dla Text Miner.	W_01
2	Wstępne przetwarzanie pozyskanych danych tekstowych w środowisku SAS. Eliminacja nieistotnych wyrazów (stop word) oraz redukcja do rdzenia (stemming) w celu doskonalenia reprezentacji danych tekstowych.	W_04 U_01
3	Wektorowa reprezentacja danych tekstowych w środowisku SAS. Metoda N – gram. Macierz term-frequency. Ocena ważności słów w macierzy – przekształcenia częstości występowania słów (podejście TF-IDF).	W_04 U_01
4	<u>Opracowanie raportu.</u> Sprawozdanie z wykonania prac z ćwiczeń nr 1-3. Porównanie wpływu testowanych metod przetwarzania dokumentów na jakość uzyskiwanych wyników. Wnioski i zalecenia.	W_02 K_01
5	Analiza dokumentów tekstowych w środowisku SAS – klasyfikacja dokumentów dotyczących pozyskanych opinii o produkcie/przedsiębiorstwie. Przygotowanie zbioru danych uczących, wykorzystanie algorytmu naiwnego klasyfikatora bayesowskiego, ocena jakości modelu.	W_02 U_02
6	Analiza dokumentów tekstowych w środowisku SAS – klasyfikacja dokumentów dotyczących pozyskanych opinii o produkcie/przedsiębiorstwie. Przygotowanie zbioru danych uczących, wykorzystanie algorytmu k najbliższych sąsiadów, ocena jakości modelu.	W_02 U_02
7	<u>Opracowanie raportu.</u> Sprawozdanie z wykonania prac z ćwiczeń nr 5-6. Porównanie testowanych algorytmów klasyfikacji. Wnioski i zalecenia.	W_02 K_01
8	Analiza dokumentów tekstowych w środowisku SAS – grupowanie dokumentów dotyczących pozyskanych opinii o produkcie/przedsiębiorstwie. Ekstrakcja cech dokumentów, wybór miary odległości. Grupowanie z wykorzystaniem algorytmów hierarchicznych.	W_02 U_03

9	Analiza dokumentów tekstowych w środowisku SAS – podział dokumentów dotyczących pozyskanych opinii o produkcie/przedsiębiorstwie. Podział z wykorzystaniem algorytmu k-średnich. Wpływ parametrów konfiguracyjnych procesu wektoryzacji dokumentów tekstowych na jakość klasteryzacji.	W_02 U_03
10	<u>Opracowanie raportu.</u> Sprawozdanie z wykonania prac z ćwiczeń nr 8-9. Porównanie testowanych algorytmów grupowania. Wnioski i zalecenia.	W_02 K_01
11	Zapoznanie ze środowiskiem Rapid Miner – wstępne przetwarzanie danych.	W_02 W_03
12	Analiza dokumentów tekstowych w środowisku Rapid Miner – klasyfikacja danych za pomocą wybranych algorytmów.	W_02 U_02
13	Realizacja projektu indywidualnego – wykonanie modelu analizy danych tekstowych. Określenie celu analizy, zebranie i przygotowanie zbioru danych tekstowych dotyczących wybranego zagadnienia. Import plików do środowiska SAS jako baza danych SAS.	W_02 U_04 K_01
14	Realizacja projektu indywidualnego – sporządzenie podstawowego modelu Text Mining, ocena wstępnych rezultatów, zastosowanie wybranej metody klasteryzacji. <u>Opracowanie raportu.</u> Sprawozdanie z wykonania prac z ćwiczeń nr 12-13. Wnioski i zalecenia.	W_02 K_01
15	Kolokwium podsumowujące.	K_01

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> <i>(sposób sprawdzenia, w tym dla umiejętności – odwołanie do konkretnych zadań projektowych, laboratoryjnych, itp.)</i>
W01	Kolokwium końcowe.
W02	Kolokwium końcowe.
W03	Kolokwium końcowe.
W04	Kolokwium końcowe.
U01	Sprawozdanie na ocenę w formie raportu z wykonanego zadania.
U02	Sprawozdanie na ocenę w formie raportu z wykonanego zadania.
U03	Sprawozdanie na ocenę w formie raportu z wykonanego zadania.
U04	Sprawozdanie na ocenę w formie raportu z wykonanego zadania.
K01	Aktywność i dyskusja na zajęciach.

## D. NAKŁAD PRACY STUDENTA

Bilans punktów ECTS			
Lp.	Rodzaj aktywności	Obciążenie studenta	Jednostka
1.	Udział w wykładach	10	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)	5	h
6.	Konsultacje projektowe		
7.	Udział w egzaminie		
8.			
9.	<b>Liczba godzin realizowanych przy bezpośrednim udziale nauczyciela akademickiego</b>	45	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>	1,7	ECTS
11.	Samodzielne studiowanie tematyki wykładów	6	h
12.	Samodzielne przygotowanie się do ćwiczeń		
13.	Samodzielne przygotowanie się do kolokwium		
14.	Samodzielne przygotowanie się do laboratoriów	6	h
15.	Wykonanie sprawozdań	20	h
16.	Przygotowanie do kolokwium końcowego z laboratorium		
17.	Wykonanie projektu lub dokumentacji	30	h
18.	Przygotowanie do egzaminu		
19.			
20.	<b>Liczba godzin samodzielnej pracy studenta</b>	62	h
21.	<b>Liczba punktów ECTS, którą student uzyskuje w ramach samodzielnej pracy</b> <i>(1 punkt ECTS=27 godzin obciążenia studenta)</i>	2,3	ECTS
22.	<b>Sumaryczne obciążenie pracą studenta</b>	107	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>	91	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>	3,4	ECTS

## E. LITERATURA

Wykaz literatury	<ol style="list-style-type: none"><li>1. <i>Getting Started with SAS® Text Miner 12.1</i>, SAS Institute Inc, 2012.</li><li>2. Kantardzic M., <i>Data mining: concepts, models, methods, and algorithms</i>, Wiley-Interscience: IEEE Press, 2003.</li><li>3. Larose D. T., <i>Metody i modele eksploracji danych</i>, Wydawnictwo Naukowe PWN, 2012.</li><li>4. Weiss S. M., Indurkha N., Zhang T., Damerau F., <i>Text Mining: Predictive Methods for Analyzing Unstructured Information</i>, Springer Science and Business Media, 2005.</li></ol>
Witryna WWW modułu/przedmiotu	