Opis **zajęć (sylabus)**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Nazwa zajęć: | | Techniki eksploracji danych | | | | | | | | **ECTS** | **4** |
| Nazwa zajęć w j. angielskim: | | Data mining techniques | | | | | | | | | |
| Zajęcia dla kierunku studiów: | | **Informatyka i Ekonometria** | | | | | | | | | |
|  | |  | | | | | | | | | |
| Język wykładowy: | | polski | | | | Poziom studiów: | | studia II stopnia | | | |
| Forma studiów: | 🗷 stacjonarne  🞎 niestacjonarne | Status zajęć: | 🞎 podstawowe  🗷 kierunkowe | 🞎 obowiązkowe  🗷 do wyboru | | Numer semestru: ……3….. | | 🗷 semestr zimowy 🞎 semestr letni | | | |
|  |  | Rok akademicki, od którego obowiązuje opis (rocznik): | | | | 2019/2020 | Numer katalogowy: | **ZIM-IE-2S-03Z-23\_1**  **ZIM-IE-2S-03Z-23\_4** | | | |
|  | | | | | | | | | | | |
| Koordynator zajęć: | |  | | | | | | | | | |
| Prowadzący zajęcia: | |  | | | | | | | | | |
| Jednostka realizująca: | |  | | | | | | | | | |
| Jednostka zlecająca: | |  | | | | | | | | | |
| Założenia, cele i opis zajęć: | | Celem kursu jest przegląd najnowszych osiągnięć w dziedzinie przetwarzania ogromnych zbiorów danych oraz rozszerzenie wiedzy zdobytej podczas pierwszej części kursu „Analiza danych masowych”.  Jednym z celów tego przedmiotu jest nabycie umiejętności niezbędnych do rozwoju wysokiej jakości oprogramowania do własnych celów naukowych lub biznesowych za pomocą języka programowania jakim jest R-CRAN.  W ramach przedmiotu zostaną szczegółowo omówione takie techniki jak drzewa decyzyjne, sztuczne sieci neuronowe oraz zaawansowane techniki uczenia maszynowego (maszyny wektorów wspierających – SVM, zespoły klasyfikatorów i predyktorów). Tematyka wykładów:  * Zagadnienia wstępne, podstawowe pojęcia (3V, 4V), problemy analizy dużych wolumenów danych; * Przetwarzanie danych, odkrywanie niespójności, osobliwości oraz zależności w dużych repozytoriach danych metody redukcji wymiaru; * Wprowadzenie do problemów klasyfikacji, regresji oraz prognozowania w praktyce gospodarczej * Techniki eksploracji danych   - drzewa klasyfikacyjne – podstawowe algorytmy, C4.5, C5.0, ID.3, CART, uczenie drzew klasyfikacyjnych, kryteria podziału drzewa, porównywanie trafności modeli;  - sieci neuronowe – historia, budowa neuronu, budowa sztucznych sieci neuronowych typu RBF i MLP, wybrane algorytmy uczenia sztucznych sieci neuronowych, optymalizacja architektury sieci neuronowej, ocena jakości działania sieci;   * Zaawansowane techniki uczenia maszynowego – maszyny wektorów wspierających (SVM), zespoły klasyfikatorów i predyktorów; * Zastosowania metod eksploracji danych w informatyce i telekomunikacji, finansach, przemyśle, medycynie, marketingu;  Tematyka ćwiczeń laboratoryjnych: Część pierwsza:   * Środowisko R i program RStudio   + Cechy języka R   + Organizacja pracy w R i RStudio * Podstawowe typy atomowe: wektory i NULL   + Podstawowe i złożone typy danych w R   + Wektory atomowe   + Tworzenie obiektów nazwanych   + Braki danych, wartości nieskończone i nie-liczby   + Typ pusty (NULL) * Operacje na wektorach   + Podstawowe operatory   + Indeksowanie wektorów. Filtrowanie danych   + Przegląd funkcji wbudowanych * Listy   + Tworzenie list   + Indeksowanie list   + Wybrane operacje * Funkcje   + Tworzenie obiektów typu funkcja   + Parametry i argumenty * Modyfikacja przepływu sterowania   + Wyrażenia warunkowe if i if...else   + Pętle   + Uwagi na temat wydajności * Typy złożone   + Macierze i tablice   + Szeregi czasowe   + Czynniki   + Ramki danych * Przetwarzanie napisów   + Standardy kodowania znaków   + Podstawowe operacje na napisach   + Wyszukiwanie wzorca   + Data i czas   Część druga:   * Techniki eksploracji danych oraz implementacja w środowisku R-CRAN: drzewa decyzyjne i sieci neuronowe do klasyfikacji i regresji; * Zaawansowane techniki uczenia maszynowego: maszyny wektorów wspierających – SVM oraz zespoły klasyfikatorów i predyktorów); * Realizacja studium przypadku na dużych wolumenach danych (kilkanaście milionów obserwacji/kilka tysięcy zmiennych), do wyboru spośród wymienionych poniżej:  1. klasyfikacja rezygnacji klientów; 2. ocena ryzyka kredytowego; 3. rozpoznawanie spamu elektronicznego; 4. klasyfikacja w marketingu bezpośrednim;  * prognozowanie szeregów czasowych i ich automatyzacja dla kilkuset linii produktowych. | | | | | | | | | |
| Formy dydaktyczne, liczba godzin: | | 1. wykład; liczba godzin ...15...; 2. ćwiczenia laboratoryjne; liczba godzin ...30...; | | | | | | | | | |
| Metody dydaktyczne: | | Prezentacje, dyskusja, ćwiczenia praktyczne, eksperymenty numeryczne, studium przypadku | | | | | | | | | |
| Wymagania formalne  i założenia wstępne: | | Podstawowe umiejętności z przetwarzania danych, systemów informacyjnych, baz danych, statystyki i analizy danych. | | | | | | | | | |
| Efekty uczenia się: | | Wiedza:  1 – Posiada wiedzę dotyczą eksploracji dużych wolumenów danych zgromadzonych w bazach danych i systemach informatycznych, z uwzględnieniem aspektów ustawy o ochronie danych osobowych, ochrony przed przestępczością elektroniczną.  2 – Posiada pogłębioną wiedzę na temat podstawowych algorytmów eksploracji danych (drzew decyzyjnych, sieci neuronowych, algorytmów) oraz zaawansowanych (zespoły klasyfikatorów) wykorzystywanych do wspomagania decyzji w praktyce biznesowej (w informatyce, przemyśle, marketingu).  3 – Posiada pogłębioną wiedzę o zjawiskach i procesach w działalności biznesowej, z uwzględnieniem specyfiki gromadzenia, przetwarzania, wydobywania danych z systemów baz danych i hurtowni danych. | | | Umiejętności:  1 – Potrafi wykorzystać poznane metody analityczne (techniki eksploracji danych) do formułowania i rozwiązywania zadań na styku zastosowań informatyki i biznesu (w informatyce, przemyśle, medycynie, marketingu).  2 – Potrafi wdrażać i testować wybrane techniki analityczne w systemach informatycznych (m.in. generowanie list scoringowych, generowanie reguł decyzyjnych) używając właściwych metod i narzędzi informatycznych. | | | | Kompetencje:  1 – Rozumie potrzebę pogłębiania swojej wiedzy oraz śledzenia trendów rozwojowych w dziedzinie systemów informatycznych, w tym systemów przetwarzania, analizy i wnioskowania na podstawie dużych wolumenów danych. | | |
| Sposób weryfikacji efektów uczenia się: | | Sprawdzian pisemno-programistyczny. | | | | | | | | | |
| Forma dokumentacji osiągniętych efektów uczenia się: | | Sprawdzian ze umiejętności przetwarzania danych oraz znajomości technik eksploracji danych | | | | | | | | | |
| Elementy i wagi mające wpływ  na ocenę końcową: | | **Kolokwium x2 – 100% (2x50%)** | | | | | | | | | |
| Miejsce realizacji zajęć: | | Wykład – sala audytoryjna, ćwiczenia laboratoryjne – laboratorium komputerowe | | | | | | | | | |
| Literatura podstawowa i uzupełniająca:  Cukier K., Mayer-Schonberger V. (2014) Big data. Rewolucja, która zmieni nasze myślenie, pracę i życie. Wydawnictwo MT Biznes  Jared D. (2014) Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners, Wiley  Hand D., Mannila H., Smyth P. (2005) Eksploracja danych, Wydawnictwo WNT  Larose T.D. (2006) Odkrywanie wiedzy z danych, Wydawnictwo Naukowe PWN, Warszawa  Larose T.D. (2008) Metody i modele eksploracji danych, Wydawnictwo Naukowe PWN, Warszawa  Gągolewski M., Programowanie w języku R, Wydawnictwo Naukowe PWN, 2014 | | | | | | | | | | | |
| UWAGI  Warunek zaliczenia przedmiotu oznacza uzyskanie co najmniej 50% punktów z każdego kolokwium. | | | | | | | | | | | |

Wskaźniki ilościowe charakteryzujące moduł/przedmiot:

|  |  |
| --- | --- |
| Szacunkowa sumaryczna liczba godzin pracy studenta (kontaktowych i pracy własnej) niezbędna dla osiągnięcia zakładanych dla zajęć efektów uczenia się - na tej podstawie należy wypełnić pole ECTS: | **120 h** |
| Łączna liczba punktów ECTS, którą student uzyskuje na zajęciach wymagających bezpośredniego udziału nauczycieli akademickich lub innych osób prowadzących zajęcia: | **2,5 ECTS** |

Tabela zgodności kierunkowych efektów uczenia się z efektami przedmiotu:

|  |  |  |  |
| --- | --- | --- | --- |
| kategoria efektu | Efekty uczenia się dla zajęć: | Odniesienie do efektów dla programu studiów dla kierunku | Oddziaływanie zajęć na efekt kierunkowy\*) |
| Wiedza 1 | Posiada wiedzę dotyczą eksploracji dużych wolumenów danych zgromadzonych w bazach danych i systemach informatycznych, z uwzględnieniem aspektów ustawy o ochronie danych osobowych, ochrony przed przestępczością elektroniczną. | K\_W05 / P7S\_WK | 1 |
| Wiedza 2 | Posiada pogłębioną wiedzę na temat podstawowych algorytmów eksploracji danych (drzew decyzyjnych, sieci neuronowych, algorytmów) oraz zaawansowanych (zespoły klasyfikatorów) wykorzystywanych do wspomagania decyzji w praktyce biznesowej (w informatyce, przemyśle, marketingu). | K\_W10 / P7S\_WG | 2 |
| Wiedza 3 | Posiada pogłębioną wiedzę o zjawiskach i procesach w działalności biznesowej, z uwzględnieniem specyfiki gromadzenia, przetwarzania, wydobywania danych z systemów baz danych i hurtowni danych. | K\_W13 / P7S\_WK | 2 |
| Umiejętności 1 | Potrafi wykorzystać poznane metody analityczne (techniki eksploracji danych) do formułowania i rozwiązywania zadań na styku zastosowań informatyki i biznesu (w informatyce, przemyśle, medycynie, marketingu). | K\_U06 / P7S\_UW | 2 |
| Umiejętności 2 | Potrafi wdrażać i testować wybrane techniki analityczne w systemach informatycznych (m.in. generowanie list scoringowych, generowanie reguł decyzyjnych) używając właściwych metod i narzędzi informatycznych. | K\_U14 / P7S\_UW | 2 |
| Kompetencje 1 | Rozumie potrzebę pogłębiania swojej wiedzy oraz śledzenia trendów rozwojowych w dziedzinie systemów informatycznych, w tym systemów przetwarzania, analizy i wnioskowania na podstawie dużych wolumenów danych. | K\_K01/ P7S\_UU | 2 |

\*)

3 – zaawansowany i szczegółowy,

2 – znaczący,

1 – podstawowy,