**karta przedmiotu**

1. Podstawowe informacje o przedmiocie

|  |  |
| --- | --- |
| Nazwa przedmiotu | Drążenie danych |
| Rocznik studiów | 2021/2022 |
| Kolegium | Informatyki Stosowanej |
| Kierunek studiów | Informatyka |
| Poziom kształcenia | Studia pierwszego stopnia |
| Profil kształcenia | praktyczny |
| Specjalność | Inżynieria danych |
| Osoba odpowiedzialna | dr inż. Jacek Jakieła |

1. Wymagania wstępne (wynikające z następstwa przedmiotów)

|  |
| --- |
| Podstawy statystyki, Bazy danych, Wstęp do programowania, Programowanie |

1. Efekty uczenia się i sposób realizacji zajęć
   1. Cele przedmiotu

|  |  |
| --- | --- |
| C1 | Przekazanie wiedzy na temat metod, technik i narzędzi z obszarów eksploracji danych oraz uczenia maszynowego. |
| C2 | Ukształtowanie umiejętności w zakresie analizy eksploracyjnej danych, projektowania, implementacji i oceny modeli uczenia maszynowego oraz ich zastosowań w rozwiązywaniu praktycznych problemów występujących we współczesnych organizacjach. |

* 1. Przedmiotowe efekty uczenia się, z podziałem na wiedzę, umiejętności i kompetencje społeczne, wraz z odniesieniem do efektów uczenia się dla kierunku

|  |  |  |
| --- | --- | --- |
| Lp. | Opis przedmiotowych efektów uczenia się | Odniesienie do efektów uczenia się dla kierunku |
| Po zaliczeniu przedmiotu student w zakresie **wiedzy** | | |
| P\_W01 | Potrafi posługiwać się aparatem pojęciowym związanym z dziedziną eksploracji danych. Zna proces oraz dobre praktyki w zakresie tworzenia rozwiązań bazujących na algorytmach z obszaru nauki o danych. | K\_W13 |
| Po zaliczeniu przedmiotu student w zakresie **umiejętności** | | |
| P\_U01 | Potrafi poddać analizie zbiór danych (analiza w oparciu o wybrane miary statystyczne oraz wizualizację) oraz przygotować dane dla analiz z wykorzystaniem algorytmów drążenia danych. | K\_U01 |
| P\_U02 | Potrafi zastosować wybrane algorytmy (dobór oraz implementacja) dla zadanych sytuacji problemowych. | K\_U10, K\_U16 |
| P\_U03 | Potrafi ocenić utworzone modele według ustalonych kryteriów oraz dobranych metod i przeprowadzić działania doskonalące. | K\_U02, K\_U03 |
| Po zaliczeniu przedmiotu student w zakresie **KOMPETENCJI SPOŁECZNYCH** | | |
| P\_K01 | Udoskonalić umiejętność pracy grupowej oraz indywidualnej związanych z warunkami pracy zawodowej biorąc pod uwagę specyfikę oraz etykę zawodową. Umiejętność pracy indywidualnej oraz komunikacji w grupie w celu konsultacji podjętych kroków w rozwiązaniu problemów. | K\_K06 |

* 1. Formy zajęć dydaktycznych oraz wymiar godzin i punktów ECTS

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Studia stacjonarne (ST) | | | | | | | |
| W | K | Ćw | L | ZP | P | eL | ECTS |
| - | - | - | 30 | - | 20 | - | 5 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Studia niestacjonarne (NST) | | | | | | | |
| W | K | Ćw | L | ZP | P | eL | ECTS |
| - | - | - | 20 | - | 20 | - | 5 |

* 1. Metody realizacji zajęć dydaktycznych

|  |  |
| --- | --- |
| Formy zajęć | Metoda realizacji |
| Laboratorium | Laboratoria realizowane w wybranym środowisku programowym w oparciu o odpowiednio dobrane studia przypadków i zbiory danych. Rozwiązywanie praktycznych problemów z wykorzystaniem algorytmów z dziedziny nauki o danych, ze szczególnym uwzględnieniem metod, technik i narzędzi eksploracji danych. |
| Projekt | Projekt realizowany zgodnie z ustalonym scenariuszem dla określonej dziedziny problemu zatwierdzonej przez prowadzącego w oparciu o odpowiednio dobrane zbiory danych. |

* 1. Treści kształcenia (oddzielnie dla każdej formy zajęć)

Laboratorium

|  |  |
| --- | --- |
| Lp. | Treści kształcenia realizowane w ramach laboratorium |
| L1 | Wprowadzenie do dziedziny drążenia danych. Obszary zastosowań. Możliwości i ograniczenia. |
| L2 | Myślenie analityczne i zadawanie właściwych pytań. Pytania deskryptywne, predykcyjne i normatywne. Przyczynowość i niepewność. |
| L3 | Badania eksploracyjne. Miary położenia i miary rozproszenia, badanie rozkładu danych, korelacja, badanie dwóch lub więcej zmiennych, eksperymenty statystyczne i badanie istotności, rozkłady danych i prób, analiza zbioru danych z wykorzystaniem wizualizacji. |
| L4 | Przygotowywanie danych pod kątem analizy z wykorzystaniem określonych algorytmów. Czyszczenie i przekształcanie danych – brakujące dane, niejednoznaczne wartości, identyfikacja wartości odstających, kodowanie danych jakościowych, zasady tworzenia testowych i treningowych zbiorów danych. |
| L5 | Klasyfikacja podobnych elementów w zbiorze danych z wykorzystaniem metod grupujących podobne przypadki. |
| L6 | Klasyfikacja z wykorzystaniem naiwnego klasyfikatora bayesowskiego. |
| L7 | Drzewa decyzyjne. Tworzenie drzew na potrzeby klasyfikacji oraz regresji. Metody przycinania drzew. |
| L8 | Analiza głównych składowych. Analiza czynnikowa. |
| L9 | Elementy eksploracji tekstu. Typy danych jako ciągi znaków. Zastosowanie worka słów. Worek słów typu n-gram. Kwantyfikowanie tekstu. Podstawowe operacje statystyczne oraz wizualizacja wyników. |
| L10 | Ocena i doskonalenie modeli. Walidacja krzyżowa. Wskaźniki oceny: metryki klasyfikacji binarnej, metryki klasyfikacji wieloklasowej, metryki regresji. |

Projekt

|  |  |
| --- | --- |
| Lp. | Treści kształcenia realizowane w ramach projektu |
| P1 | Sformułowanie problemu do rozwiązania oraz celu inżynierii i analizy danych. |
| P2 | Analiza eksploracyjna zbioru danych oraz dobór algorytmów do rozwiązania problemu. |
| P3 | Przygotowanie zbioru danych dla algorytmów wybranych do implementacji. |
| P4 | Implementacja algorytmów. |
| P5 | Ocena rozwiązania. |
| P6 | Przygotowanie dokumentacji wraz z wnioskami i zaleceniami. |

* 1. Korelacja pomiędzy efektami uczenia się, celami przedmiotu, a treściami kształcenia

|  |  |  |
| --- | --- | --- |
| Efekt uczenia się | Cele przedmiotu | Treści kształcenia |
| P\_W01 | C1 | L1-L10, P1-P6 |
| P\_U01 | C2 | L3-L4, P1-P3 |
| P\_U02 | L5-L9, P4 |
| P\_U03 | P5-P6, L5-L10 |
| P\_K01 | C1, C1 | P1-P6 |

* 1. Metody weryfikacji efektów uczenia się

|  |  |  |
| --- | --- | --- |
| Efekt  uczenia się | Metoda oceny | Forma zajęć, w ramach której następuje weryfikacja efektu |
| P\_W01 | Pytania testowe | Laboratorium |
| P\_U01 | Ocena projektu i dokumentacji | Projekt |
| P\_U02 | Ocena rozwiązań problemów przygotowywanych w ramach poszczególnych laboratoriów | Laboratorium |
| P\_U03 | Ocena projektu | Projekt |
| P\_K01 | Ocena projektu | Projekt |

* 1. Kryteria oceny stopnia osiągnięcia efektów uczenia się

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Efekt uczenia się | Na ocenę 2  student nie potrafi | Na ocenę 3  student potrafi | Na ocenę 4  student potrafi | Na ocenę 5  student potrafi |
| P\_W01 | zdefiniować podstawowych pojęć z obszaru drążenia danych | zdefiniować podstawowe pojęcia z obszaru drążenia danych uzyskując co najmniej 51% punktów | zdefiniować podstawowe pojęcia z obszaru drążenia danych uzyskując co najmniej 71% punktów | zdefiniować podstawowe pojęcia z obszaru drążenia danych uzyskując co najmniej 91% punktów |
| P\_U01 | przeprowadzić analizy eksploracyjnej zbioru danych oraz przygotować danych pod kątem analiz | przeprowadzić analizę eksploracyjną zbioru danych oraz przygotować dane pod kątem analiz popełniając niewielką liczbę błędów | przeprowadzić analizę eksploracyjną zbioru danych oraz przygotować dane pod kątem analiz nie popełniając błędów | przeprowadzić analizę eksploracyjną zbioru danych oraz przygotować dane pod kątem analiz nie popełniając błędów i popierając swoje działania dobrymi praktykami wykorzystywanymi w tym zakresie |
| P\_U02 | dobrać oraz zaimplementować wybranych algorytmów drążenia danych w podstawowym zakresie | dobrać oraz zaimplementować wybrane algorytmy drążenia danych – wykonać poprawnie co najmniej 50% zadań | dobrać oraz zaimplementować wybrane algorytmy drążenia danych – wykonać poprawnie co najmniej 75% zadań | dobrać oraz zaimplementować wybrane algorytmy drążenia danych – wykonać poprawnie wszystkie zadania |
| P\_U03 | ocenić opracowanych algorytmów z wykorzystaniem podstawowych metod | ocenić opracowane algorytmy z wykorzystaniem odpowiednich metod w stopniu podstawowym | ocenić opracowane algorytmy z wykorzystaniem odpowiednich metod w stopniu dobrym | ocenić opracowane algorytmy z wykorzystaniem większości metod oceny w sposób dogłębny |
| P\_K01 | przedstawić zastosowanych metod i narzędzi | przedstawić zastosowane metody i narzędzia w stopniu podstawowym | przedstawić zastosowane metody i narzędzia w stopniu zaawansowanym | przedstawić zastosowane metody i narzędzia w stopniu zaawansowanym oraz sformułować wnioski |

* 1. Literatura

|  |
| --- |
| Literatura podstawowa |
| Boschetti A., Massaron L.: Python: podstawy nauki o danych. Helion, 2017 lub nowsze. |
| Bruce P., Bruce A., Gedeck P.: Statystyka praktyczna w data science. 50 kluczowych zagadnień w językach R i Python. Helion, 2021 lub nowsze. |

|  |
| --- |
| Literatura uzupełniająca |
| Natingga D.: Algorytmy Data Science. Siedmiodniowy przewodnik. Helion, 2019 lub nowsze. |
| Vaughan D.: Umiejętności analityczne w pracy z danymi i sztuczną inteligencją. Wykorzystywanie najnowszych technologii w rozwijaniu przedsiębiorstwa. Helion, 2021 lub nowsze. |

1. Nakład pracy studenta - bilans punktów ECTS

|  |  |  |
| --- | --- | --- |
| **Rodzaje aktywności** | **Obciążenie studenta** | |
| **studia ST** | **studia NST** |
| Udział w C/L (UB) | 30 | 20 |
| Konsultacje do C/L (UB) | 6 | 4 |
| Samodzielne przygotowanie się do C/L, w tym przygotowanie do zaliczenia | 39 | 51 |
| Udział w i konsultacje do PS/PN/eL (UB) | 20 | 20 |
| Samodzielne przygotowanie się do zaliczenia PS/PN/eL | 30 | 30 |
| **Sumaryczne obciążenie pracą studenta** | **125** | **125** |
| **Punkty ECTS za przedmiot** | **5** | **5** |
| **Punkty ECTS za zajęcia prowadzone z bezpośrednim udziałem nauczycieli i studentów (UB)** | **2** | **2** |
| **Punkty ECTS za zajęcia kształtujące umiejętności praktyczne (PZ)** | **5** | **5** |