



Karta przedmiotu  
Metody eksploracji i analizy danych

**1. Informacje podstawowe**

<p><b>Kierunek studiów</b> telekomunikacja i technologie internetu rzeczy</p> <p>Specjalność: sieci i systemy internetu rzeczy</p> <p><b>Jednostka zarządzająca kierunkiem studiów</b> Wydział Telekomunikacji, Informatyki i Elektrotechniki</p> <p><b>Poziom studiów</b> pierwszego stopnia (inż.)</p> <p><b>Profil studiów</b> Profil ogólnoakademicki</p> <p><b>Forma studiów</b> studia stacjonarne</p>	<p><b>Cykl kształcenia (nabór)</b> 2023/24</p> <p><b>Kod przedmiotu</b> 05TTIRSIRS.PI40.1404.23</p> <p><b>Języki wykładowe</b> polski</p> <p><b>Obligatoryjność</b> Obowiązkowy</p> <p><b>Blok zajęciowy</b> Przedmioty specjalnościowe</p>	
<p><b>Wymagania wstępne</b></p>	<p>Podstawowa wiedza z zakresu programowania oraz z zakresu algorytmów</p>	
<p><b>Przedmioty wprowadzające</b></p>	<p>Algorytmy i struktury danych, Programowanie 2</p>	
<p><b>Koordinator</b></p>	<p>Agata Giełczyk</p>	
<p><b>Okres</b> Semestr 7</p>	<p><b>Forma zaliczenia</b> Zaliczenie na ocenę</p> <p><b>Forma prowadzenia i godziny zajęć</b> Wykład: 30 Ćwiczenia laboratoryjne: 45</p>	<p><b>Liczba punktów ECTS</b> 6.0</p>

**2. Efekty uczenia się dla przedmiotu**

Kod	Opis efektów uczenia się	Odniesienie do kierunkowych efektów uczenia się	Odniesienie do charakterystyk PRK
<b>Wiedza:</b>			
W1	Ma wiedzę na temat narzędzi i środowisk wytwarzania oprogramowania, w szczególności narzędzi Jupyter Notebook oraz Jupyter Lab	TTIR_O1_K_W07	P6S_WG P6S_WG_inż
W2	Posiada wiedzę o zasadach działania złożonych systemów przetwarzania i wymiany informacji	TTIR_O1_K_W12	P6S_WG P6S_WG_inż
<b>Umiejętności:</b>			
U1	Posiada umiejętności wymagane do obsługi centrum przetwarzania danych lub oprogramowania systemu rozproszonego	TTIR_O1_K_U12	P6S_UW P6S_UW_inż
U2	Potrafi posługiwać się odpowiednimi narzędziami: analizy, ewaluacji i walidacji do oceny systemów, technologii i zadań informatycznych	TTIR_O1_K_U14	P6S_UW P6S_UW_inż
<b>Kompetencje społeczne:</b>			
K1	Potrafi myśleć i działać w sposób przedsiębiorczy uwzględniając aspekt społeczny, naukowy lub etyczny	TTIR_O1_K_K02	P6S_KO

### 3. Treści programowe

Lp.	Treści programowe	Formy zajęć	Efekty uczenia się dla przedmiotu
1.	<p>Wykład: 1. Wprowadzenie – charakterystyka zbiorów danych pozyskiwanych w systemach IoT. Cele analizy danych: klasyfikacja, regresja, optymalizacja, predykcja. Podstawowe gałęzie sztucznej inteligencji wykorzystywane do analizy danych: systemy ekspertowe, uczenie maszynowe. 2. Pozyskiwanie danych - źródła danych, bazy danych 3. Ocena przydatności danych – dane źródłowe, rozkład częstości zmiennych, graficzna prezentacja danych, korelacja, reprezentatywność danych, duplikaty, szeregi czasowe 4. Metody wstępnego przetwarzania danych – uzupełnianie danych, poprawianie błędnych danych, zmienne numeryczne, zmienne katagoryczne, wyrażenia języka naturalnego, przetwarzanie danych w szeregach czasowych 5. Metody wzbogacania danych – równoważenie danych, zmienne wyliczeniowe, zastąpienie zmiennych wspólnym rozkładem prawdopodobieństwa, wydzielanie danych testowych 6. Klasyfikacja – klasyfikacja poprzez indukcję drzew decyzyjnych, klasyfikacja z użyciem maszyny wektorów nośnych, klasyfikacja probabilistyczna, klasyfikatory binarne a klasyfikacja wieloklasowa 7. Regresja – model regresji wielorakiej, zmienne katagoryczne w modelach regresji, regresja kwantylowa, regresja poprzez indukcję drzew decyzyjnych, sztuczne sieci neuronowe w zastosowaniach regresji 8. Grupowanie – algorytmy grupowania, grupowanie w celu znajdowania podobnych obiektów, grupowanie w celu kompresji, znajdowanie anomalii 9. Rekomendowanie – systemy rekomendujące, odkrywanie asocjacji, model Matchbox Recommender 10. Prognozowanie – szeregi czasowe w prognozowaniu, naiwne metody prognozowania, modele średniej ważniej, modele ARIMA, modele nieliniowe 11. Ocena i poprawa jakości modeli – reguła powrotu do średniej, kryteria oceny modeli eksploracji danych, ocena jakości modeli klasyfikacyjnych, ocena jakości modeli regresyjnych, ocena jakości modeli grupujących, ocena jakości modeli prognozujących, porównanie jakości modeli, poprawa jakości modeli 12. Publikacja modeli eksploracji danych jako usług WWW 13. Etyka przetwarzania danych – uprzedzenia w zbiorach danych, “wyjaśnialność” systemów uczenia maszynowego, ochrona danych</p> <p>Laboratorium: Ćwiczenia laboratoryjne będą polegać na implementacji algorytmów analizy danych i testowaniu ich. Kolejne laboratoria będą obejmować poszczególne algorytmy klasyfikacji, regresji, grupowania, rekomendowania oraz prognozowania. Laboratorium obejmować będzie również metody organizacji danych i wstępnego przetwarzania, np. za pomocą biblioteki numpy oraz pandas, a także wizualizacji danych za pomocą bibliotek matplotlib i seaborn. W trakcie ćwiczeń laboratoryjnych uczestnicy zapoznają się praktycznie z oceną jakości modeli oraz ich poprawą. Wśród zadań do wykonania będzie również opublikowanie modelu jako usługi www. W trakcie ćwiczeń studenci będą wykorzystywać narzędzie Jupyter Notebook oraz Jupyter Lab.</p>	Wykład, Ćwiczenia laboratoryjne	W1, W2, U1, U2, K1

#### 4. Metody prowadzenia zajęć, weryfikacji efektów uczenia się i warunki zaliczenia

Forma zajęć		
Wykład	<b>Metody prowadzenia zajęć:</b>	
	Wykład	
	<b>Metody (sposoby) weryfikacji:</b>	<b>Udział:</b>
	Zaliczenie ustne	100%
	<b>Warunki zaliczenia przedmiotu:</b>	
Uzyskanie 51% punktów podczas zaliczenia ustnego		
Ćwiczenia laboratoryjne	<b>Metody prowadzenia zajęć:</b>	
	Ćwiczenia laboratoryjne	
	<b>Metody (sposoby) weryfikacji:</b>	<b>Udział:</b>
	Sprawozdanie	100%
	<b>Warunki zaliczenia przedmiotu:</b>	
Średnia arytmetyczna ocen cząstkowych ze sprawozdań		

Efekt uczenia się dla przedmiotu	Metody (sposoby) weryfikacji	
	Zaliczenie ustne	Sprawozdanie
W1	x	
W2	x	
U1		x
U2		x
K1		x

#### 5. Literatura

##### Literatura podstawowa

1. Marcin Szeliga „Data science i uczenie maszynowe”, PWN 2017
2. Joel Grus „Data science od podstaw : analiza danych w Pythonie”, Helion 2020
3. Wes McKinney „Python w analizie danych. Przetwarzanie danych za pomocą pakietów Pandas i NumPy oraz środowiska IPython.” Wydanie II, Wydawnictwo Helion 2019

##### Literatura uzupełniająca

1. Francois Chollet “Deep learning. Praca z językiem Python i biblioteką Keras”, Helion 2019
2. Aurelien Geron “Uczenie maszynowe z użyciem Scikit-Learn i TensorFlow”, Helion 2020.

#### 6. Nakład pracy studenta - bilans godzin i punktów ECTS

Aktywność studenta		Obciążenie studenta Liczba godzin
Zajęcia prowadzone z bezpośrednim udziałem nauczyciela akademickiego lub innych osób prowadzących zajęcia	Wykład	30
	Ćwiczenia laboratoryjne	45
Praca własna studenta	Konsultacje	5
	Przygotowanie do zajęć	20
	Studiowanie literatury	30
	Przygotowanie do zaliczenia	20
<b>Łączny nakład pracy studenta</b>		150
<b>Liczba punktów ECTS</b>		6

\* Godzina (dydaktyczna) oznacza 45 minut