



## Wprowadzenie do Data Science

### Karta opisu przedmiotu

#### Informacje podstawowe

<b>Kierunek studiów</b> Geodezja i Kartografia	<b>Cykl dydaktyczny</b> 2021/2022	
<b>Specjalność</b> -	<b>Kod przedmiotu</b> DGIKS.li40K.641d8dec97fea.21	
<b>Jednostka organizacyjna</b> Wydział Geodezji Górniczej i Inżynierii Środowiska	<b>Języki wykładowe</b> polski	
<b>Poziom kształcenia</b> Studia inżynierskie I stopnia	<b>Obligatoryjność</b> Do wyboru	
<b>Forma studiów</b> Stacjonarne	<b>Blok zajęciowy</b> Przedmioty kierunkowe	
<b>Profil studiów</b> Ogólnoakademicki	<b>Przedmiot powiązany z badaniami naukowymi</b> Tak	
<b>Koordynator przedmiotu</b>	Janusz Rusek, Karol Firek	
<b>Prowadzący zajęcia</b>	Janusz Rusek, Karol Firek	
<b>Okres</b> Semestr 7	<b>Forma zaliczenia</b> Zaliczenie	<b>Liczba punktów ECTS</b> 3
	<b>Forma prowadzenia i godziny zajęć</b> Wykład: 15 Ćwiczenia projektowe: 30	

#### Cele kształcenia dla przedmiotu

C1	Zapoznanie studentów z tematyką data science
C2	Uświadomienie słuchaczom problemów związanych z zagadnieniami przeszukiwania danych oraz tworzenia modeli prognostycznych uczenia maszynowego
C3	Przekazanie wiedzy z zakresu budowy modeli klasyfikacyjnych i regresyjnych z zastosowaniem narzędzi ML

## Efekty uczenia się dla przedmiotu

Kod	Efekty w zakresie	Kierunkowe efekty uczenia się	Metody weryfikacji
<b>Wiedzy - Student zna i rozumie:</b>			
W1	- formułowanie funkcji celu wykorzystywanej w procesie uczenia sztucznych sieci neuronowych (ANN) i systemów pokrewnych, - metody regularyzacji i stabilizacji procesu uczenia, - metody optymalizacji gradientowej i bezgradientowych na przykładzie Pattern Search oraz algorytmów genetycznych i metod pokrewnych (PSO itp.), - konstruowanie funkcji celu dla metody Support Vector Machine (SVM) w problemie klasyfikacji i regresji, - metody oceny efektywności uczenia pod względem generalizacji, - metody oceny efektywności uczenia pod względem generalizacji.	GIK1A_W05, GIK1A_W06	Kolokwium
W2	- metody oceny jakości danych do dalszego przetwarzania, - metody Data-Mining wykorzystywane do redukcji wymiarowości danych i wyodrębniania istotnych cech (selekcja cech), - metody wstępnej analizy zależności liniowych między zmiennymi. Podstawowe pojęcia z obszaru modeli klasyfikacyjnych i regresyjnych.	GIK1A_W05, GIK1A_W06	Kolokwium
<b>Umiejętności - Student potrafi:</b>			
U1	- tworzyć modele regresyjne i klasyfikacyjne w oparciu o dostępne dane z wykorzystaniem środowiska programistycznego Python i platformy Matlab, - wykorzystywać modele do indywidualnych celów symulacyjnych i integracji z innymi częściami kodu, - oceniać istotność wpływu poszczególnych zmiennych wejściowych na końcową sieć neuronową lub system SVM.	GIK1A_U01, GIK1A_U07, GIK1A_U12	Wykonanie projektu
U2	- tworzyć problemy klasyfikacyjne i regresyjne na podstawie danych, - wyodrębniać zbiory treningowe, walidacyjne i testowe dla sztucznych sieci neuronowych i modeli SVM, - pisać algorytmy optymalizacyjne wykorzystywane do budowy sieci neuronowych.	GIK1A_U01, GIK1A_U07, GIK1A_U12	Wykonanie projektu
<b>Kompetencje społecznych - Student jest gotów do:</b>			
K1	pracować w zespole podczas ćwiczeń projektowych i wymieniać się informacjami.	GIK1A_K01, GIK1A_K03	Wykonanie projektu

### Treści programowe zapewniające uzyskanie efektów uczenia się dla modułu zajęć

Problemy klasyfikacji i regresji oraz redukcji wymiarowości w zbiorach danych przy użyciu metod Data-Mining i Machine Learning.

### Nakład pracy studenta

Rodzaje zajęć studenta	Średnia liczba godzin* przeznaczonych na zrealizowane aktywności
Wykład	15

Ćwiczenia projektowe	30
Dodatkowe godziny kontaktowe	5
Egzamin lub kolokwium zaliczeniowe	2
Samodzielne studiowanie tematyki zajęć	23
<b>Łączny nakład pracy studenta</b>	<b>Liczba godzin</b> 75
<b>Liczba godzin kontaktowych</b>	<b>Liczba godzin</b> 45

\* godzina (lekcyjna) oznacza 45 minut

## Treści programowe

Lp.	Treści programowe	Efekty uczenia się dla przedmiotu	Formy prowadzenia zajęć
1.	Kurs omawia metody z zakresu eksploracji danych i uczenia maszynowego. Przedstawione zostaną matematyczne podstawy tych metod oraz przykłady ich zastosowania. Omówione zostaną problemy klasyfikacji i regresji w podejściu liniowym i nieliniowym. Przedstawione zostaną metody redukcji rozmiaru pierwotnie zebranych danych. Zaprezentowane zostaną przykłady zastosowania tych metod w obszarze informacji o obiektach budowlanych inżynierii lądowej i wodnej.	W1, W2	Wykład
2.	Wykonanie ćwiczeń projektowych, w ramach których, dla dostępnego zbioru danych o charakterystykach technicznych obiektów budowlanych i oddziaływaniach eksploatacyjnych, sformułowane zostaną problemy klasyfikacyjne i regresyjne. Wykorzystane zostaną do tego metody sztucznych sieci neuronowych oraz metoda SVM. Gotowe modele zostaną przetestowane pod kątem poprawności i uogólnienia. Przetestowane zostaną różne metody optymalizacji uczenia sztucznych sieci neuronowych. W metodzie SVM analizowany będzie dobór parametrów wynikający z oryginalnej postaci funkcji celu.	U1, U2, K1	Ćwiczenia projektowe

## Informacje rozszerzone

### Metody i techniki kształcenia:

Wykład, Uczenie się zespołowe (Team based learning ), Praca grupowa, Dyskusja

Rodzaj zajęć	Metody zaliczenia	Warunki zaliczenia przedmiotu
Wykład	Kolokwium	
Ćwiczenia projektowe	Wykonanie projektu, Kolokwium	

## **Warunki i sposób zaliczenia poszczególnych form zajęć, w tym zasady zaliczeń poprawkowych, a także warunki dopuszczenia do egzaminu**

attendance at classes: obligatory (2 unjustified absences) attendance at lectures: indicated (optional)

### **Sposób obliczania oceny końcowej**

Degree of the test (lecture) X 0.5 + Degree of the project X 0.25 + Degree of the test (classes) X 0.25

### **Sposób i tryb wyrównywania zaległości powstałych wskutek nieobecności studenta na zajęciach**

in the student's own scope

## **Wymagania wstępne i dodatkowe**

- podstawowa wiedza z zakresu analizy matematycznej, algebry i probabilistyki,
- umiejętność działania na wektorach i macierzach.

### **Zasady udziału w poszczególnych zajęciach, ze wskazaniem, czy obecność studenta na zajęciach jest obowiązkowa**

Wykłady: Nie określono, ponieważ wykłady nie są obowiązkowe

Zajęcia projektowe: Obecność i aktywne uczestnictwo w wykonywaniu poszczególnych partii programistycznych uwzględnionych w treści sylabusu

## **Literatura**

### **Obowiązkowa**

1. OSOWSKI, S. Sieci neuronowe do przetwarzania informacji, Oficyna Wydawnicza Pol. Warszawskiej, Warszawa, 2006

### **Dodatkowa**

1. ŁĘSKI, Jacek. Systemy neuronowo-rozmyte. Wydawnictwa Naukowo-Techniczne, 2008.

## **Badania i publikacje**

### **Publikacje**

1. RUSEK, Janusz. Modelowanie stopnia zużycia technicznego budynków na terenach górniczych z wykorzystaniem wybranych metod sztucznej inteligencji. Wydawnictwa AGH, 2013.
2. RUSEK, Janusz. Creating a model of technical wear of building in mining area, with utilization of regressive SVM approach. Archives of Mining Sciences, 2009, 54.3: 455-466.
3. FIREK, K.; RUSEK, Janusz; WODYŃSKI, Aleksander. Wybrane metody eksploracji danych i uczenia maszynowego w analizie stanu uszkodzeń oraz zużycia technicznego zabudowy terenów górniczych. Przegląd Górniczy, 2016, 72.1: 50—55.

## Kierunkowe efekty uczenia się

Kod	Treść
GIK1A_K01	krytycznej oceny posiadanej wiedzy i stałego samokształcenia i samorozwoju zawodowego
GIK1A_K03	wypełniania zobowiązań społecznych i działalności na rzecz społeczeństwa i interesu publicznego
GIK1A_U01	pozyskiwać, integrować i interpretować informacje z literatury polskiej i obcej, samokształcić się, wyciągać wnioski oraz formułować i uzasadniać opinie z pozyskanych informacji, szczególnie w zakresie geodezji i kartografii i dziedzin pokrewnych
GIK1A_U07	zastosować analizę statystyczną i odpowiednie algorytmy do oceny wyników obserwacji, rozwiązywania zadań geodezyjnych i prognoz badanych zjawisk przestrzennych
GIK1A_U12	przygotować proste opracowanie naukowe w języku polskim i krótką informację naukową w języku obcym, przedstawiające wyniki własnych badań naukowych oraz przygotować i przedstawić prezentację zagadnień z zakresu geodezji i kartografii oraz wybranych zagadnień specjalistycznych
GIK1A_W05	podstawy technologii informacyjnych, programowania komputerowego i baz danych
GIK1A_W06	podstawy statystycznych i probabilistycznych metod analizy wyników obserwacji