**Nazwa przedmiotu:**

Techniki analizy i przetwarzania danych

**Koordynator przedmiotu:**

dr hab. inż. Maciej Kozłowski, prof. uczelni, Wydział Transportu Politechniki Warszawskiej, Zakład Systemów Informatycznych i Mechatronicznych w Transporcie

**Status przedmiotu:**

Obowiązkowy

**Poziom kształcenia:**

Studia I stopnia

**Program:**

Transport

**Grupa przedmiotów:**

Specjalnościowe

**Kod przedmiotu:**

**Semestr nominalny:**

5 / rok ak. 2022/2023

**Liczba punktów ECTS:**

5

**Liczba godzin pracy studenta związanych z osiągnięciem efektów uczenia się:**

150 godz., w tym: praca na wykładach 30 godz., praca na zajęciach komputerowych 30 godz., zapoznanie się ze wskazaną literaturą w zakresie wykładu 30 godz., przygotowanie do zaliczenia wykładu 20 godz., konsultacje w zakresie wykładu 6 godz., udział w zaliczeniu wykładu 2 godz., przygotowanie do kolokwiów z zajęć komputerowych 20 godz., konsultacje w zakresie zajęć komputerowych 6 godz., udział w zaliczeniu zajęć komputerowych 2 godz.

**Liczba punktów ECTS na zajęciach wymagających bezpośredniego udziału nauczycieli akademickich:**

2,5 pkt. ECTS (76 godz., w tym: praca na wykładach 30 godz., praca na zajęciach komputerowych 30 godz., konsultacje w zakresie wykładu 6 godz., udział w zaliczeniu wykładu 2 godz., konsultacje w zakresie zajęć komputerowych 6 godz., udział w zaliczeniu zajęć komputerowych 2 godz.)

**Język prowadzenia zajęć:**

polski

**Liczba punktów ECTS, którą student uzyskuje w ramach zajęć o charakterze praktycznym:**

2,0 pkt. ECTS (58 godz., w tym: praca na zajęciach komputerowych 30 godz., przygotowanie do kolokwiów z zajęć komputerowych 20 godz., konsultacje w zakresie zajęć komputerowych 6 godz., udział w zaliczeniu zajęć komputerowych 2 godz.)

**Formy zajęć i ich wymiar w semestrze:**

|  |  |
| --- | --- |
| Wykład:  | 30h |
| Ćwiczenia:  | 0h |
| Laboratorium:  | 0h |
| Projekt:  | 0h |
| Lekcje komputerowe:  | 30h |

**Wymagania wstępne:**

 brak

**Limit liczby studentów:**

Wykład: 100 osób, zajęcia komputerowe: 15 osób.

**Cel przedmiotu:**

Celem przedmiotu jest nabycie przez studenta wiedzy z zakresu technik analizy i przetwarzania danych stosowanych w transporcie

**Treści kształcenia:**

Wykład:
Wstęp do baz danych. Współczesne modele baz danych. Klasyczne relacyjne bazy danych. Bazy danych NoSQL oraz ich zastosowanie. Formaty przechowywania danych. Integracja danych heterogenicznych. Procesy przetwarzania danych. Przetwarzanie transakcyjne OLTP. Podstawy języka SQL. Analizy analityczna danych OLAP. Hurtownie danych. Narzędzia oraz metody analizy wielowymiarowej danych. Metody i narzędzia generowania raportów w bazach danych. Zagadnienia związane z pojęciem BigData. Perspektywy rozwoju nowoczesnych systemów baz danych.
Metody maszynowego uczenia w analizie dużych zbiorów danych, algorytmy uczenia: uczenie nadzorowane i nienadzorowane, funkcja celu, metoda optymalizacji gradientu opadania, kondycjonowanie danych, MNK, PCA, regresja (uogólniony model liniowy), klasyfikacja, grupowanie, sieci neuronowe płaskie, propagacja wsteczna błędu, sieci neuronowe głębokie, metoda transfer learning
- środowiska analizy danych.
Zajęcia komputerowe:
Podstawy komunikacji z bazą danych. Import danych do bazy. Integracja danych różnorodnych. Przetwarzanie transakcyjne z wykorzystaniem języka SQL. Wyszukiwanie oraz analiza danych w trybie OLTP. Projektowanie hurtowni danych. Zasilanie hurtowni danymi. Analiza analityczna OLAP z zastosowaniem wybranych narzędzi. Generowanie raportów wynikowych. Podstawy budowy oraz operowania bazą danych NoSQL.
Zapoznanie się z metodami kondycjonowania danych, MNK, wieloczynnikowa PCA, regresja (uogólniony model liniowy), klasyfikacja, grupowanie. Zastosowania sieci neuronowych płaskich. Sieci neuronowe głębokie Metoda transfer learning.

**Metody oceny:**

Wykład:
Ocena jest wystawiana na podstawie liczby punktów uzyskanych przez studenta na zaliczeniu; zaliczenie przeprowadzane jest w formie testu komputerowego zamkniętego składającego się z 15..30 pytań z zakresu zagadnień omawianych na wykładach. Pytania obejmują każdy z efektów kształcenia w zakresie wiedzy. W celu zaliczenia wykładów należy uzyskać pozytywną (> 50% poprawnych odpowiedzi) ocenę dla każdego z efektów.
Zajęcia komputerowe:
Ocena jest wystawiana na podstawie sumy punktów uzyskanych na kolokwium zaliczeniowym. Kolokwium jest oceniane w skali od 0 do 100 punktów. Punkty są przyznawane zgodnie z ustaloną listą obejmującą efekty kształcenia w zakresie umiejętności, która zawiera nazwę ocenianej cechy rozwiązania oraz liczbę przyznawanych punktów. W celu zaliczenia zajęć komputerowych należy uzyskać pozytywną (>50% punktów możliwych do zdobycia) ocenę dla każdego z efektów kształcenia w zakresie umiejętności.
Ocena zintegrowana:
Ocena końcowa z przedmiotu jest średnią arytmetyczną ocen z wykładu oraz zajęć komputerowych, pod warunkiem, że są one pozytywne.

**Egzamin:**

nie

**Literatura:**

Literatura podstawowa:
1) Larose Daniel T. Metody i modele eksploracji danych. Wydawnictwo PWN, Warszawa, 2021.
2) Ian Goodfellow, Yoshua Bengio, Aaraon Couville, Deep Learning, Systemy uczące się, PWN 2018,
3) Szeliga Marcin, Data Science I uczenie maszynowe, PWN 2018

Wybrane źródła internetowe:
4) Maszynowe uczenie, wykłady i ćwiczenia, Uniwersytet Warszawski – Kapitał Ludzki: https://brain.fuw.edu.pl/edu/index.php/Uczenie\_maszynowe\_i\_sztuczne\_sieci\_neuronowe
5) Internetowy podręcznik statystyki
https://www.statsoft.pl/textbook/stathome\_stat.html?https%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstnonlin.html
6) UFLDL Tutorial: http://deeplearning.stanford.edu/wiki/index.php/UFLDL\_Tutorial
7) Deep Learning Tutorial: http://ufldl.stanford.edu/tutorial/
Uczenie maszynowe i głębokie w rozpoznawaniu obrazów, 2) Uczenie maszynowe w Matlabie Nagrane webinaria do pobrania, (3) Wprowadzenie do statistics toolbox, (4) Dopasowanie krzywych w Matlab. Oprogramowanie Naukowo Techniczne, Kraków: http://www.ont.com.pl/do-pobrania/nagrane-webinaria/, Matlab: Statistics and Machine Learning Toolbox Examples

**Witryna www przedmiotu:**

www.wt.pw.edu.pl

**Uwagi:**

O ile nie powoduje to zmian w zakresie powiązań danego przedmiotu z kierunkowymi efektami uczenia się w treściach kształcenia mogą być wprowadzane na bieżąco zmiany związane z uwzględnieniem najnowszych osiągnięć naukowych.

## Charakterystyki przedmiotowe

### Profil ogólnoakademicki - wiedza

**Charakterystyka W01:**

Zna współczesne modele baz oraz hurtowni danych, języki oraz narzędzia przetwarzania i analizy dużych zbiorów danych

Weryfikacja:

Od 5 do 10 pytań zamkniętych na komputerowym teście, wymagana jest poprawna odpowiedź na co najmniej 50% z tych pytań.

**Powiązane charakterystyki kierunkowe:** Tr1A\_W06

**Powiązane charakterystyki obszarowe:** P6U\_W, I.P6S\_WG.o

**Charakterystyka W02:**

Zna najważniejsze algorytmy przetwarzania i analizy dużych zbiorów danych

Weryfikacja:

Od 5 do 10 pytań zamkniętych na komputerowym teście, wymagana jest poprawna odpowiedź na co najmniej 50% z tych pytań.

**Powiązane charakterystyki kierunkowe:** Tr1A\_W07

**Powiązane charakterystyki obszarowe:** P6U\_W, I.P6S\_WG.o

**Charakterystyka W03:**

Zna metody maszynowego uczenia w analizie dużych zbiorów danych

Weryfikacja:

Od 5 do 10 pytań zamkniętych na komputerowym teście, wymagana jest poprawna odpowiedź na co najmniej 50% z tych pytań.

**Powiązane charakterystyki kierunkowe:** Tr1A\_W12

**Powiązane charakterystyki obszarowe:** P6U\_W, I.P6S\_WG.o

### Profil ogólnoakademicki - umiejętności

**Charakterystyka U01:**

Potrafi dokonać analizy zbiorów danych przy rozwiązywaniu problemów transportowych

Weryfikacja:

Indywidualne zadanie zaliczeniowe na kolokwium wykonywane na komputerze, wymagane jest uzyskanie co najmniej 50% wartości punktów z zakresu efektu

**Powiązane charakterystyki kierunkowe:** Tr1A\_U11

**Powiązane charakterystyki obszarowe:** P6U\_U, I.P6S\_UW.o, III.P6S\_UW.o

**Charakterystyka U02:**

Potrafi wykorzystać metody maszynowego uczenia w analizie dużych zbiorów danych

Weryfikacja:

Indywidualne zadanie zaliczeniowe na kolokwium wykonywane na komputerze, wymagane jest uzyskanie co najmniej 50% wartości punktów z zakresu efektu

**Powiązane charakterystyki kierunkowe:** Tr1A\_U10

**Powiązane charakterystyki obszarowe:** P6U\_U, I.P6S\_UW.o, III.P6S\_UW.o

### Profil ogólnoakademicki - kompetencje społeczne

**Charakterystyka KS01:**

Jest gotów do identyfikacji oraz rozstrzygnięcia problemów związanych z informatyzacją w transporcie z uwzględnieniem zmieniających się potrzeb społecznych

Weryfikacja:

Ocena aktywności podczas zajęć – wymagane jest rozwiązanie co najmniej 50% zadań wykonywanych podczas zajęć praktycznych

**Powiązane charakterystyki kierunkowe:** Tr1A\_K03

**Powiązane charakterystyki obszarowe:** P6U\_K, I.P6S\_KR