



KARTA OPISU PRZEDMIOTU - SYLABUS

Nazwa przedmiotu

Analiza danych [S2Teleinf2>AD]

Przedmiot

Kierunek studiów

Teleinformatyka

Rok/Semestr

1/2

Studia w zakresie (specjalność)

Sztuczna inteligencja i uczenie maszynowe

Profil studiów

ogólnoakademicki

Poziom studiów

drugiego stopnia

Język oferowanego przedmiotu

polski

Forma studiów

stacjonarne

Wymagalność

obligatoryjny

Liczba godzin

Wykład

24

Laboratorium

30

Inne

24

Ćwiczenia

0

Projekty/seminaria

0

Liczba punktów ECTS

4,00

Koordynatorzy

dr inż. Sławomir Maćkowiak

slawomir.mackowiak@put.poznan.pl

Wykładowcy

Wymagania wstępne

1. Student posiada usystematyzowaną podstawową wiedzę z zakresu analizy matematycznej, algebry, rachunku prawdopodobieństwa i analizy statystycznej. 2. Posiada uporządkowaną i podbudowaną matematycznie wiedzę w zakresie teorii sygnałów jedno i wielowymiarowych niezbędną do rozumienia reprezentacji i analizy sygnałów w dziedzinie czasu i częstotliwości. 3. Zna zasady konstrukcji programów komputerowych, posiada wiedzę z zakresu informatyki i zna składnię wybranych języków programowania wysokiego poziomu (np.: C, C++, C#, Python, MatLab). 4. Potrafi swobodnie porozumiewać się w języku angielskim, potrafi rozmawiać w j. angielskim o sprawach zawodowych, potrafi ze zrozumieniem korzystać z literatury fachowej w j. angielskim. 5. Zna ograniczenia własnej wiedzy i umiejętności, rozumie konieczność dalszego kształcenia się. 6. Potrafi realizować projekty zespołowe.

Cel przedmiotu

Analiza danych sprawia, że dzięki ich dużym i mniejszym kolekcjom uzyskujemy wartościową wiedzę, która pozwala na podejmowanie najlepszych decyzji. Dzieje się to poprzez odkrywanie wzorców lub trendów. Możemy łatwo wydobywać wiedzę z danych. Przedmiot skupia się na głębokim zrozumieniu procesów analizy i interpretacji danych w różnych kontekstach. Jest to kurs interdyscyplinarny, który integruje statystykę, programowanie, matematykę oraz zagadnienia związane z bazami danych. Studenci uczą się zbierania, przetwarzania, wizualizacji oraz interpretacji danych przy użyciu różnorodnych narzędzi i technik, takich jak języki programowania, narzędzia do analizy statystycznej i narzędzia do eksploracji danych. Głównym celem tego przedmiotu jest przygotowanie studentów do efektywnego wykorzystywania danych w podejmowaniu decyzji i rozwiązywaniu problemów w różnych dziedzinach, takich jak nauka, biznes, medycyna i wiele innych.

Przedmiotowe efekty uczenia się

Wiedza:

Rozumienie koncepcji i teorii związanych z analizą danych, w tym podstaw statystyki, teorii prawdopodobieństwa i metod analizy danych. K2_W01

Znajomość różnych narzędzi, technik i języków programowania stosowanych w analizie danych, takich jak Python, R, SQL itp. K2_W04, K2_W11

Wiedza na temat zagadnień związanych z etyką i bezpieczeństwem danych w kontekście analizy danych. K2_W08

Umiejętności:

Umiejętność zbierania, czyszczenia i przygotowania danych do analizy. K2_U06

Zdolność do tworzenia wizualizacji danych i prezentowania wyników analizy w sposób zrozumiały dla innych. K2_U03, K2_U07, K2_U16

Umiejętność stosowania technik analizy danych, w tym regresji, klasyfikacji, analizy skupień i innych metod, do rozwiązywania problemów i podejmowania decyzji. K2_U01, K2_U06, K2_U15

Kompetencje społeczne:

Umiejętność pracy w zespole przy rozwiązywaniu problemów związanych z analizą danych. K2_K01

Zdolność do efektywnej komunikacji i prezentacji wyników analizy danych przed innymi osobami, zarówno technicznie jak i nietechnicznie. K2_K04

Świadomość etycznych i społecznych aspektów związanych z pracą z danymi oraz zdolność do podejmowania odpowiedzialnych decyzji w kontekście analizy danych. K2_K02, K2_K06

Metody weryfikacji efektów uczenia się i kryteria oceny

Efekty uczenia się przedstawione wyżej weryfikowane są w następujący sposób:

1. Wykład

Egzamin pisemny i/lub ustny. Egzamin składa się z kilku - kilkunastu pytań (w zależności od przyjętego charakteru pytań) i dotyczy treści przedstawionych podczas wykładów. Dokładny charakter pytań egzaminacyjnych zostanie studentom przedstawiony podczas jednego z ostatnich wykładów. Próg zdania egzaminu: 50% punktów.

W przypadku zaliczenia pisemnego i ustnego punkty są sumowane.

Skala ocen: <50% - 2,0 (ndst); 50% do 59% - 3,0 (dst); 60% do 69% - 3,5 (dst+) ; 70% do 79% - 4,0 (db); 80% do 89% - 4,5 (db+); 90% do 100% - 5,0 (bdb).

2. Laboratorium

Umiejętności osiągnięte w laboratorium określa się na podstawie raportów (sprawozdań) z przeprowadzonych ćwiczeń laboratoryjnych (OL) oraz zaliczenia końcowego (ZK) w formie samodzielnie realizowanego ćwiczenia lub projektu.

Kompetencje społeczne (KS) ocenia się na podstawie oceny umiejętności aktywnego słuchania, umiejętności współpracy i efektywnego udziału w dyskusjach zespołowych oraz poziomu zaangażowania w procesy rozwiązywania problemów .

Wyznacza się średnią ważoną: $OK = 0,5 \times OL + 0,3 \times ZK + 0,2 \times KS$ i wystawia oceny:

5,0 dla $OK > 4,75$;

4,5 dla $4,75 > OK > 4,25$;

4,0 dla $4,25 > OK > 3,75$;

3,5 dla $3,75 > OK > 3,25$;

3,0 dla $3,25 > OK > 2,75$;

2,0 dla OK < 2,75.

Treści programowe

Wykład obejmuje następujące zagadnienia:

Wprowadzenie do analizy danych (Definicja analizy danych i jej znaczenie. Różnice między danymi strukturalnymi a niestructuralnymi. Proces analizy danych: zbieranie, czyszczenie, eksploracja i modelowanie danych.)

Statystyka w analizie danych (Podstawy statystyki: miary tendencji centralnej i rozproszenia. Rozkłady prawdopodobieństwa i testowanie hipotez. Regresja liniowa i nieliniowa w analizie danych.)

Techniki wizualizacji danych (Tworzenie wykresów i grafik. Wizualizacja danych wielowymiarowych.

Mapy cieplne i wykresy interaktywne.)

Eksploracja danych i przygotowanie danych (Czyszczenie danych: usuwanie brakujących danych, obsługa wartości odstających., ekstrakcja cech: wybór odpowiednich cech do modelowania., Inżynieria cech: tworzenie nowych cech na podstawie istniejących danych.)

Metody eksploracji danych (Redukcja wymiarowości danych., Analiza skupień (klasyfikacja) danych.

Analiza głównych składowych (PCA). Wykorzystanie algorytmów uczenia maszynowego.)

Bazy danych i SQL w analizie danych (Modele baz danych i normalizacja. Język SQL do przetwarzania danych. Łączenie i agregowanie danych z różnych źródeł.)

Analiza dużych zbiorów danych (Big Data) (Wyzwania związane z przetwarzaniem dużych zbiorów danych. Narzędzia i technologie do analizy dużych danych, Przetwarzanie strumieni danych (streaming data).)

Etyka i bezpieczeństwo w analizie danych (Kwestie prywatności danych i ochrona danych osobowych.

Etyczne aspekty analizy danych, takie jak bias i dyskryminacja. Zasady bezpieczeństwa danych i zarządzanie ryzykiem w analizie danych.)

Laboratorium

Podstawy analizy danych i korzystanie z narzędzi przetwarzania

Praca z danymi w różnych formatach

Interaktywna wizualizacja danych

Inżynieria cech, analiza szeregów czasowych i przetwarzanie sygnałów

Zaawansowana analiza danych tekstowych i obrazów

Tematyka zajęć

Wykład obejmuje następujące zagadnienia:

Wprowadzenie do analizy danych (Definicja analizy danych i jej znaczenie. Różnice między danymi strukturalnymi a niestructuralnymi. Proces analizy danych: zbieranie, czyszczenie, eksploracja i modelowanie danych.)

Statystyka w analizie danych (Podstawy statystyki: miary tendencji centralnej i rozproszenia. Rozkłady prawdopodobieństwa i testowanie hipotez. Regresja liniowa i nieliniowa w analizie danych.)

Techniki wizualizacji danych (Tworzenie wykresów i grafik. Wizualizacja danych wielowymiarowych.

Mapy cieplne i wykresy interaktywne.)

Eksploracja danych i przygotowanie danych (Czyszczenie danych: usuwanie brakujących danych, obsługa wartości odstających., ekstrakcja cech: wybór odpowiednich cech do modelowania., Inżynieria cech: tworzenie nowych cech na podstawie istniejących danych.)

Metody eksploracji danych (Redukcja wymiarowości danych., Analiza skupień (klasyfikacja) danych.

Analiza głównych składowych (PCA). Wykorzystanie algorytmów uczenia maszynowego.)

Bazy danych i SQL w analizie danych (Modele baz danych i normalizacja. Język SQL do przetwarzania danych. Łączenie i agregowanie danych z różnych źródeł.)

Analiza dużych zbiorów danych (Big Data) (Wyzwania związane z przetwarzaniem dużych zbiorów danych. Narzędzia i technologie do analizy dużych danych, Przetwarzanie strumieni danych (streaming data).)

Etyka i bezpieczeństwo w analizie danych (Kwestie prywatności danych i ochrona danych osobowych.

Etyczne aspekty analizy danych, takie jak bias i dyskryminacja. Zasady bezpieczeństwa danych i zarządzanie ryzykiem w analizie danych.)

Laboratorium

Podstawy analizy danych i korzystanie z narzędzi przetwarzania

Praca z danymi w różnych formatach

Interaktywna wizualizacja danych

Inżynieria cech, analiza szeregów czasowych i przetwarzanie sygnałów

Metody dydaktyczne

1. Techniki aktywnego uczenia się: Strategie aktywnego uczenia się, takie jak dyskusje w grupach, rozwiązywanie problemów i studia przypadków, aby aktywnie zaangażować studentów w proces uczenia się. Zachęcanie do wspólnego uczenia się i interakcji, aby wspierać krytyczne myślenie i stosowanie wiedzy.
2. Integracja technologii: Wykorzystanie narzędzi i platformy technologicznej, aby poprawić jakość nauki. Korzystanie z narzędzi do współpracy online podczas sesji burzy mózgów, wirtualnych symulacji do rozwiązywania problemów oraz prezentacji multimedialnych, aby dostarczać wciągające treści. Ponadto wykorzystanie internetowych forów dyskusyjnych lub systemów zarządzania nauczaniem do asynchronicznego uczenia się i udostępniania zasobów.
3. Uczenie się oparte na przypadkach: Włączenie rzeczywistych studiów przypadków do wykładów i laboratoriów, aby zademonstrować praktyczne zastosowanie kreatywnego myślenia w rozwiązywaniu problemów technicznych. Zachęci to do analizowania i omawiania przypadków, identyfikowania kreatywnych rozwiązań i refleksji nad procesem podejmowania decyzji.
4. Informacja zwrotna i nauczanie od studentów: Wprowadzenie mechanizmów informacji zwrotnej od studentów, w ramach których uczniowie przekazują konstruktywne informacje zwrotne na temat podejść do rozwiązywania problemów lub rozwiązań projektowych swoich rówieśników. Zachęcanie do sesji nauczania studenckiego, podczas których studenci mogą dzielić się swoją wiedzą i kreatywnymi technikami z kolegami.
5. Nauka oparta na projektach: Włączenie nauki opartej na projektach do programu nauczania, w której studenci pracują nad rzeczywistymi problemami lub projektują wyzwania wymagające kreatywnego myślenia. Takie podejście pozwala zastosować swoje umiejętności, przeprowadzić dogłębne badania i opracować innowacyjne rozwiązania poprzez praktyczne, empiryczne uczenie się.

Literatura

Podstawowa:

- Brandt, Analiza danych. Metody statystyczne, Wydawnictwa Naukowe PWN,
- Avinash Navlani, Armando Fandango, Ivan Idris, Python i praca z danymi. Przetwarzanie, analiza, modelowanie i wizualizacja. Wydanie III, Helion
- Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, 2009,
- Wes McKinney, Python for Data Analysis, O'Reilly Media, 2017,
- Wes McKinney, Python w analizie danych. Przetwarzanie danych za pomocą pakietów Pandas i NumPy oraz środowiska IPython. Wydanie II, Helion
- Tamara Dinevski, Wilfried Gansterer, Data Analysis in the Cloud: Models, Techniques and Applications, Springer, 2019,

Uzupełniająca:

- Klonecki W.: Statystyka dla inżynierów. Wydawnictwo Naukowe PWN SA, Warszawa, 1999 - Sobczyk M.: Statystyka. Wydawnictwo Naukowe PWN SA, Warszawa, 2002
- Zięba, Analiza danych w naukach ścisłych i technicznych, Wydawnictwo naukowe PWN, Warszawa, 2013.
- Wawrzyński P., Podstawy sztucznej inteligencji, Oficyna Wydawnicza Politechniki Warszawskiej, 2019

Bilans nakładu pracy przeciętnego studenta

	Godzin	ECTS
Łączny nakład pracy	104	4,00
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	54	2,00
Praca własna studenta (studia literaturowe, przygotowanie do zajęć laboratoryjnych/ćwiczeń, przygotowanie do kolokwium/egzaminu, wykonanie projektu)	50	2,00