

PROGRAM

Ilość godzin dydaktyczny: 215 godz.

MODUŁ 1 Inżynieria danych (134 godz.)

1. Wprowadzenie do technologii Big Data

- czym jest big data – filozofia, technologia, analityka
- model 42V
- podejścia do problemów skali Big Data, algorytm MapReduce, koncepcja BigTable, NoSQL oraz rozproszone systemy plikowe HDFS
- projektowanie rozwiązań klastrowych i chmurowych oraz integracja systemów Big Data z istniejącymi systemami i oprogramowaniem

2. Podstawy programowania

Język programowania R

- wprowadzenie do R (Praca w R , Przestrzeń robocza w RStudio, Pakiety)
- struktury danych (Wektor, Macierz, Faktor/czynnik, Lista,)Ramka danych
- przetwarzanie danych (Pakiet tidyverse, Import danych, Filtrowanie, Wybieranie kolumn, Tworzenie nowej zmiennej, Zmiana nazwy zmiennej, Podsumowanie danych, Grupowanie, Sortowanie, Łączenie zbiorów, Wąska i szeroka reprezentacja danych, Eksport danych)
- wizualizacja danych (Korzystanie z wbudowanych funkcji, Pakiet ggplot2, Wykres punktowy, Wykres pudełkowy, Wykres liniowy, Wykres kolumnowy)
- programowanie w R (Tworzenie funkcji, Pętle, Instrukcje warunkowe)

Język programowania Python

- wprowadzenie do języka Python
- składnia języka Python
- wyrażenia regularne w języku Python
- klasy, obiekty, moduły
- skrypty i kontrola przepływu programu
- zastosowanie bibliotek naukowych: NumPy, SciPy, Matplotlib
- rozwiązywanie praktycznych problemów Big Data z wykorzystaniem języka Python

3. Metody przetwarzania i analizy danych

- eksploracja danych - wstęp, techniki eksploracji danych (zmienne jakościowe i numeryczne, relacje wielowymiarowe, dyskretyzacja, regresja wielokrotna, algorytm najbliższych sąsiadów, funkcje decyzyjne, drzewa klasyfikacyjne i regresyjne), metody przetwarzania i analizy danych (regresja logistyczna, regresja liniowa, uogólnione modele liniowe (GLM), modele mieszane (Mixed models), algorytm locality-sensitive hashing)
- przetwarzanie Big Data za pomocą Apache Hadoop, implementacja algorytmów w paradygmacie MapReduce, wykorzystanie narzędzi analitycznych Pig i Hive
- przetwarzanie Big Data z użyciem Apache Spark, praca z platformą analityczną SAS LASR Server

4. Algorytmy i struktury danych

- podstawowe definicje, schematy zapisu algorytmów
- podstawowe struktury danych: zmienne, wskaźniki i referencje, rekordy, tablice i macierze. Struktury listowe: lista, stos, kolejka. Struktury grafowe: graf, drzewo. Struktury słownikowe. Struktury programistyczne: klasa, wyliczenie, struktura (rekord)
- algorytmy iteracyjne i rekurencyjne. Sortowanie: bąbelkowe, przez selekcję i wstawianie, quicksort. Wyszukiwanie: liniowe i binarne. Grafowe: przeszukiwanie w głąb i wszerz, cykle: Eulera i Hamiltona, przeszukiwanie drzew

5. Bazy danych

- relacyjny model baz danych
- wprowadzenie do języka strukturalnego SQL: składnia SQL: DML, DDL, DCL, DQL
- praktyczne zastosowanie SQL – zapytania analityczne
- metody efektywnego tworzenia modeli danych i ich pozyskiwania z wykorzystaniem SQL
- czym jest baza danych i hurtownia danych. Rys historyczny, silniki baz danych; konceptualne modelowanie danych z wykorzystaniem diagramów ER, zbiory encji, atrybuty, związki, klucze, krotność i obligatoryjność
- modelowanie danych, relacyjny model danych, podstawowe operacje na bazie danych z wykorzystaniem języka SQL: dodawanie, suwanie, modyfikowanie, wyświetlanie danych, określanie schematu bazy danych, rozszerzenia relacyjnych baz danych, elementy baz danych NoSQL
- definicja hurtowni danych wg. B. Inmona, R. Kimballa, analiza wymiarów hurtowni danych, zasady budowania hurtowni danych, model koncepcyjny, model logiczny oraz implementacja fizyczna hurtowni danych, architektura hurtowni danych, kopie zapasowe hurtowni, archiwizacja, ekstrakcja i ładowanie danych, Aplikacje CRT – analiza przykładowego oprogramowania oraz jego obsługa, przygotowanie danych dla hurtowni danych, proces ETL, wymiary w hurtowni danych, analiza danych (OLAP cube, analiza w bazach NoSQL)

6. Machine Learning

- charakterystyka sztucznej inteligencji
- metody i zastosowania uczenia maszynowego
- sieć neuronowa jako klasyczny przykład SI w zakresie ML
- reprezentacja wiedzy i jej konsekwencje
- błędy ludzi, które powielają lub których unikają maszyny
- dobór metody z zakresu SI i ML do klasy zadania
- optymalizacja i walidacja rozwiązań

MODUŁ 2 Analityka Big Data (60 godz.)

1. Analiza sieci społecznych

- SNA jako metoda i narzędzie badawcze
- zastosowanie w biznesie (obserwowanie zmian społeczno-gospodarczych, analiza aktywności użytkowników internetu)
- oprogramowanie badawcze i jego praktyczne zastosowanie

-
- definicja sieci społecznych, aktora sieci społecznej, połączenia i zbioru relacji w sieciach społecznych, zastosowanie w biznesie (obserwowanie zmian społeczno-gospodarczych, analiza aktywności użytkowników internetu)
 - analiza sieci społecznej jako metoda badawcza, sieci społeczne a wirtualne sieci społeczne – różnice i podobieństwa, znaczenie analizy sieci społecznych (SNA), poziomy analizy sieci społecznych (makro, mezo i mikro)
 - podstawy analizy sieci społecznych (matematyczna teoria grafów, algebra macierzowa, statystyka, miary w analizie sieci społecznych: centralność, prestiż, pośrednictwo, bliskość, gęstość, kapitał społeczny, zastosowanie analizy sieci społecznych – case study: wyszukiwanie kluczowych użytkowników, marketing skierowany do grupy docelowej, systemy rekomendacyjne, tworzenie grup projektowych w sieciach społecznych, propagacja nowych idei, kolaboratywne zarządzanie wiedzą

2. Wizualizacja danych

- co to jest wizualizacja danych i dlaczego jest potrzebna,
 - czym jest Business Intelligence,
 - narzędzia wizualizacji danych
 - ogólne zasady i dobre praktyki wizualizacji danych,
 - metody wizualizacji danych: wykresy, infografiki, dashboardy i story,
 - typy wykresów używanych do wizualizacji danych - kiedy jakiego używać, kiedy i jakich nie powinno się używać
 - przygotowywanie danych do wizualizacji – rodzaje zmiennych, cech uporządkowania danych, praktyczne przykłady porządkowania zbiorów danych
 - Tableau - prezentacja aplikacji wraz z praktycznym tworzeniem: wykresów - 13 typów wykresów, kiedy używać, jak budować, dashboardów, story
 - Alteryx – prezentacja aplikacji i jej możliwości – metody podłączania danych oraz rodzaje źródeł danych, podstawowe narzędzia (Tools), konstruowanie workflow'u, zapisywanie tabel wynikowych do plików, ćwiczenia praktyczne

3. Internetowe narzędzia analityczne

- narzędzia służące do: monitoringu mediów, zarządzania social media, influencer marketingu, badania zaangażowania użytkowników, content marketingu itp.
- analityka internetowa: wprowadzenie, konfiguracja narzędzi analitycznych, budowa kompetencji w zespole, tworzenie ciągłych procesów analizy danych
- przegląd wybranych narzędzi służących do: monitoringu mediów, zarządzania social media, influencer marketingu, badania zaangażowania użytkowników, content marketingu
- analityka internetowa w działaniu: Google Analytics; Moz Analytics; TrueSocial Metrics; Woopra; BrightEdge; CrazyEgg; Heatmap; Yandex Metrica; Adobe Analytics; Webtrends; IBM Digital Analytics; Piwik/Matomo.

4. Badania marketingowe i społeczne

- typy badań w marketingu i naukach społecznych (np. badania ilościowe, jakościowe, badania satysfakcji klientów, badanie opinii publicznej, badania fokusowe, eye-tracking)
- proces badawczy
- metody, narzędzia i techniki badawcze

-
- wnioski i raportowanie

MODUŁ 3 Człowiek w świecie technologii (12 godz.)

1. Psychologia Big Data - mariaż Big Data i psychologii empirycznej

- psychologiczny model OCEAN i jego wykorzystanie w tworzeniu psychograficznych profili konsumentów/wyborców w powiązaniu z danymi z FB. Przewidywanie cech i atrybutów na podstawie cyfrowej rejestracji ludzkich zachowań
- wybrane badania wykorzystujące Big Data w analizie ludzkich zachowań i ich konfrontacja z dotychczasowymi ustaleniami psychologii
- laboratorium psychologiczne oparte na Big Data: The Psychometric Centre, University of Cambridge
- redukcjonistyczne ujęcie (psychologii) człowieka jako skończonego zbioru danych behawioralnych i neurodanych – możliwości i konsekwencje
- wybrane elementy cyberpsychologii
- interakcja człowiek-media-technologia w ujęciu psychologicznym i neurokognitywnym. Multitasking i efektywność w zakresie przetwarzania i zapamiętywania informacji. Konsekwencje zacierania granic pomiędzy światem wirtualnym i realnym, psychologiczne i społeczne funkcjonowanie w rzeczywistości hybrydowej. Hipoteza cyfrowej demencji. Wpływ cyfrowych reprezentacji ludzi i środowiska na postawy i zachowania społeczne. Media społecznościowe: następstwa psychologiczne w kontekście najnowszych badań.
- poznawcze, emocjonalne i społeczne zagrożenia wynikające z naużywania mediów cyfrowych; wybrane zaburzenia: m.in. FOMO ("fear of missing out"), DA (Disconnectivity Anxiety), nomofobia ("no mobile phobia")
- procesy poznawcze – jak nasz mózg radzi sobie z procesami poznawczymi i informacjami
- media, technologie i duże zbiory danych a procesy poznawcze, emocjonalne

2. Prawne i etyczne aspekty gromadzenia i przetwarzania danych

- Big Data a prawo do prywatności
- Ustawa o ochronie danych osobowych versus praktyka przetwarzania danych
- tajemnice sektorowe
- Implikacje etyczne gromadzenia i przetwarzania informacji
- ochrona danych osobowych w prawie krajowym i prawie europejskim
- przedstawienie regulacji prawnych dotyczących gromadzenia i przetwarzania danych
- prawa i obowiązki administratorów danych osobowych
- zabezpieczenie danych osobowych przetwarzanych w wersji papierowej i elektronicznej
- działalność Głównego Inspektora Danych Osobowych
- dokumentacja wymagana przez ustawę o ochronie danych osobowych.
- zasady ochrony danych osobowych
- przygotowanie do wdrożenia przepisów Rozporządzenia UE (RODO)
- odpowiedzialność karna, cywilna i administracyjna zakresu ochrony danych osobowych

-
- Mass surveillance. Europejska Konwencja Praw Człowieka i Karta praw podstawowych UE
 - Mass surveillance. Orzecznictwo ETPC
 - Mass surveillance, a przesyłanie danych do państw trzecich

MODUŁ 4 Seminarium (9 godz.)

Sposób weryfikacji osiągnięcia efektów uczenia się:

- ✓ Aktywność i obecność na zajęciach (wykonywanie zadań i aktywności).
- ✓ Projekt indywidualny/zespołowy – praca dyplomowa w formie projektu zaliczeniowego z wykorzystaniem poznanych metod i narzędzi.
- ✓ Absolwenci uzyskują, przewidziane ustawą, świadectwo ukończenia Studiów Podyplomowych CDV.