

## ***STUDIA STACJONARNE DRUGIEGO STOPNIA, specjalizacja DATA SCIENCE***

### **Wprowadzenie do Big Data      WPBD**

W ramach przedmiotu studenci zostaną zapoznani z podstawową problematyką i terminologią związaną z tym obszarem w technologii informacyjnej. Opiszemy kompetencje potrzebne w zawodach analityka danych, inżyniera Big Data oraz profesjonalisty Data Science. Przedstawione zostaną dane dotyczące przyrostu danych generowanych przez maszyny, ludzi oraz organizacje. Omówimy strategie integracji dużych danych, podstawowe założenia algorytmiki przetwarzania oraz metod ich przechowywania. Przedstawione zostaną najważniejsze narzędzia związane z przetwarzaniem dużych zbiorów danych, ich pozyskiwanie, instalacja oraz konfiguracja. Zaprezentujemy metody tworzenia własnych środowisk do analizy Big Data jak również możliwości korzystania z komercyjnych rozwiązań chmurowych. Oprócz rozwiązań softwarowych omówione zostaną możliwości rozwiązań sprzętowych.

### **Eksploracja i wizualizacja danych      EWD**

Celem kursu jest zapoznanie z podstawowym cyklem i metodami eksploracji oraz wizualizacji danych. Po omówieniu metod wstępnej obróbki, wizualizacji, podsumowywania i wstępnej (eksploracyjnej) analizy danych zaprezentowane są podstawowe techniki takie jak: regresja, klasyfikacja czy grupowanie oraz metody oceny modeli. Ponadto zaprezentowane są wybrane przykładowe praktyczne zadania takie jak eksploracja tekstu, sieci WWW czy grafów. Kurs wzbogacony jest o praktyczne ćwiczenia laboratoryjne na danych z użyciem pakietu R.

### **Programowanie dla analityki danych      PAD**

Celem przedmiotu jest zapoznanie studentów z najpopularniejszymi narzędziami programistycznymi stosowanymi do analiz danych, z naciskiem szczególnym na kontekst analiz dużych zbiorów danych, oraz uczenia maszynowego. W ramach wykładu, studenci zapoznają się z dwoma najważniejszymi językami służącymi do analiz danych. W pierwszej połowie kursu przedstawiony zostanie język programowania Python, oraz jego możliwości w analizie danych. Po wprowadzeniu do podstaw języka, studenci zapoznają się z bibliotekami dedykowanymi do analiz danych (Numpy, Pandas), oraz uczenia maszynowego (Sklearn) z uwzględnieniem sieci neuronowych (Theano, Gensim). W drugiej połowie kursu, studenci zapoznają się z podstawami języka R. Po zapoznaniu się ze składnią języka studenci poznają biblioteki dedykowane do analiz danych oraz uczenia maszynowego.

### **Integracja danych i hurtownie danych      IDH**

Integracja danych jest niezwykle istotnym zagadnieniem, stąd celem pierwszej części wykładu jest zapoznanie studentów z tą tematyką. W drugiej części wykładu omawiana jest tematyka

hurtowni danych, czyli dziedziny informatyki obejmującej tworzenie zintegrowanych magazynów danych, zawierających ujednoczone dane historyczne dotyczące przedsiębiorstwa. Bazy takie tworzone są w celach analitycznych (w tym wykorzystujących narzędzia eksploracji danych i odkrywania wiedzy) i archiwizacyjnych. Ze względu na ilość przechowywanych danych i odrębne dziedziny zastosowań, zagadnienia projektowania, utrzymywania i rozbudowy hurtowni danych wymagają podejścia zasadniczo odmiennego niż w przypadku klasycznych baz danych. Celem tej części wykładu jest zapoznanie studentów z metodami projektowania, tworzenia i optymalizacji hurtowni danych.

### **Wprowadzenie do uczenia maszynowego WUM**

Przedmiot ma za zadanie wprowadzić lub przypomnieć podstawowe pojęcia i techniki matematyczne niezbędne do później-szego opanowania bardziej zaawansowanego materiału dotyczącego uczenia maszynowego, oraz, wprowadzić podstawowe ogólne pojęcia dotyczące uczenia maszynowego ze szczególnym uwzględnieniem uczenia głębokiego, np. metoda gradientowa. Następnie omawiane są wybrane zagadnienia dotyczące sztucznych sieci neuronowych: m.in. perceptron, metoda wstecznej propoagacji błędu, sieci konwolucyjne, rekurencyjne, wybrane zagadnienia regularyzacji i praktyczne aspekty uczenia sieci neuronowych.

### **Bezpieczeństwo Big Data BZBD**

W ramach przedmiotu studenci poznają zagadnienia ochrony danych w chmurze obliczeniowej od strony technicznej oraz prawnej. Aspekty techniczne będą skupione na kryptograficznych metodach ochrony danych poprzez zapewnienie poufności, prywatności lub anonimowości. Inne aspekty techniczne będą dotyczyły modelowania oraz analizy systemów IT wykorzystujących dane typu Big Data, ze szczególnym uwzględnieniem najlepszych praktyk w dziedzinie cyberbezpieczeństwa. Aspekty prawne skupiają się na kwestiach ochrony danych osobowych w świetle RODO.

### **Nierelacyjne bazy danych NBD**

Rosnąca popularność baz określanych jako NoSQL oraz języków funkcyjnych stawia informatyków przed nowymi wyzwaniami. Dobranie odpowiedniego narzędzia do wymagań projektu wymaga zrozumienia stojących za różnymi typami baz danych założeń, ich wad i zalet oraz zastosowań, do których są przeznaczone. Rosnąca rola języków funkcyjnych wymaga znajomości przynajmniej podstawowych koncepcji takich, jak niemutowalny stan, pattern matching czy różne modele przetwarzania równoległego. Celem tego przedmiotu jest zapoznanie studentów z najważniejszymi typami nierelacyjnych baz danych i ich zastosowaniami oraz podstawowymi koncepcjami stosowanymi w jednym z najbardziej popularnych języków pozwalających programować w paradygmacie funkcyjnym - Scala. W ramach przedmiotu studenci będą również rozwijać swoje umiejętności programistyczne.

## **Big Data - modelowanie, zarządzanie, przetwarzanie i integracja      BGD**

Podczas wykładu studenci zapoznają się z problemami typu Big Data występującymi w ich codziennym otoczeniu, nauczą się je identyfikować oraz skutecznie nimi zarządzać na potrzeby organizacji. Nauczą się korzystać z praktycznie nieograniczonych już w chwili obecnej źródeł dużych danych, odkrywać nowe, projektować narzędzia do ich pozyskiwania. Odkryją na poziomie praktycznym podstawowe techniki Data Science do modelowania klas abstrakcji Big Data, dobierania odpowiednich operacji do pracy z tymi modelami, nauczą się rozróżniać klasyczne techniki zarządzania danymi od technik zarządzania Big Data. Ponadto przedstawione zostaną narzędzia przetwarzania oraz integracji danych na przykładzie analizy procesu od etapu wydobycia interesujących danych z bazy po wykorzystanie realnych wyników uzyskanych przez zastosowanie kolekcji aplikacji. Oprogramowanie, którego zastosowanie zostanie przedstawione w praktycznych przykładach będzie obejmować m.in. Ekosystem Apache Hadoop, Apache Spark, AsterixDB, HP Vertica, Impala, Neo4j, Redis, SparkSQL.

## **Nauczanie łamigłówkowe - przedmiot internetowy      PUZ**

Nauczanie łamigłówkowe jest znaną od wielu lat metodą nauczania, polegającą na uczeniu poprzez rozwiązywanie zagadek. Celem wykładu jest nauczenie studentów myślenia o formułowaniu i rozwiązywaniu nieustrukturyzowanych problemów, z którymi każdy człowiek spotyka się w życiu codziennym. Dodatkowo wykład ma na uwadze zwiększenie matematycznej świadomości studentów i pokazanie, że świat matematyki i algorytmiki jest światem fascynującym, wciągającym, a jednocześnie bardzo przydatnym.

## **Etyka sztucznej inteligencji      ESI**

Zajęcia poświęcone będą omówieniu najważniejszych zagadnień etycznych w kontekście dążenia do uregulowania AI za pomocą standardów etycznych na poziomie firm, państw i organizacji ponadnarodowych. Standardy etyczne dla AI pełnią obecnie funkcję miękkich regulacji (*soft law*), których zaletami w porównaniu z tradycyjnymi regulacjami są elastyczność procesu regulacyjnego, podatność na interpretację oraz łatwość dostosowania do zmieniających się realiów technicznych i ekonomicznych. Głównym punktem odniesienia będą europejskie regulacje etyczne dotyczące AI oraz rozwiązania służące ich implementacji i egzekwowaniu w kontekście dążenia do odpowiedzialnej AI (*responsible AI*). Podczas zajęć dokonamy również przeglądu istniejących narzędzi kontroli zgodności ze standardami etycznymi.

## **Zastosowania Ucznia Maszynowego      ZUM**

Uczenie maszynowe (ML) jest podstawą sztucznej inteligencji, a jego rozwój pozwolił na stworzenie systemów, które potrafią same uczyć się i wnioskować na podstawie zdobytej wiedzy. Zakres zastosowań takich systemów stale się poszerza, a celem tego wykładu jest zapoznanie słuchaczy z najciekawszymi aplikacjami ML, szczególnie takimi, które bezpośrednio kontaktują się z ludźmi (systemy dialogowe, rozpoznawania mowy,

maszynowego tłumaczenia, itp.). Przedstawione zostaną także specyficzne techniki stosowane w takich systemach, a w szczególności techniki przygotowania danych, stosowane systemy wnioskujące i algorytmy ich działania oraz miary jakości stosowane do oceny aplikacji ML.

## **Grafy i ich zastosowania    GIZ**

Ze względu na wszechobecność reprezentacji grafowej w większości współczesnych zagadnień praktycznych informatyki (exemplum: sieci komputerowe i telekomunikacyjne, sieć WWW, sieci społeczne, semantyczne grafy wiedzy, etc.) wykład ma za zadanie dostarczyć podstawowych narzędzi pojęciowych, algorytmicznych i matematycznych związanych z grafami przydatnych do późniejszego zastosowania przy rozwiązywaniu i modelowaniu praktycznych współczesnych problemów obliczeniowych w przemyśle i nauce. Przekrojowo omówione będą zarówno fundamentalne pojęcia teoretyczne jak i przykłady ich praktycznych zastosowań współczesnych z użyciem komputerów.

## **Technologie i platformy Chmury    TPC**

Platformy cloudowe i aplikacje rozproszone działające w chmurze zyskują obecnie coraz większą popularność chociażby ze względu na łatwość dopasowania infrastruktury do aktualnego obciążenia, czy brak konieczności bezpośredniego zarządzania tą infrastrukturą. W ramach kursu studenci poznają podstawowe pojęcia związane z tematyką platform cloudowych, jak również nauczą się (w podstawowym zakresie) stosować wykorzystywane w tego typu systemach narzędzia i standardy. Podczas zajęć praktycznych szczególny nacisk położony zostanie na konteneryzację (na przykładzie technologii Docker), komunikację sieciową pomiędzy kontenerami (na przykładzie technologii VXLAN), zarządzanie kontenerami w klastrze (na przykładzie technologii Docker Swarm i Kubernetes) oraz zagadnienia związane z projektowaniem rozproszonych aplikacji działających w chmurze (model mikroserwisów).

## ***WYKAZ PRZEDMIOTÓW BLOKOWYCH***

### **Systemy wiedzy    SWI**

Na wykładzie przedstawionych będzie kilka klasycznych metod wydobywania wiedzy w przypadku danych kompletnych, niekompletnych, jak również bardzo dużych danych. Przedstawione będą metody wydobywania reguł akcji oraz meta-akcji. Omówione będą metody odkrywania danych ukrytych w niekompletnych bazach danych (w szczególności metody Chase) oraz metody szczepienia danych. Wprowadzony zostanie język zapytań dla danych rozproszonych i różne jego semantyki. Metody analizy tekstu i analizy sentymentu będą przedstawione. Również automatyczne metody konstrukcji meta-akcji w oparciu o

folksonomie i analizę sentymentu. Będą omówione. Na koniec, metody budowy systemów rekomendujących w medycynie oraz biznesie w przypadku dużych danych będą opisane.

## **Sieciowe systemy zorientowane usługowo SSU**

Tematyką tego wykładu są zaawansowane podstawowe zasady sieciowych systemów zorientowanych na usługi, studiowanie zarówno podstaw jak i zastosowań w tej dziedzinie. Rozpatrywane zagadnienia dotyczą działania protokołów sieciowych, ewolucji istniejących generacji zdalnego wywołania procedur, sieci pokryciowe (overlay), ich zasady implementacji i praktyki, w końcu zaawansowane dynamiczne sieci zorientowane usługowo. Celem jest przyswojenie wiedzy na temat podstaw/technik sieciowych i sieci pokryciowych niezbędnych do konstrukcji systemów zorientowanych na usługi. Ukończenie tego wykładu umożliwia dogłębne zrozumienie różnorodnych pojęć i protokołów komunikacji wraz ze zdobyciem szerokiej wiedzy wymaganej do rozwoju wyrafinowanych aplikacji zorientowanych na usługi. Analizy takich przypadków, programowania projektów i artykułów badawczych wspierają ten wykład.

## **Zaawansowane techniki multimedialne ZTM**

Celem tego przedmiotu jest zapoznanie studentów z koncepcją tzw. *uczenia głębokiego* oraz metodami *sztucznej inteligencji* dotyczącymi szeroko pojętych multimedialnych. Przedmiot pozwala doświadczyć uproszczonego procesu projektowania inteligentnych aplikacji opartych o uczenie maszynowe dla danego rozwiązania. Szczególna uwaga jest poświęcona klasyfikacji danych, modelom generatywnym, analizie audio oraz przetwarzaniu języka naturalnego w kontekście tłumaczenia maszynowego. W ramach przedmiotu studenci rozwijają również swoje umiejętności programistyczne podczas budowy oraz tuningu przygotowanych ćwiczeń.

## **Jak mózg oblicza to co widzimy i jak koordynuje nasze ruchy MKR**

Interaktywne wykłady prowadzone przez AI-badacza mózgu rozpatrujące kompleksowe procesy obliczeniowe mózgu. Będziemy porównywać obliczenia mózgu z obliczeniami komputerowymi pod względem ich kreatywności. Zaczniemy od opisu przetwarzania informacji przez pojedyncze neurony aż do sieci neuronalnych/ANNs i będziemy się starali zrozumieć gdzie mogą być źródła kreatywności. Następnie zajmiemy się zrozumieniem obliczeń i źródeł ich kreatywności związanych z różnymi strukturami mózgu. Zaczniemy od reprezentacji zawartości i zrozumieniem obrazów w układzie wzrokowym ponieważ przez układ wzrokowy uzyskujemy najwięcej informacji. Później omówimy zrozumienie funkcji i organizacji różnych struktur mózgu aż do roli półkul mózgu i ich kreatywności. Od fantomów poprzez reprezentacje ciała będziemy poszerzać analizę aż do reprezentacji siebie (self) i przeanalizujemy moralne tego konsekwencje. Zauważmy że kreatywność wiąże się z procesami fizjologicznymi a patologiczne zmiany w mózgu są ich przeciwieństwem. Głównie skupione na chorobach degeneracyjnych jak choroba Parkinsona ale także będziemy omawiać

patologie związane z innymi uszkodzeniami mózgu. Poprzez analizę patologii mózgu i ich objawów będziemy się starać znaleźć ich mechanizmy i ta droga dojść do kreatywności.

### **Teoria i zastosowanie analizy falkowej AFA**

Tematyką wykładu są teoretyczne podstawy oraz metody projektowania systemów przetwarzania sygnałów za pomocą falek. Studenci będą wykorzystywali oprogramowanie środowiska obliczeniowego Matlab, a w szczególności narzędzia analizy falkowej z interfejsem graficznym użytkownika (*ang.* GUI). Uczestnicy kursu otrzymają dodatkowe oprogramowanie opracowane w ciągu wielu lat przez wykładowcę i jego studentów. Zapoznają się oni z techniką zajmującą centralne miejsce w nowoczesnych pracach badawczych i zastosowaniach w dziedzinie przetwarzania sygnałów. Rozpatrywane zagadnienia dotyczą między innymi przetwarzania obrazów medycznych przy wykorzystaniu techniki wyznaczania maksimów modułu transformaty falkowej, programowania za pomocą transformaty falkowej (wykorzystującej splajny wielomianowe) trajektorii robotów naśladowujących ruchy człowieka z uwzględnieniem ograniczeń prędkości i przyspieszenia, odpornego znakowania wodnego obrazów, badania dynamiki interakcji sygnałów w systemach złożonych z zastosowaniem do analizy skorelowanych sygnałów medycznych, jak również analizy integralności sygnałów. Zagadnienia te będą także przedmiotem projektów końcowych.

### **Cyfrowe przetwarzanie obrazów: algorytmy i zastosowania CPO**

Studenci poznają podstawy teoretyczne przetwarzania sygnałów dwuwymiarowych, obejmujące dwuwymiarowa transformate Fouriera (2-D FFT), transformate Z i równania różnicowe. Omówione zostanie projektowanie dwuwymiarowych filtrów dla zastosowań przetwarzania obrazów. Omawiane zagadnienia będą również obejmowały techniki poprawiania jakości i rekonstrukcji obrazów. Przedstawione będą podstawy przetwarzania obrazów przy wykorzystaniu transformaty falkowej i ich reprezentacja za pomocą krawędzi, ilustrowane przykładami analizy obrazów biomedycznych.

### **Inżynieria oprogramowania INN**

Wykład ten jest wprowadzeniem do teorii i praktyki inżynierii oprogramowania włączając techniki związane z analizą i modelowaniem (design), wytwarzaniem (construction), i testowaniem systemów informatycznych, w szczególności jakościowego oprogramowania dla dużych i złożonych systemów obliczeniowych. Studenci zapoznają się z modelowaniem oprogramowania w UML i kilkoma zunifikowanymi procesami wytwarzania oprogramowania. Używać będziemy narzędzia do budowy, modelowania i testowania oprogramowania w kontekście wolnych i otwartych programów. Studenci biorą udział w zespołowym projekcie badawczym. Wykład ten jest czasochłonny i wymaga wykonania trzech praktycznych zadań poza udziałem w wykładzie.