Informator dla kandydatów na studia podyplomowe

BIG DATA

PRZETWARZANIE I ANALIZA DUŻYCH ZBIORÓW DANYCH

DATA SCIENCE

ALGORYTMY, NARZĘDZIA I APLIKACJE DLA PROBLEMÓW TYPU BIG DATA

WIZUALNA ANALITYKA DANYCH

Wydział Elektroniki i Technik Informacyjnych
POLITECHNIKA WARSZAWSKA

sages

Spis treści

BIG DATA

PRZETWARZANIE I ANALIZA DUŻYCH ZBIORÓW DANYCH

198h
wykładów
i warsztatów

12 500 pln
z możliwością płatności
w dwóch ratach

w cenie kurs
e-learnignowy
z Pythona

Poznaj potencjał studiów

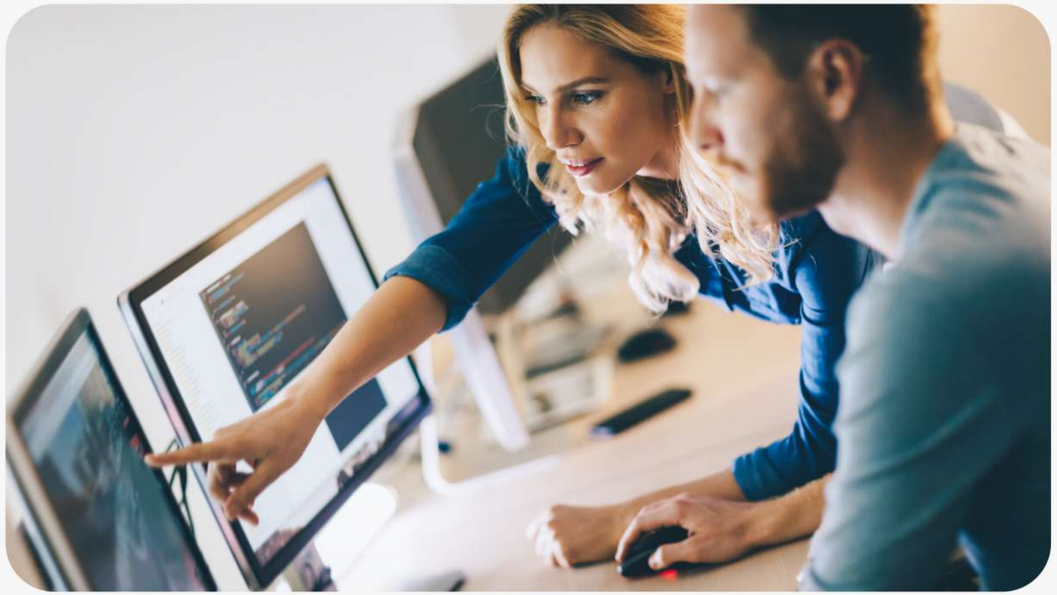
Dlaczego warto studiować duże dane?

Celem studiów jest zdobycie **praktycznych umiejętności analizy dużych zbiorów danych**, zrozumienie podstaw, celu i obszaru zastosowania rezultatów takiej analizy.

W czasie studiów słuchacze zapoznają się z najważniejszymi współczesnymi narzędziami i technologiami związanymi z zagadnieniami Big Data: **Apache Hadoop** i **Spark** w ujęciu programistycznym (**MapReduce**), analitycznym (**Pig** i **Hive**) i administracyjnym, a także bazy **NoSQL**, elementy programowania współbieżnego w językach funkcyjnych oraz **podstawy uczenia maszynowego** w kontekście przetwarzania dużych ilości danych.

Dla kogo są studia Big Data?

Studia przeznaczone są dla osób, które zainteresowane są wykorzystaniem potencjału analizy dużych zbiorów danych w celu wspierania procesu podejmowania decyzji: w biznesie, nauce i innych obszarach działalności.



Doświadczenie w pracy z technologiami **jest mile widziane** (np. podstawowa umiejętność programowania w dowolnym języku, podstawowa znajomość zagadnień związanych z bazami danych i językiem SQL), ale nie jest wymagane.

Szczegółowy program studiów

PROJEKTOWANIE ROZWIĄZAŃ BIG DATA

Zajęcia seminaryjne, poświęcone pracy nad projektami końcowymi.

STUDIUM PRZYPADKU

Zajęcia prowadzone przez przedstawicieli środowisk biznesowych. Prezentacje w ramach przedmiotu obejmują przegląd komercyjnego wykorzystania wybranych metod z obszaru Data Science i Big Data.

WPROWADZENIE DO TECHNOLOGII BIG DATA

- Wstęp do Big Data
- Infrastruktura i narzędzia systemów Big Data
- Wprowadzenie do baz NoSQL
- Uczenie maszynowe w Big Data

Słuchacze poznają historię oraz definicję zagadnienia Big Data, ekosystem stosowanych narzędzi oraz powszechnie wykorzystywanych języków programowania, podział ról i obowiązków spotykany w rozwiązaniach Big Data, różnice pomiędzy przetwarzaniem wsadowym a strumieniowym oraz ich zastosowania.

PROGRAMOWANIE ZORIENTOWANE NA DANE

- Wprowadzenie do systemu Linux
- Wprowadzenie do Python
- Podstawy programowania w Python
- Biblioteka NumPy
- Biblioteka Matplotlib
- Podstawy programowania w Python II
- Współbieżność i równoległość a praca z danymi
- Przegląd biblioteki SciPy
- Anonimizacja danych

Słuchacze nauczą się obsługi najważniejszych poleceń oraz narzędzi w systemie Linux, struktury i składni języka programowania **Python**, wykorzystania kolekcji oraz dedykowanych bibliotek do efektywnego przetwarzania danych takich jak **NumPy, Pandas i Matplotlib**.

WPROWADZENIE DO BAZ NOSQL **CASSANDRA**

- Warsztat z instalacji Cassandry
- Warsztat: Praca z danymi i architektura bazy Cassandra
- Warsztat: mały klaster Cassandry
- Struktura danych
- Oficjalny sterownik do Cassandry dla języka Python/Java
- Konfiguracja i optymalizacja
- Dodatkowe źródła wiedzy
- Demo i przegląd możliwości DataStax Enterprise

Uczestnicy w trakcie zajęć poznają rozwiązanie Apache Cassandra - pokrewne **Google Bigtable** lub **Amazon Dynamo**. Zarówno na poziomie czysto praktycznym - jak również zagłębiając się w architekturę systemów rozproszonych i analizując jak konieczność zapewnienia wysokiej dostępności wpływa na cały proces modelowania danych. W trakcie zajęć Słuchacze poznają takie technologie jak **Cassandra, Docker, Python / Jupyter notebooks**.

WPROWADZENIE DO BAZ NOSQL: **MONGODB**

- Wstęp do MongoDB
- Podstawowe zapytania oraz struktura danych
- Zaawansowane zapytania
- Aggregation Framework
- GridFS
- Replikacja
- Sharding
- Elasticsearch

Po przeprowadzonych zajęciach słuchacze zdobędą umiejętności pozwalające na samodzielną instalację oraz konfigurację bazy MongoDB. Zostaną zapoznani z hierarchicznym modelem danych oraz jego obsługą poprzez wbudowany w MongoDB język zapytań. Uczestnicy zdobędą umiejętności z zakresu używania **Aggregation Framework**, który pozwoli im na manipulacje na dużych zbiorach danych. Po zakończonych zajęciach słuchacze zdobędą również wiedzę pozwalająca im na rozpraszanie zbioru danych MongoDB za pomocą replikacji oraz shardingu.

PRZETWARZANIE BIG DATA Z WYKORZYSTANIEM CHMUR OBLICZENIOWYCH

- Wprowadzenie do chmur obliczeniowych
- Wprowadzenie do AWS
- Pierwsze kroki
- Podstawowe usługi
- Big Data i analityka danych
- Sztuczna Inteligencja
- Serverless
- Bazy danych
- Wyszukiwanie
- Data Warehouse & Business Intelligence
- ETL
- Integracja
- Strumienie danych
- Konteneryzacja
- Zarządzanie
- Przegląd innych rozwiązań dostępnych w chmurze publicznej

Słuchacze nauczą się implementować infrastrukturę jako kod, przetwarzać dane wsadowe i strumieniowe używając usług chmurowych Amazon Web Services. Poznają podstawowe techniki projektowania architektury z użyciem usług chmurowych na przykładzie środowiska AWS. W trakcie zajęć Słuchacze poznają takie technologie jak **AWS** (EC2, EMR, S3, Athena, Lambda, Glue, SageMaker, usługi kognitywne i AI) przeglądowo **Google, Azure**.

PRZETWARZANIE BIG DATA
SPARK

- Apache Spark
- RDD
- DataFrame
- Streaming

W ramach przedmiotu słuchacze zapoznają się z Apache Spark w sposób praktyczny i kompleksowy. Poznają problemy w rozwiązaniu których pomaga ta technologia. Uczestnicy nauczą się pracować z danymi wsadowymi i strumieniowymi. Posiądą praktyczną umiejętność przetwarzania dużych danych w sposób szybki i wydajny pisząc zwięzłe i klarowne aplikacje. W trakcie zajęć Słuchacze poznają takie technologie jak Spark (RDD, DF, streaming), Jupyter, Kafka, EMR i S3.

PRZETWARZANIE BIG DATA
APACHE HADOOP

- HDF
- Wprowadzenie do rozproszonego systemu plików
- YARN
- Apache Hive
- MapReduce oraz Tez
- Oozie

Słuchacze poznają w praktyce Hive, będą tworzyli tabele partycjonowane oraz kubełkowane, jak również będą przetwarzać rozproszone dane przy pomocy silników MapReduce oraz Tez. Słuchacze zapoznają się także z najważniejszymi poleceniami rozproszonego systemu plików Hadoop Distributed File System (HDFS), dowiedzą się czym jest YARN oraz jak używać zarządzanych przez niego zasobów oraz zdobędą umiejętności z zakresu tworzenie workflowów w Oozie.

POZYSKIWANIE DANYCH ORAZ PRZETWARZANIE STRUMIENIOWE

- Dane strumieniowe
- Ważne terminy i pojęcia
- Przegląd narzędzi używanych w przetwarzaniu danych strumieniowych
- Docker - perspektywa użytkownika i podstawowe komendy
- Apache NiFi
- Apache Kafka
- Apache Flink
- Architektury do przetwarzania danych strumieniowych

Słuchacze nauczą się budować efektywny system pobierający, przetwarzający i wprowadzający strumienie danych do systemu Big Data.

UCZENIE MASZYNOWE W ROZWIĄZANIACH BIG DATA

- Teoria i praktyka uczenia maszynowego
- Metody ewaluacji modeli uczenia maszynowego
- „Klasyczne" uczenie maszynowe jako problem uczenia z nadzorem: algorytmy naive bayes, drzewa decyzyjne, support vector machines
- Uczenie bez nadzoru, analiza skupień
- Głębokie uczenie maszynowe (deep learning): sieci rekurencyjne, sieci konwolucyjne, sieci typu transformer, modele wykorzystujące transfer learning

W ramach zajęć Słuchacze nauczą się **trenować i oceniać modele uczenia maszynowego we współczesnych środowiskach big data**. Metody będą mieć po części charakter uniwersalny, ale w ramach zajęć skupimy się przede wszystkim na problemach przetwarzania tekstów i obrazów. Przykładowe problemy, z jakimi zmierzymy się na laboratoriach to **rozpoznawanie obiektów na zdjęciach** oraz **rozpoznawanie wydźwięku tekstu** (opinion mining albo sentiment analysis).

Warunki ukończenia studiów

Warunkiem ukończenia jest zaliczenie wszystkich przedmiotów oraz napisanie i obrona indywidualnej pracy końcowej.

Sprawdź zasady rekrutacji

Celem pracy końcowej jest opracowanie, implementacja i opisanie systemu Big Data do przetwarzania danych. Wykonujący projekt powinien postawić **cel biznesowy** z użyciem systemu analizy danych, przeprowadzić **analizę postawionej hipotezy biznesowej** i opisanie jej **wyniku**.

Niezależnie od wielkości wybranego zbioru danych, analiza powinna być wykonana za pomocą **narzędzi do przetwarzania Big Data**. Zaprojektowany system może skupiać się zarówno na problemie analitycznym jak i technicznym przetwarzania danych.

Nie ma wymagań odnośnie technologii użytych w wykonaniu projektu. Muszą być to technologie Big Data, a ich wybór powinien być umotywowany. Końcowy etap analizy danych zagregowanych czy wizualizacji, może być wykonany przy pomocy narzędzi do małych danych (np Python, Pandas), ale znacząca część przetwarzania danych powinna być wykonana w technologii Big Data.

DATA SCIENCE

ALGORYTMY, NARZĘDZIA I APLIKACJE DLA PROBLEMÓW TYPU BIG DATA

198h
wykładów
i warsztatów

12 500 pln
z możliwością płatności
w dwóch ratach

w cenie kurs e-learnignowy
z Pythona

Poznaj potencjał studiów

Dlaczego warto studiować data science?

Celem studiów jest przekazanie słuchaczom praktycznych umiejętności z zakresu przetwarzania i analizy dużych danych i zapoznanie ich z rolą danych w procesach podejmowania decyzji oraz wykorzystaniem danych jako cennych aktywów strategicznych.

Na rynku pracy istnieje znaczne, ciągle rosnące zapotrzebowanie na rozumiejących dane profesjonalistów w instytucjach biznesowych, publicznych, organizacjach non-profit. Podaż profesjonalistów mogących pracować efektywnie z dużymi wolumenami danych jest ograniczona i znajduje odzwierciedlenie w rosnących pensjach inżynierów danych, badaczy danych, statystyków i analityków danych.

Dla kogo są studia Data Science?

Studia są przeznaczone dla osób chcących wykorzystywać wiedzę zawartą w dużych wolumenach danych w celu wspierania podejmowania decyzji, w szczególności dla analityków i decydentów z obszaru finansów, bankowości, ubezpieczeń, produkcji, marketingu, handlu, usług, opieki zdrowotnej, branży energetycznej, nauki i innych obszarów działalności.

Doświadczenie w pracy z technologiami **jest mile widziane** (np. podstawowa umiejętność programowania w dowolnym języku, podstawowa znajomość zagadnień związanych z bazami danych i językiem SQL), ale nie jest wymagane.

Szczegółowy program studiów

STUDIUM PRZYPADKU

Zajęcia prowadzone przez przedstawicieli środowisk biznesowych. Prezentacje w ramach przedmiotu obejmują przegląd komercyjnego wykorzystania wybranych metod z obszaru Data Science i Big Data.

ANALIZA DANYCH
PODSTAWY STATYSTYCZNE

- Statystyka opisowa
- Podstawowe rozkłady prawdopodobieństwa
- Prawa wielkich liczb i twierdzenia graniczne
- Estymacja punktowa
- Estymacja przedziałowa i testy parametryczne
- Regresja
- Korelacja
- Testowanie normalności
- Testy nieparametryczne

Celem zajęć jest przybliżenie podstaw wnioskowania statystycznego, a w szczególności budowy modelu statystycznego, estymacji punktowej i przedziałowej, teorii weryfikacji hipotez oraz metod badania zależności między cechami. Słuchacze poznają metody i narzędzia eksploracji danych. Podstawowe rozkłady prawdopodobieństwa. Metody konstrukcji estymatorów punktowych i badania ich własności. Przedziały ufności. Podstawowe testy parametryczne. Testowanie zgodności i niezależności cech. Podstawy analizy regresji.

BIG DATA
KONCEPCJE I TECHNOLOGIE

- Wprowadzenie do Big Data
- Skalowalne systemy baz danych na przykładzie Apache Cassandra
- Składowanie plików na przykładzie Hadoop File System
- Analiza danych przy użyciu Hadoop MapReduce i Apache Spark
- Zarządzane zasobami
- Harmonogramowanie i zarządzanie danymi
- Integracja
- Bezpieczeństwo

W ramach zajęć Słuchacze poznają następujące technologie: ekosystem Hadoopa, HDFS, formaty plików: text, sequence files, RC, ORC, Parquet, Key-value stores: HBase, Accumulo, Cassandra, In-memory stores: Tachyon, Ignite, Paradygmat MapReduce, Hive, Spark, Kafka.

PROGRAMOWANIE
W JĘZYKU R

- Środowisko pracy
- Typy danych
- Struktury danych
- Operatory
- Funkcje
- Instrukcja warunkowa i operator trójargumentowy
- Obliczenia iteracyjne
- Podstawy pracy z pakietem dplyr
- Podstawy pracy z grafiką w R

Przedmiot koncentruje się na zaprezentowaniu słuchaczom składni języka R, zaczynając od zagadnień podstawowych, a na średniozaawansowanych kończąc. W trakcie zajęć omawiane będą zarówno tematy ogólnoprogramistyczne, znajdujące zastosowanie w różnych językach programowania, jak również zagadnienia specyficzne dla języka R. Zajęcia poruszać będą również kluczowe kwestie związane z przetwarzaniem danych w R, przy wykorzystaniu najczęściej stosowanej do tego celu biblioteki dplyr.

PROGRAMOWANIE
W JĘZYKU PYTHON

- Środowisko pracy
- Wstęp do programowania w Python dla programistów R
- Podstawy programowania obiektowego
- Elementy przetwarzania danych z Pandas
- Elementy wizualizacji danych z Matplotlib

Przedmiot skupia się na omówieniu kluczowych, z punktu widzenia pracy w Data Science, aspektów składni języka Python. Porusza ponadto najważniejsze zagadnienia związane z przetwarzaniem oraz analizą danych, przy wykorzystaniu najpopularniejszej biblioteki Pythona przeznaczonej do tego celu, czyli Pandas.

ZAAWANSOWANA ANALITYKA SAS VIYA

- Supervised learning:
- Rodzaje modelowania ze względu na charakter zmiennej objaśnianej
- Metody selekcji zmiennych
- Techniki imputacji braków danych
- Transformacje zmiennych
- Drzewa decyzyjne
- Regresja
- Podstawy sieci neuronowych
- Klasyfikacja modeli
- Proces scoringu
- Wykorzystanie zewnętrznych modeli, np. R
- Unsupervised learning:
- Techniki klasteryzacji danych
- Profilowanie segmentów

Oprogramowanie SAS Viya, Visual Data Mining and Machine Learning - stanowi uniwersalny framework, który pozwala na łatwe wykorzystanie szeregu metod analitycznych przy budowie liniowych i nieliniowych modeli predykcyjnych oraz technik segmentacyjnych. Tworzone modele wspierają proces podejmowania decyzji z m.in. następujących obszarów: prawdopodobieństwo zajścia zdarzenia, oczekiwana wartość zdarzenia oraz szacowany czas wystąpienia zdarzenia.

METODY SZTUCZNEJ INTELIGENCJI

- Metoda automatycznego wnioskowania
- Logika rozmyta
- Konstrukcja systemów eksperckich
- Algorytmy ewolucyjne i genetyczne
- Metody przeszukiwania przestrzeni stanów
- Sieci neuronowe

Przedmiot obejmuje przegląd gałęzi sztucznej inteligencji i oferowanych przez nie metod przetwarzania dużych zbiorów danych. Działanie poszczególnych klas metod jest badane w trakcie zajęć laboratoryjnych. Słuchacze nauczą się algorytmów ewolucyjnych i genetycznych, metod przeszukiwania przestrzeni stanów, działania sieci neuronowych i konstrukcji systemów eksperckich.

DATA MINING - METODY EKSPLORACJI DANYCH

- Środowisko pracy
- Wstęp do programowania w Python dla programistów R
- Podstawy programowania obiektowego
- Elementy przetwarzania danych z Pandas
- Elementy wizualizacji danych z Matplotlib

Przedmiot obejmuje przegląd metod eksploracji danych. W szczególności są tu prezentowane metody odnajdywania zbiorów częstych i reguł asocjacyjnych, metody klasyfikacji, znajdowania wzorców sekwencyjnych i grupowania (clustering). Funkcjonowanie poszczególnych klas prezentowanych metod eksploracji jest badane w trakcie zajęć laboratoryjnych.

TEXT MINING
ODKRYWANIE WIEDZY Z TEKSTOWYCH ZBIORÓW DANYCH

- Wprowadzenie (co to jest NLP i dlaczego jest ważne)
- Wiadomości podstawowe (statystyka, teoria informacji, lingwistyka)
- Parsing i struktura języka (text corpora, słowa i zdania, tokenization, fleksja)
- Analiza statystyczna (modele dokumentów, modele języka, collocations, word sense disambiguation)
- Analiza gramatyczna (POS tagging, parsing, PCFG)
- Wyszukiwanie informacji
- Kategoryzacja i grupowanie dokumentów
- Streszczanie dokumentów
- Tłumaczenie automatyczne
- Wykrywanie słów kluczowych
- Budowanie ontologii
- Analiza dokumentów hipertekstowych (page rank, SEO)
- Profilowanie użytkowników, analiza sieci społecznościowych
- Analiza zachowania użytkowników (sentiment analysis)

Przedmiot obejmuje przegląd metod automatycznego przetwarzania danych tekstowych. W szczególności są tu prezentowane zagadnienia z zakresu parsowania i struktury języka, analizy statystycznej, analizy gramatycznej, kategoryzacji i grupowania dokumentów, automatycznego tłumaczenia, budowania ontologii, analizy dokumentów hipertekstowych.

WSTĘP DO WIZUALIZACJI DANYCH

- Najlepsze praktyki wizualizacji danych
- Najczęściej popełniane błędy w wizualizacji danych oraz aktualne trendy w Business Intelligence, między innymi Self-Service Analytics
- Najważniejsze koncepcje psychologiczne mające zastosowanie w wizualizacji danych
- Sygnały biznesowe (Outliers)
- Opowiadanie historii przy użyciu danych (Storytelling)

Przedmiot obejmuje zagadnienia dotyczące najlepszych praktyk wizualizacji danych. W trakcie zajęć omawianych jest 8 typów wizualizacji danych: (i) porównanie części do całości, (ii) analiza w czasie, (iii) analiza relacji, (iv) analiza korelacji, (v) analiza porównawcza, (vi) dystrybuanta, (vii) analiza hierarchii, (viii) analiza przepływu. Rozważane są najczęściej popełniane błędy w wizualizacji danych oraz aktualne trendy w Business Intelligence, między innymi Self-Service Analytics. Podczas zajęć duży nacisk jest kładziony na omówienie najważniejszych koncepcji psychologicznych mających zastosowanie w wizualizacji danych, takich jak: psychologia poznawcza, psychologia Gestalt, prawo Hicksa, psychologia koloru, dopasowanie wzorców, rozpoznawanie twarzy, wpływ społeczny, progresywne ujawnianie, hierarchia potrzeb Maslowa, brzytwa Ockhama, efekt von Restorff. W trakcie zajęć są wyjaśniane pojęcia sygnałów biznesowych (Outliers) oraz koncepcja opowiadania historii przy użyciu danych (Storytelling).

Słuchacze nauczą się jak w prosty i zrozumiały sposób pokazać to, co ważne i ukryć to, co nieistotne; budować dashboard tak, aby użytkownik szybko dotarł do najważniejszych informacji; pobierać odpowiednie wizualizacje w zależności od typu analizy w oparciu o zasady User Experience Design, koncepcji wizualizacji danych.

Warunki ukończenia studiów

Warunkiem ukończenia studiów jest zaliczenie wszystkich przedmiotów oraz zdanie egzaminu końcowego.

Sprawdź zasady rekrutacji

WIZUALNA ANALITYKA DANYCH

196h
wykładów
i warsztatów

12 500 pln
z możliwością płatności
w dwóch ratach

w cenie kurs e-learnignowy
z Pythona

Poznaj potencjał studiów

Korzyści wynikające z ukończenia studiów

Studia mają na celu zbudowanie kompetencji w zakresie zastosowania wizualizacji danych w analizie, skutecznej komunikacji wyników analiz oraz podejmowania decyzji biznesowych w oparciu o dane.

Po zakończonej nauce kompetencje absolwentów wzbogacą się o umiejętność:

- posługiwania się wybranymi narzędziami do wizualizacji danych i narzędziami Business Intelligence
- pracy z danymi przy użyciu programowania
- stosowania metod eksploracji danych i odróżniania podejścia opartego na statystyce i podejścia opartego na sztucznej inteligencji
- doboru rodzaju wizualizacji do danych w kontekście wymagań użytkownika końcowego
- selekcji komponentów wizualizacji danych dostosowanej do charakteru przekazywanej informacji

Dla kogo są studia WAD?

Studia podyplomowe Wizualna analityka danych skierowane są do osób pracujących w działach sprzedaży, marketingu, finansów, analiz na stanowiskach takich jak analityk danych, programista Business Intelligence, specjalista data science, statystyk, które chcą rozwijać kompetencje związane z analizą danych w obszarze pracy z danymi, wykorzystując rozwiązania wykraczające poza narzędzia typu excel. Ponadto, chcą rozwijać swoje umiejętności w obszarze prezentacji danych, poznając zasady tworzenia wizualizacji użytecznych dla odbiorców.

Szczegółowy program studiów

STUDIUM PRZYPADKU

Prezentacje w ramach przedmiotu obejmują przegląd zastosowań biznesowych wizualnej analityki danych:

1. Przykłady przejścia z analityki opartej o narzędzia tabelaryczne do analityki opartej o narzędzia wizualne w kontekście:
   - zmian w efektywności komunikacji
   - zarządzania taką transformacją.
2. Wpływ komunikowania analiz przez obrazy vs komunikowania przez tabele na procesy w organizacjach.
3. Wpływ infografiki na odbiór informacji.

PROGRAMOWANIE W JĘZYKU PYTHON

- Środowisko pracy i wprowadzenie do programowania.
- Podstawy programowania w języku Python.
- Przetwarzanie danych tabelarycznych w bibliotece pandas.
- Wczytywanie i zapis danych w formatach typowych dla programu Excel.
- Podstawowe instrukcje do wizualizacji danych.
- Połączenie z bazą danych SQL z poziomu Python.
- Zapisywanie, wczytywanie i zastosowanie zbudowanych modeli.

Słuchacze zdobędą podstawy programowania w języku Python w zakresie potrzebnym do pracy z danymi. Nauczą się jak zaimplementować w języku Python rozwiązania prostych problemów.oraz przeprowadzić typowe operacje na danych w języku Python.

ZAAWANSOWANE PRZYGOTOWANIE DANYCH DO ANALIZY BIZNESOWEJ

- Koncepcja pracy z danymi, omówienie różnych scenariuszy, praca z dużymi zbiorami danych,
- Budowa procesu na danych, podstawowe narzędzia (oczyszczanie, parsowanie, filtrowanie, grupowanie, advanced join, append union)
- Tworzenie formuł (podstawowe, multi-row, multi-field, narzędzia regex)
- Przetwarzania danych geograficznych i użycie ich w procesach
- Narzędzia developerskie, budowa makr
- Krótkie wprowadzenie do języka R
- Zaawansowana analityka statystyczna (predictive, time series, grouping), użycie r i python.

METODY ANALIZY DANYCH: EKSPLORACJA DANYCH I SZTUCZNA INTELIGENCJA

- Wprowadzenie do uczenia maszynowego i sztucznej inteligencji
- Podstawy matematyczne uczenia maszynowego
  - Operacje na macierzach i wektorach
  - Statystyka oraz prawdopodobieństwo
  - Testowanie hipotez statystycznych
- Metoda spadku gradientu
- Algorytmy uczenia z nadzorem - regresja liniowa, regresja logistyczna, sieci neuronowe, svm
- Algorytmy uczenia bez nadzoru - algorytm k-średnich
- Trendy i kierunki rozwoju uczenia maszynowego i sztucznej inteligencji

Zakres przedmiotu obejmuje wprowadzenie do uczenia maszynowego i sztucznej inteligencji. Zostaną przedstawione podstawy matematyczne uczenia maszynowego, operacje na macierzach i wektorach. Ponadto, program obejmuje zagadnienia ze statystyki oraz prawdopodobieństwa, a także testowanie hipotez statystycznych. Słuchacze poznają algorytmy uczenia z nadzorem oraz algorytmy uczenia bez nadzoru.

WIZUALNA ANALITYKA DANYCH W BIZNESIE (POWERBI® + TABLEAU®)

Program zajęć z Tableau:

- Podłączenie danych w różnych układach, porządkowanie danych, tworzenie modelu
- Praca ze źródłami big data, ekstrakty, bazy analityczne
- Podstawy Tableau desktop i tworzenia wizualizacji w biznesie, omówienie kluczowych typów wykresów i przypadków ich użycia
- Tworzenie podstawowych kalkulacji arytmetycznych, użycie agregacji i kalkulacji tabelarycznych
- Użycie technik statystycznych w analizie danych
- Analiza geoprzestrzenna, wizualizacja na mapach, geokodowanie
- Budowanie interaktywnych dashboardów, najlepsze praktyki wizualizacji na przykładach biznesowych

Program zajęć z Power BI:

- Podłączenie do danych źródłowych.
- Transformacja danych do modelu analitycznego.
- Wzbogacanie modelu raportowego o hierarchie i miary.
- Budowa raportu z użyciem wizualizacji.
- Wykorzystanie wizualizacji zbudowanych przy pomocy R i Python.
- Publikacja raportu i korzystanie z niego w organizacji.
- Raporty mobilne i scenariusze prezentacji.

W ramach zajęć prowadzonych w trybie warsztatu słuchacze zdobędą praktyczną umiejętności pracy z dwoma narzędziami do wizualizacji danych w zakresie: przygotowania danych źródłowych, ich transformacji do modelu analitycznego, wykorzystania w modelu raportowym hierarchiii i miar, budowy raportu z użyciem interaktywnych dashboardów oraz publikacji raportu.

INŻYNIERIA WIZUALIZACJI DANYCH

- Prowadzenie do wizualizacji danych w języku Python.
- Przedstawienie bibliotek do wizualizacji statycznej danych (matplotlib, seaborn, pandas)
- Tworzenie dashboardów w bibliotece bokeh oraz uruchamianie ich jako osobnej aplikacji
- Wprowadzenie do bibliotek SQL Alchemy oraz pymongo, pobieranie danych z baz SQL oraz NoSQL
- Wykorzystanie biblioteki plotly wraz z biblioteką dash do tworzenia i publikowania aktywnych dashboardów
- Cykliczne przetwarzanie danych (airflow, dash)
- Budowa kompletnego dashboardu na wybranym zbiorze danych.

Słuchacze poszerzą znajomość języka Python o jego zastosowanie w wizualizacji danych. Poznają szereg bibliotek wykorzystywanych w tym obszarze i nauczą się tworzenia z ich użyciem dashboardów. Na zajęciach w ramach warsztatu słuchacze zbudują kompletny dashboard na wybranym zbiorze danych.

WSTĘP DO WIZUALIZACJI DANYCH

- Historia wizualizacji danych.
- Teoria wizualizacji danych.
- Procesy i standardy wizualizacji danych.
- Nowoczesne techniki wizualizacji danych.
- Wizualizacja danych w procesie decyzyjnym.
- Percepcja i kolor w wizualizacji danych.
- UX (user experience).

Słuchacze poznają teorię wizualizacji danych, procesy i standardy z nią związane oraz rolę, jaką odgrywa w procesie decyzyjnym. Nauczą się korzystania z nowoczesnych technik wizualizacji danych i ich doboru w kontekście wymagań użytkownika końcowego. Zakres przedmiotu obejmuje także zagadnienia związane z user experience.

TECHNIKI EFEKTYWNEJ KOMUNIKACJI WYNIKÓW

- Skuteczna komunikacja.
- Wystąpienia publiczne.
- Ludzka percepcja i wzorce przetwarzania informacji.
- Zasady estetyczne projektowania.
- Storytelling:
  - filary Data Storytellingu
  - cele, strategie, etyka
  - techniki data storytellingu.

Słuchacze rozwiną umiejętności w zakresie komunikowania wyników analiz. W szczególności poznają zasady skutecznej komunikacji oraz prowadzenia narracji w oparciu o kontekst oraz odbiorców stosując odpowiednią strategię storytellingu. Zdobędą wiedzę o ludzkiej percepcji i różnych wzorcach przetwarzania informacji. Będą potrafili przygotować prezentację w oparciu o najnowsze trendy w projektowaniu. Poznają techniki storytellingu, w szczególności techniki data storytellingu.

Warunki ukończenia studiów

Ukończenie studiów następuje po przygotowaniu i obronie pracy końcowej.

Obrona pracy końcowej składa się z krótkiej prezentacji pracy oraz odpowiedzi na dwa pytania dotyczące obszaru tematycznego pracy końcowej.

Sprawdź zasady rekrutacji

FAQ

Jaka jest różnica między ścieżkami Data Science, Big Data i WIzualna Analityka Danych?

Wszystkie kierunki przygotowują do pracy z danymi, przy czym różnią się rozłożeniem akcentów.

Na ścieżce Data Science nacisk położony jest na analizę danych i jej metody. Na ścieżce Big Data uczymy pracy z dużymi danymi od etapu ich pozyskania i przygotowania do momentu dalszej pracy analitycznej. Natomiast na kierunku Wizualna Analityka Danych rozwijane są umiejętności związane z przygotowaniem danych, ale ścieżka pracy z nimi kończy się na etapie przygotowania komunikacji z nimi związanej, czyli odpowiedniej ich wizualizacji i przestawienia zgodnie z najnowszymi trendami opartymi na storytellingu.

Czy poradzę sobie, jeżeli ukończyłem/łam studia nie techniczne, np. ekonomiczne?

Każde dodatkowe kompetencje techniczne to łatwiejsza droga przez studia. Ale to nie poziom Twoich umiejętności na wstępie decyduje o sukcesie, o tym, jak sobie na nich poradzisz. Więcej zależy od tego, czy nie boisz się wyzwań i jesteś zdeterminowany.

Czy program studiów został jakoś ograniczony ze względu na tryb zdalny?

Materiał realizowany w ramach studiów pozostaje taki sam, niezależnie od tego, czy jest to tryb nauki zdalnej czy stacjonarnej. Zakładane kompetencje absolwenta każdej edycji są takie same.

Czy jak skończę studia to znajdę pracę? Czy znamy dalsze losy absolwentów?

Znajomość programowania nie jest wymagana, na studiach przewidziany jest przedmiot Programowanie zorientowane na dane, w ramach którego prowadzona jest nauka programowania w Pythonie w zakresie potrzebnym do pracy z danymi. Niemniej, wcześniejsza znajomość programowania w dowolnym języku na pewno ułatwia wejście w tematykę pracy z danymi.

Czy studia kończą się egzaminem czy projektem?

Studia na ścieżce Data Science kończą się egzaminem. Studia na ścieżce Big Data kończą się projektem. Najczęstsze obszary, których dotyczą prace końcowe to: system do zbierania i agregowania danych w czasie rzeczywistym, implementacja ETL/ELT na klastrze, pozyskanie, czyszczenie, eksploracja i wizualizacja danych. Prace są konsultowane w ramach przedmiotu seminaryjnego: Projektowanie rozwiązań Big Data. Studia na ścieżce Wizualna Analityka Danych kończą się projektem wymagającym zaprezentowania kompetencji z obszaru wykorzystania narzędzi i technik wizualizacyjnych.

Czy na ścieżce Big Data wymagana jest znajomość programowania?

Uczelnia nie prowadzi tego rodzaju statystyk, ale nasi wykładowcy spotykali często na konferencjach branżowych swoich byłych studentów, którzy pracują już z danymi. Natomiast wszystkie publikacje dotyczące prognoz nt rynku pracy wskazują na rozwój technologii AI i Big Data (w tym raporty Deloitte, Forbes, Gartner).

Jaki jest zakres tematów związanych z architekturą i wdrożeniami na studiach Big Data?

Jeżeli przyjmiemy, że architektura to układanie klocków tak, żeby wyjściowy system był możliwie najlepszy, ważne jest w pierwszym kroku poznanie tych “klocków”. Temu służą wszystkie zajęcia, pokazują rozwiązania i ich zastosowanie. Tematyka architektury pojawia się w szczególności na przedmiotach Projektowanie rozwiązań Big Data oraz Przetwarzanie Big Data z wykorzystaniem chmur obliczeniowych.

Na jakim poziomie zaawansowania jest absolwent studiów (mid czy junior)?

Studia stwarzają warunki do zdobycia wiedzy i kompetencji wymaganych na stanowisku samodzielnego specjalisty, niemniej osiągnięcie tego poziomu zależy od zaangażowania uczestnika i wkładu jego pracy własnej w pełne opanowanie programu. Zależy to także od wcześniejszego doświadczenia zawodowego.

Sprawdź zasady rekrutacji na ds.ii.pw.edu.pl

Wydział Elektroniki i Technik Informacyjnych
POLITECHNIKA WARSZAWSKA

sages