

# Big Data, inżynieria i analiza danych z wykorzystaniem języka Python

STUDIA PODYPLOMOWE



**9**

**176**

**11**

**2**

Liczba miesięcy nauki Liczba godzin zajęć Liczba zjazdów Liczba semestrów

## **Program studiów**

### **Program studiów**

#### **Programowanie w języku Python (24 godz.)**

- Podstawowe koncepcje
- Pandas, numpy, statystyka w Pythonie
- Systemy kontroli wersji
- Podstawy testowania
- Analiza porównawcza algorytmów na podstawie złożoności obliczeniowej

#### **Formaty danych (8 godz.)**

- Formaty danych: csv, json, avro, parquet, xml

#### **Programowanie obiektowe w języku Python (16 godz.)**

- Atrybuty, klasy, konstruktor
- Metody, dziedziczenie, „metody magiczne”

#### **Orkiestratory (8 godz.)**

- Cron
- Airflow

#### **Procesy CICD (8 godz.)**

- Github Actions lub AirFlow

#### **Apache Kafka (8 godz.)**

- Interfejs Apache Kafka – szybki start do strumieniowego przetwarzania danych

#### **NoSQL (Microsoft Azure) (16 godz.)**

- Podstawowe koncepcje baz NoSQL - HBase, Cassandra, Impala, Neo4j



## **Zaawansowane bazy danych i hurtowanie danych (32 godz.)**

- Podstawowe oraz zaawansowane aspekty języka SQL
- Koncepcje modelowania hurtowni danych (ROLAP, MOLAP, HOLAP)
- Technologie ETL/ELT
- Elementy prezentacji danych np. Power BI

## **Narzędzia Big Data (Microsoft Azure) (24 godz.)**

- Apache Hadoop & Apache Spark

## **Wprowadzenie do sztucznej inteligencji (16 godz.)**

- Podstawowe pojęcia i definicje
- Zadania regresji, klasyfikacji, detekcji, klasteryzacji i optymalizacji
- Uczenie nadzorowane i nienadzorowane
- Atrybuty danych, ich typy i właściwości
- Zbiory danych (uczący, testujący, walidacja, etc.)
- Metody klasyfikacji, klasteryzacji i estymacji

## **Projekt zaliczeniowy (16 godz.)**

- Seminarium projektowe

## **Forma zaliczenia**

- Egzamin końcowy
- Praca projektowa pisana w grupach