

kod szkolenia: PYTH_ANALIZA_PODST / PL AA 3d

Analiza danych w Pythonie – poziom podstawowy

Szkolenie: Podstawy analizy danych Python - Poziom podstawowy



Przeznaczenie szkolenia

Szkolenie przeznaczone jest dla osób chcących poznać podstawowe metody **analizy danych** w języku **Python**.



Korzyści wynikające z ukończenia szkolenia

Uczestnicy szkolenia zdobędą, ugruntują lub wzbogacą swoją wiedzę z zakresu analizy danych. Zapoznają się z podstawami środowiska **PyCharm** oraz dwoma głównymi pakietami do operacji na danych – **NumPy** oraz **pandas**. Słuchacze dowiedzą się, jak przygotować dane, aby były gotowe do dalszego **modelowania** z użyciem **uczenia maszynowego** oraz poznają podstawowe metody ich wizualizacji. Kontynuację szkolenia stanowi kurs „**Analiza danych w Pythonie – poziom średniozaawansowany**”.



Oczekiwane przygotowanie słuchaczy

Python na poziomie podstawowym



Język szkolenia

- Szkolenie: polski



Czas trwania

3 dni / 21 godzin

Agenda szkolenia

1. Wprowadzenie

- Podstawowe statystyki
- Typy **danych**
- Problemy z danymi
- Tworzenie nowego projektu w środowisku **PyCharm**
- Import bibliotek
- Przegląd podstawowej funkcjonalności środowiska

2. Podstawy NumPy

- Wektory, macierze, tensory
- Generowanie liczb pseudolosowych
- Tworzenie tablic
- Właściwości tablicy, **konwersja typów**
- **Działania na indeksach**
- Operacje arytmetyczne i **logiczne** na tablicach
- Zmiana rozmiaru i łączenie **tablic**
- **Agregacja danych**
- Płytkie i **głębokie kopiowanie**
- Wbudowane funkcje
- Sortowanie **tablic**

3. Podstawy pandas

- Seria i ramka danych
- Indeksy i kolumny
- **Tworzenie obiektów**
- Import danych (excel, csv)
- Zapis danych
- **Atrybuty ramki**
- Dodawanie i usuwanie wierszy
- **Indeksery**

- Filtracja danych
- Operacje **na ramce**
- Grupowanie i **agregacja**

4. **Analiza eksploracyjna** i przygotowanie danych

- Typy i rozmiar danych
- Statystyki
- Kodowanie danych
- Rodzaje **braków danych**
- Wykrywanie braków **danych**
- Podstawowe metody uzupełniania braków danych
- Niespójności w **danych**
- Wartości odstające
- Korelacja danych
- Standaryzacja i **skalowanie**

5. Wizualizacja danych w pakiecie Matplotlib

- Konfiguracja wykresów
- Podwykresy
- Wykres liniowy
- **Wykres rozrzutu**
- Histogram
- Wykres słupkowy
- Diagram kołowy
- Heatmap
- Wyświetlanie obrazów