**Szkolenie pn. „Data Mining – kurs podstawowy”**

**realizowane dla pracowników dydaktycznych i badawczo-dydaktycznych**

**Wydziału Nauk Przyrodniczych**

**Cele szkolenia**

Celem kursu jest zapoznanie się z podstawowymi ideami i zasadami Data Mining (zgłębiania danych), takimi jak:

* dostęp do baz danych
* przygotowanie danych do analiz
* dobór odpowiedniej metody analitycznej
* ocena uzyskanego modelu
* oraz wykonywanie analiz i stosowanie uzyskanych modeli (deployment)

dla samodzielnego realizowania projektów Data Mining.

**Efekty kształcenia i korzyści dla uczestnika**

Na szkoleniu uczestnik zdobędzie podstawową wiedzę na temat wybranych technik Data Mining. Wiedza dotyczyć będzie rozwiązań dostępnych w środowisku Statistica. Po ukończeniu kursu uczestnik będzie poprawnie wykorzystywać środowisko Statistica do realizacji projektów Data Mining wybranymi metodami.

**Forma realizacji**

Szkolenie dla 8-osobowej grupy realizowane w formie warsztatów wspartych mini wykładem (16h). W trakcie szkolenia wykorzystane zostaną co najmniej następujące metody dydaktyczne: wykład, prezentacja z praktycznymi przykładami, case studies, dyskusja, wymiana doświadczeń między uczestnikami, ćwiczenia praktyczne. Ćwiczenia zostaną przeprowadzone w programie Statistica. Miejsce szkolenia: sala komputerowa B-204, w Instytucie Biologii, Biotechnologii i Ochrony Środowiska WNP UŚ przy ul. Jagiellońskiej 28 w Katowicach.

**Program szkolenia**

1. Podstawowe idee Data Mining
   * Modele Data Mining
   * Rodzaje zadań Data Mining
   * Przeuczenie i podział na próby
2. Środowisko Statistica Data Miner
   * Wprowadzenie
   * Dane wejściowe
   * Przeglądarka węzłów
   * Przykład prostej analizy w systemie Statistica Data Miner
3. Specjalistyczne moduły systemu Statistica Data Miner (przegląd)
   * Wstępne przetwarzanie danych
   * Metody predykcyjnego Data Mining (uczenie z nauczycielem)
   * Odkrywanie wiedzy (uczenie bez nauczyciela)
   * Moduły stosowane po uzyskaniu modelu
4. Wprowadzenie do wybranych metod Data Mining
   * Regresja logistyczna
   * Drzewa klasyfikacyjne i regresyjne
5. Przykłady analiz w systemie Statistica Data Miner
   * Wstępna obróbka danych – czyszczenie i przekształcenia
   * Problem klasyfikacyjny – modelowanie zdolności kredytowej
   * Zadanie regresyjne
   * Analiza skupień (segmentacja)

**Informacje o wykonawcy i trenerce realizującej szkolenie**

Organizator: **Instytut Innowacyjnej Edukacji**

Prowadząca: **Natalia Apolinarska**

Trener szkoleń w obszarze Data Mining i Data Science. Absolwent matematyki na Politechnice Wrocławskiej. Ponad 6 lat zawodowo związana z uczeniem maszynowym oraz sztuczną inteligencją. Na co dzień realizuje projekty z obszaru zaawansowanej analizy danych, data science. Od kilku lat prowadzi szkolenia w obszarze statystyki, Data Mining, Data Science z wykorzystaniem popularnych narzędzi. Specjalistka z zakresu statystyki, Data Mining, Data Science, Python, R Project, Statistica, KNIME.

**Harmonogram szkolenia w terminie: 13 i 14 czerwca 2022 r.**

***Dzień 1***

|  |  |
| --- | --- |
| ***Godziny zajęć*** | ***Zakres tematyczny*** |
| *8.30 – 8.45* | Test sprawdzający (pre test) |
| * 1. *– 10.45* | * *Wstęp do Data Mining* * *Metodyki Data Mining* * *Narzędzia wykorzystywane w Data Mining* * *Organizacja środowiska Statistica Data Miner w którym będziemy przeprowadzać praktyczne ćwiczenia eksploracji danych* |
| *10.45 – 11.00* | *Przerwa* |
| *11.00 – 13.00* | * *Integracja i wybór danych oraz konstrukcja targetu* * *Czyszczenie danych* * *Tworzenie danych pochodnych* * *Wstępna analiza danych* |
| *13.00 – 13.15* | *Przerwa* |
| * 1. *– 15.00* | * *Główne zadania Data Mining oraz rodzaje metod* * *Podział próby do modelowania* |

***Dzień 2***

|  |  |
| --- | --- |
| ***Godziny zajęć*** | ***Zakres tematyczny*** |
| * 1. – 10.45 | * *Metody w Data Mining - opis* |
| 10.45 – 11.00 | Przerwa |
| 11.00 – 13.00 | * *Metody nadzorowane w Data Mining - modelowanie* * *Metody nienadzorowane w Data Mining - modelowanie* |
| 13.00 – 13.15 | Przerwa |
| 13.15 – 14.45 | * *Poznanie miar do oceny modelu* * *Ocena modeli* |
| 14.45 – 15.00 | PODSUMOWANIE, DYSKUSJA, PYTANIA I ODPOWIEDZI.  Test sprawdzający (post test), Ankieta podsumowująca szkolenie |