

## Sztuczna inteligencja

### Analiza dużych zbiorów danych metodą k-średnich

Przebieg ćwiczeń laboratoryjnych i warunki zaliczenia przedmiotu  
semestr V rok akademicki 2021/22

#### Ćwiczenia laboratoryjne

- na studiach dziennych 10 zajęć \* 3 godziny = 30 godzin,

Przedmiotem ćwiczeń laboratoryjnych jest analiza dużych zbiorów danych **metodą k-średnich**.

Studenci otrzymują trzy moduły, autorstwa prowadzącego przedmiot, zawierające definicje funkcji niezbędnych do realizacji algorytmu metodą **k-średnich** zapisane w języku **Python 3.9**. Oprócz tego: duży zbiór danych **TopBabyNamesbyState** w postaci 10 500 pięcio-atrybutowych rekordów zapisanych w formacie **.csv**.

Studenci pracują w jedno- lub dwuosobowych zespołach.

Ćwiczenia laboratoryjne przebiegają według niżej podanego schematu:

#### 1. Zajęcia nr 1 i 2:

Podział na zespoły. Zapoznanie się szczegółowo z przebiegiem ćwiczeń laboratoryjnych, oraz budową i własnościami dostarczonego programu w języku **Python 3.9**.

#### 2. Zajęcia nr 3 i 4 :

Badania w zespołach własności dostarczonej przykładowej bazy danych za pomocą dostarczonego oprogramowania. Będzie to przede wszystkim **analiza wpływu wybranej ilości klastrów na możliwości wydobywania wiedzy z poszczególnych klastrów**. Odbywać się to będzie poprzez pokazanie w wersji elektronicznej i dyskusja z prowadzącym zajęcia otrzymanych wyników. Wymagane będzie:

- **dobrze rozumienie działania algorytmu k-średnich,**
- **rozumienie własności, struktury i wzajemnych zależności między atrybutami dostarczonego zbioru danych,**
- **umiejętność zaplanowania i przeprowadzenia odpowiednich badań, jak również**
- **umiejętność prezentacji i interpretacji osiągniętych wyników.**

Za wykonania tego zadania ( etap I) student może otrzymać **10-20 punktów** (jeśli zaliczy I-szy etap), lub 0 punktów w przypadku nie zaliczenia. Ocena będzie indywidualna, mimo ewentualnej pracy w zespole.

#### 3. Zajęcia nr 5 - 9:

**Przed zajęciami nr 6** poszczególne zespoły mają **obowiązek pozyskania z Internetu** do dalszych badań innej, odpowiednio dużej bazy danych opisującej realny obszar rzeczywistości.

Do zadań studentów należeć będzie (w kontakcie z prowadzącym przedmiot):

- **przeprowadzenie przed rozpoczęciem badań normalizacji danych** (normalizacja wymaga zatwierdzenia przez prowadzącego zajęcia),

- przeprowadzenie badań wpływu wyboru liczby klastrów na efektywność algorytmu z wyborem optymalnej liczby klastrów,
- przeprowadzenie badań i analiza wpływu metody liczenia odległości (euklidesowa, Manhattan) na efektywność algorytmu.

Do zaliczenia należy przygotować sprawozdanie końcowe w wersji papierowej. Konieczna będzie praca w domu. Za wykonania etapu drugiego (opisanego w pk. 3) student może otrzymać **20-70 punktów**.

#### 4. Zajęcia nr 10:

Zajęcia przeznaczone są na ocenę i dyskusję osiągniętych przez poszczególne zespoły wyników, oraz końcową ocenę indywidualną studentów.

Promowana będzie: jakość i terminowość realizacji poszczególnych etapów projektu, organizacja i współdziałanie poszczególnych członków w zespole, wartość osiągniętych rezultatów, przy czym brany będzie pod uwagę indywidualny wkład każdego z członków zespołu. Wcześniejsze zaliczenie etapu drugiego promowane będzie dodatkowymi punktami.

Przedmiotem oceny będzie przede wszystkim: **umiejętność dostosowania dostarczonego oprogramowania do własności nowej bazy danych**, ale też - sposób prezentacji osiągniętych wyników w sprawozdaniu końcowym, dokumentującym osiągnięte wyniki.

Zawartość sprawozdania z laboratorium "Sztucznej inteligencji" na temat : **"Inteligentna analiza dużych zbiorów danych"**:

- strona tytułowa,
- zbiór danych **etapu drugiego**,
- moduły kodów programu w języku *Python* dostosowane do aktualnego zbioru danych,
- wyniki badań **wpływu liczby klastrów**, oraz **wyboru metody liczenia odległości** między rekordami danych (**odległość euklidesowa, odległość Manhattan**) na **efektywność wydobywania wiedzy ze zbioru danych**.

#### Warunki zaliczenia laboratorium

Za ćwiczenia laboratoryjne student może uzyskać na zaliczeniu maksymalnie 90 punktów. Ćwiczenia uznaje się za zaliczone, jeśli student uzyska **nie mniej niż 30 punktów**.

#### Tabela ocen końcowych:

**30 – 45 pk. ocena 3.0**

**46 – 50 pk. ocena 3.5**

**51 – 65 pk. ocena 4.0**

**66 – 80 pk. ocena 4.5**

**81 – 90 pk. ocena 5.0**

Wyróżniające się prace będą oceniane na **6.0**

Ostatni dzień zajęć laboratoryjnych, wskazany w harmonogramie zajęć, jest ostatecznym terminem zaliczenia ćwiczeń.

**Uwaga:** Celem dobrego przygotowania teoretycznego do ćwiczeń laboratoryjnych **na pierwszych wykładach** omawiane będą w ujęciu praktycznym zagadnienia ściśle wiążące się z tematyką ćwiczeń. **Dlatego też wykłady te są obowiązkowe.**

opr. dr Feliks Kurp