**BJ: Zadania do Lab 5** (+ zadania ogólne do Lab 6, 7)

***Cele Lab-5***: Implementacja w Orange:

G5.1 Algorytmy **Random Forest** (**RF**) dla problemów klasyfikacji;

G5.2 Algorytmy **klastrowania**

G5.3 **Analiza regresji**: Liniowa, Wielomianowa (jako liniowa ze zmodyfikowanymi kolumnami), Nieliniowa

**Zadania do Lab-5:**

1. Przeczytaj pliki (od BJ) dotyczący klastrowania, RF i analizy regresji w Orange;

2. Zaimplementuj RF dla problemu klasyfikacji (np. z Iris Dataset);

3. Porównaj wyniki RF/różne wskaźniki z innymi metodami, np. NNs, Rank, PCA, SVM.

*(porównaj wyniki zbliżone do „najlepszej”/dobrej klasyfikacji dla każdej Metody)*

4. Zaimplementuj kilka algorytmów klastrowania (k-średnich, DBSCAN, Louvain Cl) i porównaj wyniki.

[+ Wygeneruj (w Pythonie) własny zestaw danych do klastrowania, pobierz i zaimplementuj go w Orange za pomocą widżetu pliku].

5. Analiza regresji: Liniowa, Wielomianowa (jako Liniowa ze zmodyfikowanymi kolumnami) wraz z innymi widżetami (Rank, PCA, RF), Implementacja regresji nieliniowej za pomocą widżetów NN i SVM. Porównaj wyniki.

6. *Przygotuj raport* (*krótki* z wnioskami)

**7. Dodatkowa praca domowa** *do późniejszej (efektywnej) realizacji w ramach* ***Lab-6***

7.1 Pobierz 10-20 (100) zdjęć dla każdej klasy/kategorii, np. klasa kotów i klasa psów.

[mały rozmiar i jednorodność do zadania klasyfikacji obrazu; np. 10 zdjęć psów i 10 zdjęć kotów, jeśli masz mocny komputer/laptop, możesz pobrać większą liczbę zdjęć i większą liczbę klas obiektów]. Np. od

<https://pixabay.com/images/search/pet/>

7.2 Pobierz zbiór danych MNIST w formacie CSV, np. pliki mnist\_train.csv, mnist\_test.csv:

<https://www.kaggle.com/datasets/oddrationale/mnist-in-csv?resource=download>

Aby efektywnie pracować w Orange (dla każdego, szczególnie dla posiadaczy laptopów o niskim poborze mocy) ***polecam zrobić*** (w Pythonie) podpróbki z każdego zbioru danych, np. pierwszy/ostatni/random 3000-5000 wierszy dla mnist\_ train i 500- 1000 dla plików mnist\_test.

**8. Schemat Laboratorium nr 5:**

0. BJ: Przeglądanie pracy domowej nr 4 każdego ucznia

1. Zaleca się zapoznanie się z plikami 5\_BJ\_PUM\_Lab-5.... .

2. Realizacja zadań 2-6, Lab. 5.

3. BJ: ostatnie 10-15 minut zajęć: Analiza wykonania przez studentów zadań w Lab. 5.

**Oceny lekcji**: 5 – wszystko zostało zrobione, 4 – częściowo, 2 – nic nie zostało zrobione.

(Na następnej lekcji: w wyniku wykonania zadania w domu, 2 można poprawić na 3, 4 - na 4.5).

**Ogólny plan pracy Laboratorium 6 i 7**

***Laboratorium-6:***

6.1 Klasyfikacja obrazów w Orange

6.2 Klasyfikacja zbioru danych MNIST w Orange

***Lab*-7:**

**7.1 *Prezentacje ustne studenta +*** *analiza wykonania pracy domowej (dla* *pozostałych)* ***:***

:= Plusy i minusy Orange w porównaniu do Pythona na przykładzie:

-- analizy obrazu,

-- analizy zbioru danych MNIST,

-- regresji liniowej i nieliniowej,

*-- lub innych zadań wybranych przez studenta (np. Analiza ciejawego case-study)*

*Prezentacja tablicowa z analizą i wnioskami*, 6-10 minut (*zależy od ilości prezentacji*).

7.2 *Wystawianie ocen końcowych* (za zajęcia laboratoryjne):

7.2.1 Ocena 5 dla tych, *którzy dokonali ciekawej prezentacji*;

7.2.2 Dla *pozostałych*: Rozmowa z każdym po zaliczeniu laboratoriów 1-4, a zwłaszcza 5-6

(będę oglądał wyniki i raport dla każdego laboratorium).

[***According to Prof. W. Dzwinel***, **Final grade = 0.7\*grade(Lab) + 0.3\*grade (lecture-assignments)**]