

Zadania do ćwiczeń laboratoryjnych:

1. Dla zbioru danych dotyczących cukrzycy (*diabetes*)
 1. otwórz zbiór,
 2. wypisz nazwy kolumn,
 3. policz ile przebadanych pacjentów opisano w zbiorze,
 4. narysuj histogram dla badanej zmiennej
2. Dla zbioru danych dotyczącego kwiatków (*iris*)
 1. otwórz zbiór,
 2. wypisz nazwy kolumn,
 3. policz i wypisz nazwy badanych gatunków kwiatków,
 4. policz i wypisz ile przypadków jest w zbiorze dla poszczególnych gatunków,
3. wygeneruj zbiór dla binarnej klasyfikacji dla 5 kolumn i 500 przypadków,
4. wygeneruj zbiór dla 5 klasowej klasyfikacji dla 10 kolumn i 1000 przypadków,
5. wygeneruj zbiór dla regresji dla 400 przypadków i 4 kolumn,
6. wygeneruj zbiór dla regresji dla 600 przypadków i 6 kolumn.
7. zbiór z zadania 1 podziel losowo na treningowy (250 pacjentów) i walidacyjny (reszta)
8. zbiór z zadania 2 podziel losowo na treningowy (70% przypadków) i walidacyjny (reszta), w powtarzalny sposób (*random_state*)
9. zbiór z zadania 3 podziel losowo na treningowy (50% przypadków) i walidacyjny (reszta), tak aby obie klasy były równo reprezentowane w podzbiorach (stratyfikacja). Wyrysuj histogram dla obu klas z obu zbiorów podzbiorów.
10. zbiór z zadania 6 podziel losowo na walidacyjny (100 przypadków) i treningowy (wszystkie pozostałe), tak aby modelowana zmienna była równo reprezentowana w podzbiorach (stratyfikacja). Wyrysuj histogram dla zmiennej modelowanej z obu podzbiorów.

- scikit-learn.org/stable/modules/generated/sklearn.datasets.load_diabetes.html
- scikit-learn.org/stable/modules/generated/sklearn.datasets.load_iris.html
- pl.wikipedia.org/wiki/Kosaciec
- scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html