

Wymagania projektu

Projekt polega na stworzeniu modułu, który może być zbiorem funkcji i/lub obiektów z metodami i ta druga metoda wydaje się lepszym rozwiązaniem. Moduł będzie operował na wybranym przez studenta datasetcie pobranym ręcznie przez studenta na dysk. Zbiory danych można znaleźć pod adresem <https://archive.ics.uci.edu/ml/index.php>. Moduł powinien umożliwiać jego zaimportowanie w interaktywnej konsoli i uruchomienie jego funkcji z wiersza poleceń.

Funkcjonalność, którą ma dostarczać moduł:

- **Wczytanie datasetu** – funkcja, która po podaniu ścieżki (nazwa pliku, jeżeli w tym samym folderze) wczyta dane z pliku do listy (można użyć modułu csv). Dodatkowo funkcja przyjmuje parametr określający czy pierwszy wiersz pliku zawiera etykiety kolumn czy nie. Jeżeli tak to etykiety wczytywane są do oddzielnej listy.
- **Wypisanie etykiet** – funkcja wypisująca etykiety lub komunikat, że etykiet nie było w danym datasetcie.
- **Wypisanie danych datasetu** – funkcja wypisuje kolejne wiersze datasetu. Bez podania parametrów wypisywany jest cały dataset, ale możliwe też podanie 2 parametrów, które określają przedział, który ma zostać wyświetlony (na wzór slice),
- **Podział datasetu na zbiór treningowy, testowy i walidacyjny**. Funkcja przyjmuje 3 parametry określające procentowo jaka część głównego zbioru danych trafia do poszczególnych zbiorów.
- **Wypisz liczbę klas decyzyjnych** – wypisanie krotek gdzie pierwsza wartość to wartość klasy (np. nazwa irysa, dla binarnych 0 lub 1 itd.), a druga to liczebność (kardynalność) tej klasy.
- **Wypisz dane dla podanej wartości klasy decyzyjnej** – wypisuje wiersze z zadaną wartością klasy decyzyjnej.
- **Zapisanie danych do pliku csv** – jako parametr przyjmowana jest dowolna lista, która może być podzbiorem datasetu, zmienną przechowującą dane treningowe, itp. Dodatkowo podawana jest nazwa pliku, do którego dane zostaną zapisane.