

Wydział Elektroniki i Technik Informacyjnych
Politechnika Warszawska

Wprowadzenie do sztucznej inteligencji

Sprawozdanie z ćwiczenia nr 7

Tymon Kobylecki

Warszawa, 2022

Spis treści

1. Wstęp	2
2. Ćwiczenie	3
2.1. Środowisko - dane o kosaćcach	3
2.2. Eksperymenty	3
2.3. Wyniki	3
2.4. Analiza wyników	5
2.5. Wnioski	5

1. Wstęp

W niniejszym sprawozdaniu opisane zostało rozwiązanie zadania oraz eksperymenty dotyczące zadania nr 6 polegającego na implementacji naiwnego klasyfikatora Bayesa. Algorytm ten został zastosowany do klasyfikacji odmian kosaćców na podstawie parametrów takich jak długość działki kielicha czy płatków. Dane zostały pobrane z <https://archive.ics.uci.edu/ml/datasets/iris> za pomocą pakietu `sklearn`.

2. Ćwiczenie

2.1. Środowisko - dane o kosaćcach

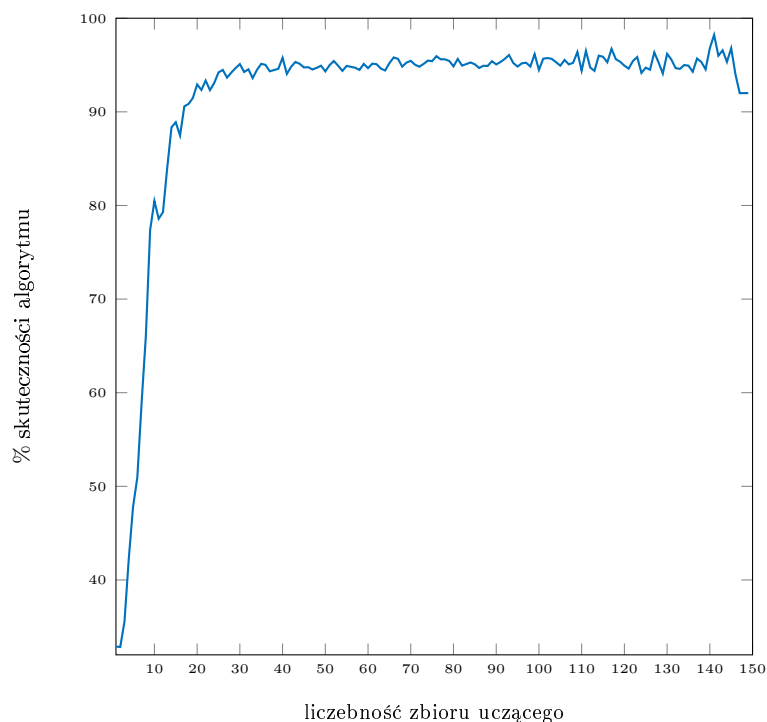
Dostarczone dane dotyczyły wymiarów działek kielichów oraz płatków 150 kosaćców z trzech odmian. Klasą była więc odmiana kosaćca przedstawiona w formie liczbowej (0 - kosaciec szczeniowy, 1 - kosaciec różnobarwny, 2 - kosaciec wirginijski). Podczas wszelkich eksperymentów dane podzielone były na zbiór uczący oraz testowy.

2.2. Eksperymenty

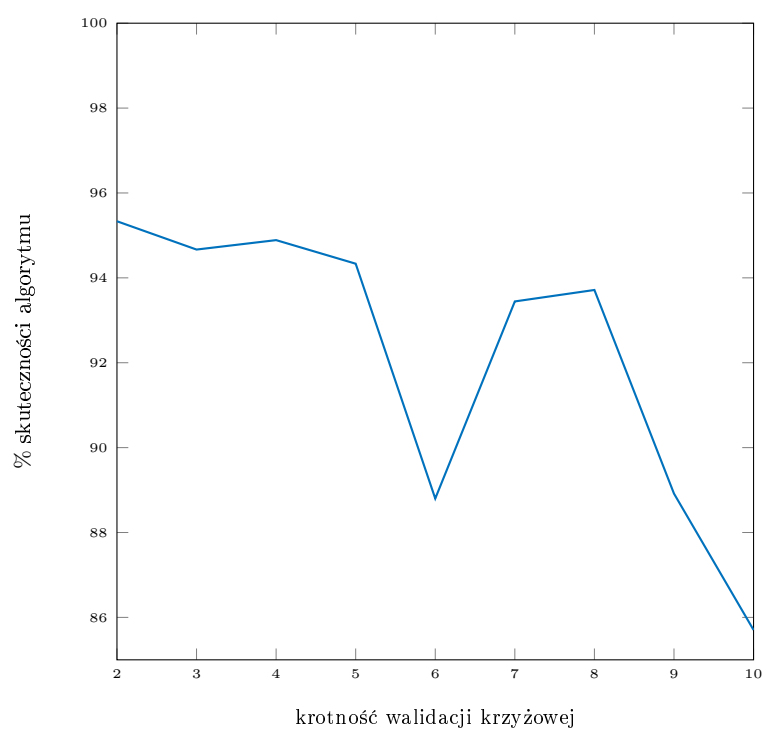
Jedynym parametrem poddanym eksperymentom była wielkość zbioru testowego, zmieniana w zakresie od 1 do 149, co oznaczało zmianę liczebności zbioru testowego odpowiednio od 149 do 1. W przypadku walidacji krzyżowej modyfikowana była krotność walidacji, w zakresie od 2 do 10 podzbiorów.

2.3. Wyniki

Na wykresie 2.1 widoczna jest skuteczność algorytmu w zależności od liczebności zbioru uczącego, zaś na wykresie 2.2 przedstawiono wyniki walidacji krzyżowej dla różnej wartości krotności.



Rys. 2.1. Wykres przedstawiający uśrednione wyniki dla wszystkich wielkości zbioru uczącego



Rys. 2.2. Wykres przedstawiający uśrednione wyniki dla wszystkich krotności walidacji krzyżowej

2.4. Analiza wyników

Łatwo zauważyć, że przy podziale na zbiór uczący i testowy, nie warto zwiększać liczebności zbioru uczącego powyżej 40, gdyż dalsze wyniki nie są istotnie odmienne. W przypadku podziału przy pomocy k-krotnej walidacji krzyżowej również widać tendencję spadkową, gdy zwiększany jest parametr k (a zarazem zmniejszany jest zbiór uczący). Wynika to w prosty sposób z niedouczenia algorytmu, który może czasem nawet „nie poznać” pewnej klasy w swoim zbiorze uczącym.

2.5. Wnioski

Po zastosowaniu odpowiedniej liczebności zbioru uczącego algorytm bardzo dobrze rozwiązuje postawione zadanie, osiągając skuteczność niekiedy przekraczającą 95%. Warto zauważyć, że są to wyniki istotnie lepsze, niż w przypadku zastosowania algorytmu ID3 w ćwiczeniu 4, który osiągał skuteczność na poziomie 70-80%. Należy jednak zaznaczyć, że nie działały one na tym samym zbiorze danych.