

WSI21Z Piotr Szmurło (303785)

Ćw 4. (7 pkt), data oddania: do 06.12.2021 - Regresja i klasyfikacja

Zaimplementować naiwny klasyfikator Bayesa (Gaussowski).

Do eksperymentów wykorzystać zbiór danych dot. jakości wina (białe).

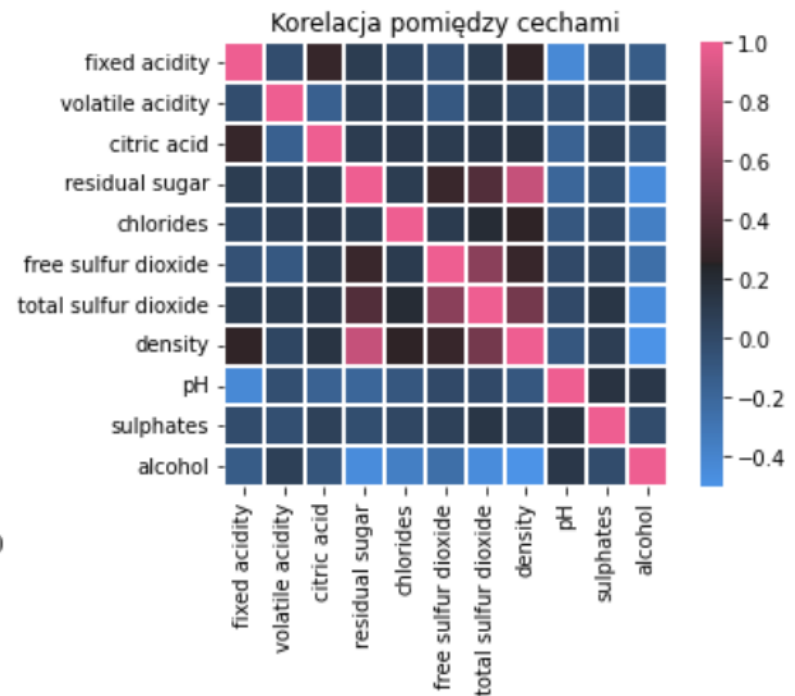
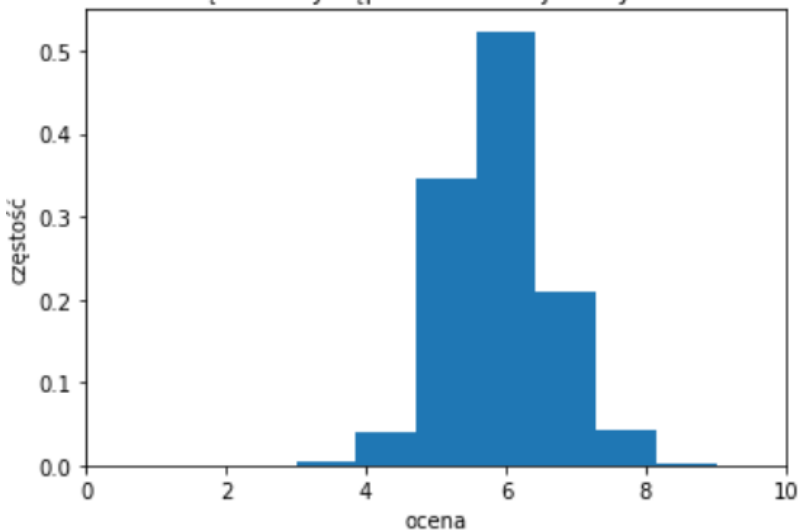
Do weryfikacji jakości modelu wykorzystać:

- k-krotną walidację krzyżową ($k=5$)
- oraz podział na zbiór treningowy i testowy (60/40)

hint: dwa razy się zastanowić jaka jest finalna jakość modelu przy walidacji krzyżowej.

Rozkłady cech przybliżam rozkładem normalnym.

Częstość występowania danej oceny wina



Cechy density i residual sugar są silniej skorelowane niż inne, dlatego przeprowadziłem również próbę klasyfikacji bez jednej z tych cech (Naiwny klasyfikator Bayesa zakłada, że atrybuty są niezależne). Poprawa okazała się być niewielka.

Wyniki:

	Zbiór treningowy/testowy 60/40	Zbiór treningowy/testowy 60/40 (bez cechy density)	5-krotna walidacja krzyżowa (uśredniony wynik)	5-krotna walidacja krzyżowa (uśredniony wynik, bez cechy density)
poprawnie sklasyfikowanych (accuracy)	46.0%	47.2%	44.4%	45.4%
ważona statystyka F1	0.455	0.466	0.435	0.449

Statystykę F1 wybrałem ze względu na lepszą ocenę modelu przy nierównym rozkładzie klas w zbiorze danych.

Rozkład klas w zbiorze danych jest bardzo nierówny, dlatego trudno ocenić jakość otrzymanego modelu. Model mógłby zwracać jakość równą 6 dla wszystkich próbek, a i tak poprawnie „sklasyfikowałby” ok. 40-50% próbek.

Jakiego podzbioru danych (z tych którymi dysponujemy) użyjemy do zbudowania docelowego modelu na potrzeby klasyfikowania nowych próbek (czyli dla tych dla których budujemy klasyfikator)?

Do zbudowania docelowego modelu użyjemy całego dostępnego zbioru danych.

Jak zinterpretować różnice/brak różnic w wynikach z weryfikacji jakości modelu obu metod (k-krotna walidacja vs zbiór treningowy i testowy)?

Różnice są nieznaczne, więc model jest stabilny (klasyfikacja nie zmienia się mocno przy zmianie danych treningowych).