Sprawozdanie - Lab 3

Zadania wstępne 1

1. Po usunięciu kolumny "sex" widać wyraźną zmianę w widgecie rank, która wygląda następująco:

Przed usunięciem:

	#	Info.gain 🗸	Gain ratio	Gini	χ²	ReliefF	FCBF
N sex		0.215	0.226	0.140	189.966	0.006	0.288
N pclass		0.075	0.049	0.050	62.733	0.058	0.064
N fare		0.066	0.033	0.044	79.188	0.023	0.000
N parch		0.030	0.025	0.020	30.688	0.031	0.000
N sibsp		0.019	0.015	0.013	0.417	0.018	0.000
N age		0.002	0.001	0.002	1.319	0.054	0.000

Po usunięciu:

	# Info.gain ~		Gain ratio	Gini	χ²	ReliefF	FCBF	
N pclass		0.075	0.049	0.050	62.733	0.000	0.064	
N fare		0.066	0.033	0.044	79.188	0.007	0.000	
N parch		0.030	0.025	0.020	30.688	0.006	0.028	
N sibsp		0.019	0.015	0.013	0.417	0.006	0.000	
N age		0.002	0.001	0.002	1.319	0.022	0.000	

2. W regresji logistycznej po zmianie z Lasso na Ridge model ma niższe precision oraz F1 i nieco wyższe Recall. Model można dodatkowo osłabić zwiększając siłę (np. do C = 10). Wówczas model ma też nieco gorsze wyniki.

Lasso i C = 2:

Model	~	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
Tree		0.059	0.000	0.953	0.868	0.866	0.870	0.868	0.253	0.839
Logistic Regress	ion	0.015	0.001	0.732	0.695	0.684	0.691	0.695	0.590	0.632



3. W regresji logistycznej również możemy zmienić Lasso na Ridge i zmienić siłę.

Dla Lasso i alpha = 0.003:



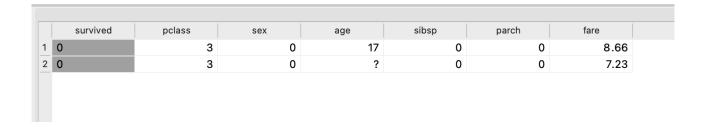
Dla Ridge i alpha = 4:



Widoczna jest niewielka zmiana w MAE.

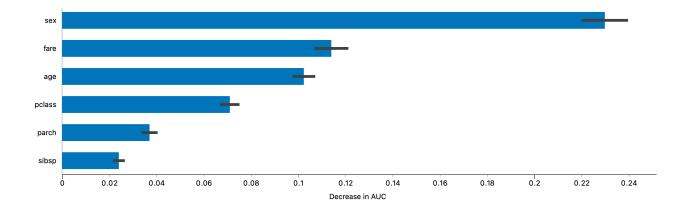
Zadania wstępne 2

Wybieram 2 przykłady:

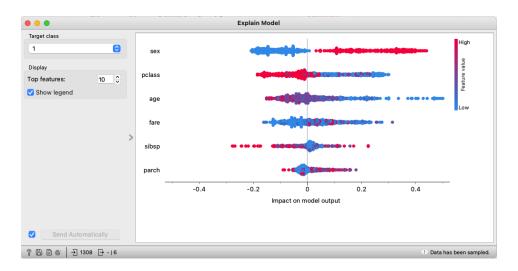


Dla Random Forest:

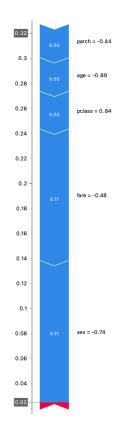
Możemy zauważyć, że najważniejsze cechy w zbiorze to "sex" oraz "fare" (w metodzie klasycznej były sex oraz pclass). Możemy to zobaczyć w widgecie Feature Importance:



Wyjaśnienie modelu (Widget explain model):



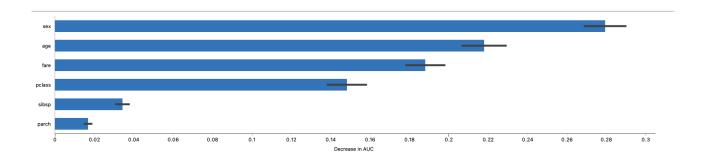
Explain prediction:



Wybrane przeze mnie przykłady otrzymały kategorię 0, ponieważ zadecydowały o tym w głównej mierze "sex" (female) oraz "fare" (niska opłata).

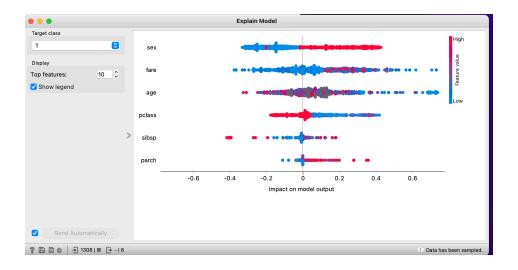
Dla Tree:

Feature importance:

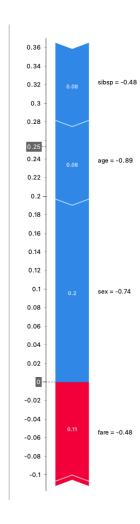


Tutaj największą rolę odgrywają sex, age i fare.

Wyjaśnienie modelu (Widget explain model):



Explain prediction:

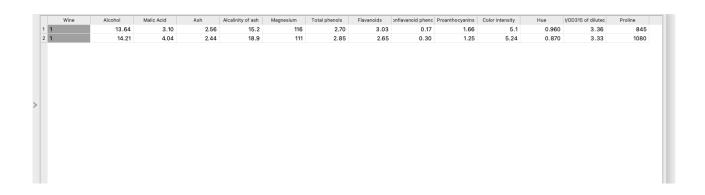


Ponownie wybrane przeze mnie przykłady otrzymały kategorię 0, ponieważ zadecydowały o tym w głównej mierze "sex" (female), "age" (osoba młoda) oraz "fare" (niska opłata).

Zadania finalne:

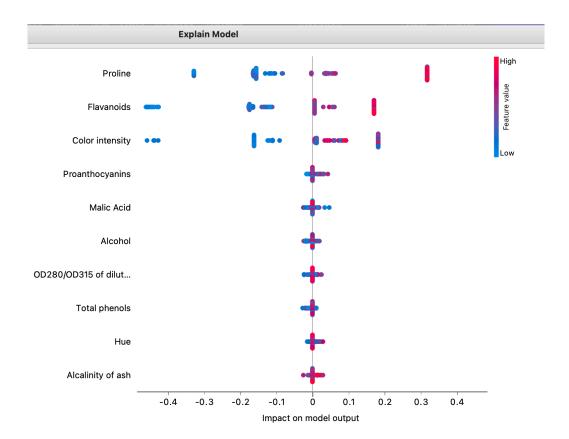
Wybieram zbiór Wine.

Ponownie wybieram 2 przykłady:

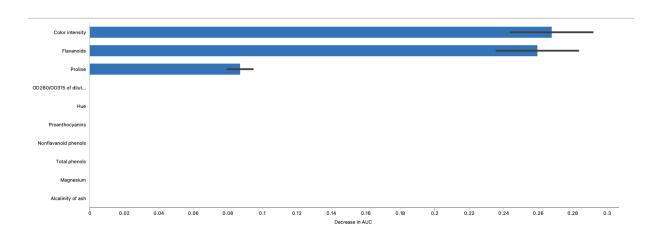


Dla Tree (model słabszy):

Wyjaśnienie modelu:

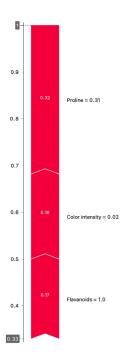


Feature importance:



Najważniejsze cechy to Color intensity oraz flavanoids. Co ciekawe, tylko 3 cechy mają tu znaczenie.

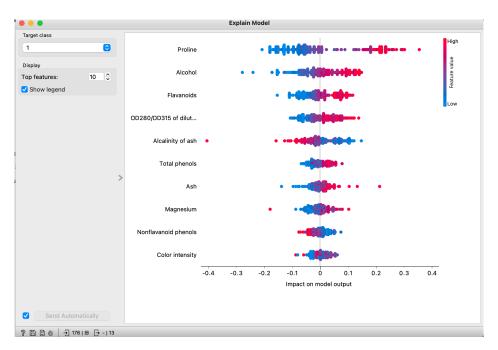
Explain prediction:



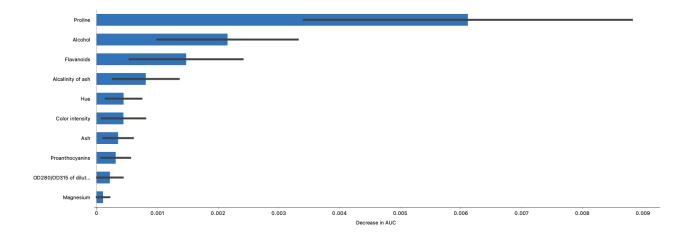
Na zaklasyfikowanie do grupy jeden wpłynęły Color intensity, proline oraz flavanoids.

Dla SVM (model silniejszy):

Wyjaśnienie modelu:

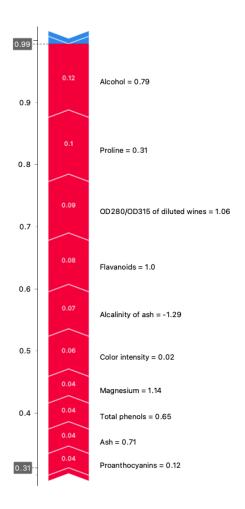


Feature importance:



Tutaj największe znaczenie mają proline oraz alcohol. W odróżnieniu od Tree, Tutaj na wynik wpływa znacznie więcej cech.

Explain prediction:



Jak możemy zauważyć na klasyfikację miały wpływ głównie alcohol oraz proline.