

Zadanie 3.3, 228884, Mateusz Guściora

3.3.1

a)

Zapoznajemy się z procesem Credit Risk Modelling dotyczącym oceny ryzyka kredytowego. Zastosowana jest w niej (cross) walidacja a w niej algorytm kwalifikacyjny SMV. Zapoznajemy się z parametrami i zapisujemy proces i wyniki domyślnego modelu.

b)

Przed zmianą parametrów dla decision tree:

Optimize Parameters (Grid) (33 rows, 4 columns)

iteration	Decisio...	Decisio...	acc... ↓
12	accuracy	31	0.938
24	accuracy	70	0.936
14	informati...	41	0.936
15	accuracy	41	0.936
16	gain_ratio	51	0.936
1	gain_ratio	1	0.933
3	accuracy	1	0.933
2	informati...	1	0.926
8	informati...	21	0.921
5	informati...	11	0.915
23	informati...	70	0.910
22	gain_ratio	70	0.821
25	gain_ratio	80	0.751
26	informati...	80	0.751
28	gain_ratio	90	0.751

Po zmianie parametrów dla decision tree (przy minimal leaf size)

na

Grid/Range

Min	Max	Steps	Scale
<input type="text" value="1.0"/>	<input type="text" value="50"/>	<input type="text" value="5"/>	<input type="text" value="linear"/>

Value List

<input type="text"/>	+ 1
	11
	21
	30
-	40
-	50

Otrzymaliśmy poniższe wyniki:

Optimize Parameters (Grid) (18 rows, 4 columns)

iteration	Decisio...	Decisio...	acc... ↓
10	gain_ratio	30	0.941
13	gain_ratio	40	0.938
6	accuracy	11	0.938
11	informati...	30	0.938
12	accuracy	30	0.938
7	gain_ratio	21	0.938
9	accuracy	21	0.938
18	accuracy	50	0.938
17	informati...	50	0.938
3	accuracy	1	0.936
14	informati...	40	0.936
15	accuracy	40	0.936
16	gain_ratio	50	0.936
4	gain_ratio	11	0.936
1	gain_ratio	1	0.931

Możemy zauważyć że accuracy przy prawie wszystkich wierszach jest większa niż 90% czego nie widzieliśmy w poprzednim przykładzie. Wiąże się to ze zmianą parametrów, zmniejszeniem Max i zmniejszeniem step.

c)

Optimize Parameters (Grid) (33 rows, 4 columns)

iteration	Decisio...	Decisio...	accuracy
1	gain_ratio	1	0.476
9	informati...	21	0.471
17	accuracy	51	0.451
2	accuracy	1	0.425
10	gain_ratio	31	0.476
18	informati...	51	0.450
19	gain_ratio	60	0.476
11	accuracy	31	0.466
3	informati...	1	0.462
4	gain_ratio	11	0.476
20	accuracy	60	0.455
12	informati...	31	0.460

d)*

W zadaniu 2.2.2 najlepsze otrzymane wyniki(accuracy, kappa) były różne w zależności od podpunktu. Zdecydowałem się skorzystać z algorytmu Multilayer Perceptron i skorzystamy z weka extensions aby zrobić to zadanie w Rapid Miner. Można zauważyć długi czas trwania procesu(Nie wiem czy dobrze ustawiłem parametry tego procesu). Accuracy wychodzi na poziomie podobnym co w przypadku pracy w Wece, najwięcej 48,3%.

3.3.2

a)

Operatorem **Remember** możemy zapisać w dowolnej części procesu np. wyniki z jakiegoś podprocesu. Natomiast operatorem **Recall** możemy odczytać/odwołać się do tych wyników w późniejszej części procesu. Operator Default Model Pozwala nam zamienić nim inne operatory predykcyjne i wybrać domyślną metodę(np. medianę czy średnią dla zmiennych numerycznych!). Jest stosowany w modelach do nauki ale nie powinien być stosowany w prawdziwych modelach predykcyjnych. Przykład ułatwiający zrozumienie operatorów: Remember, Recall, Default Model „TutorialIntroduction to Remember and Recall” został zapisany.

b)

Dla zbioru danych klienci6 oraz dla zbioru bank6 nie udało się dodać wszystkich algorytmów z poprzedniego zadania.(Proces nie działał).

Zostały użyte trzy algorytmy :rule induction, decission tree, random forest. Oraz operatory Optimation parameter grid, recall, remember. Dla klienci6 należało zmienić do pliku csv klienci6binominal z dwoma klasami dla przedziału kwotowego. Krzywe ROC dla klientów nie wyglądają sensownie natomiast dla banku6 zostały przypisane poprawnie. Najlepszy okazał się decission tree.

