Zadanie 3.3, 228884, Mateusz Guściora

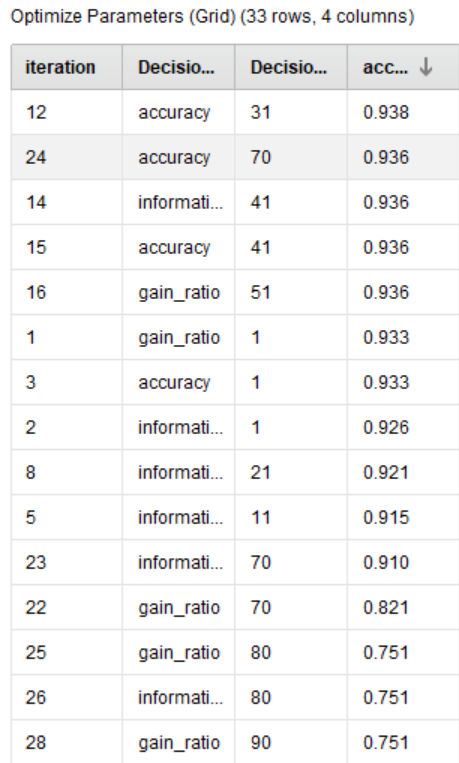
3.3.1

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 decission tree:



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

naObraz zawierający zrzut ekranu

Opis wygenerowany automatycznie

Otrzymaliśmy poniższe wyniki:

Obraz zawierający biały

Opis wygenerowany automatycznie

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)

Obraz zawierający biały, duży

Opis wygenerowany automatycznie

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 metode(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 przedzialu kwotowego. Krzywe ROC dla klientów nie wyglądają sensownie natomiast dla banku6 zostały przypisane poprawnie. Najlepszy okazał się decission tree.

