

Zadanie 3.2

3.2.1 analiza porównawcza klasyfikatorów w trybie Design aplikacji RapidMiner

Korzystając z danych bank6 został stworzony nowy proces w aplikacji RapidMiner. Jako zmienna wyjściowa zostało wybrane pole „y”. Aby podzielić proces na kilka strumieni zastosowano blok Multiply, dzięki czemu możliwe będzie uruchomienie kilku klasyfikatorów. Dla klasyfikatorów Naive Bayes oraz Neural Net utworzone zostały strumienie z walidacją krzyżową, natomiast dla Rule Induction oraz Decision Tree je pominięto. Wyniki przeprowadzonego procesu zostały zaprezentowane w Tabeli 1.

Algorytm	% poprawność klasyfikacji	Wartość Kappa
Decision Tree	70,70%	0,318
Rule Induction	71,98%	0,384
Naive Bayes	65,68%	0,283
Neural Net	63,89%	0,242

Tabela 1 - Wyniki klasyfikacji dla danych bank6

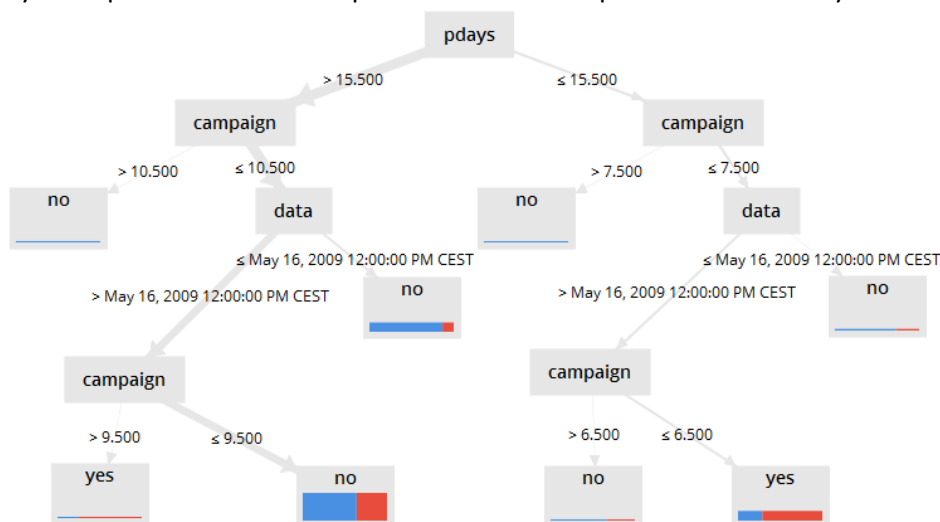
Jak widać, najlepszym klasyfikatorem okazał się algorytm Rule Induction, który osiągnął 71,98% poprawności. Wartość Kappa jednak nie przekroczyła 0,4 zatem klasyfikacja jest niedostateczna.

Wynikiem algorytmu Decision Tree było bardzo szerokie drzewo decyzyjne dlatego też podjęto próbę dostosowania go za pomocą parametrów aby otrzymać łatwiej interpretowalne wyniki. Zmiana parametrów została przedstawiona w Tabeli 2.

Parametr	Wartość przed zmianą	Wartość po zmianie
Minimal gain	0,01	0,03
Minimal leaf size	2	3
Minimal laf for split	4	2
Numer of prepring alternatives	3	4

Tabela 2 - Wartości parametrów dla klasyfikatora Decision Tree przed i po zmianie

Drzewo otrzymane po zmianie wartości parametrów zostało przedstawione na Rysunku 1.



Rysunek 1 - Drzewo otrzymane po zmianie wartości parametrów

Aby otrzymać mniejszy zbiór reguł dla klasyfikatora Rule Induction również zastosowano zmianę parametrów (Tabela 3).

Parametr	Wartość przed zmianą	Wartość po zmianie
Sample ratio	0,9	0,7
Pureness	0,9	0,8
Minimal Prune benefit	0,25	0,2

Tabela 3 - Wartości parametrów dla klasyfikatora Rule Induction przed i po zmianie

Dla algorytmu Neural Net sprawdzono jak zmiana parametrów wpływa na jakość klasyfikacji. Wyniki klasyfikacji przy określonych wartościach poszczególnych parametrów przedstawiono w Tabeli 4.

% poprawność klasyfikacji	63,89%	69,50 %	62,04 %	69,50%
Wartość Kappa	0,242	0,329	0,251	0,331
Training cycles	200	2	2	2
Learning rate	0,01	0,01	0,2	0,2
Momentum	0,9	0,9	0,9	0,4

Tabela 4 - Wyniki klasyfikacji przy użyciu algorytmu Neural Net dla różnej wartości parametrów

Jak można zauważyć zmniejszenie liczby cykli pozytywnie wpłynęło na poprawność klasyfikacji, natomiast przy zwiększeniu wartości parametru Learning rate odnotowano spadek poprawności klasyfikacji oraz wartości statystyki Kappa. Przy zwiększeniu parametru Momentum również dostrzeżono poprawę klasyfikacji.

Do procesu dodany został blok Write as Text aby wyeksportować wyniki klasyfikacji do pliku tekstowego. Wyniki zostały zapisane w pliku wyniki_3.2e.res. Dodatkowo wyniki klasyfikatora Rule Induction zostały zapisane do pliku wyniki_RuleInduction.xlsx.

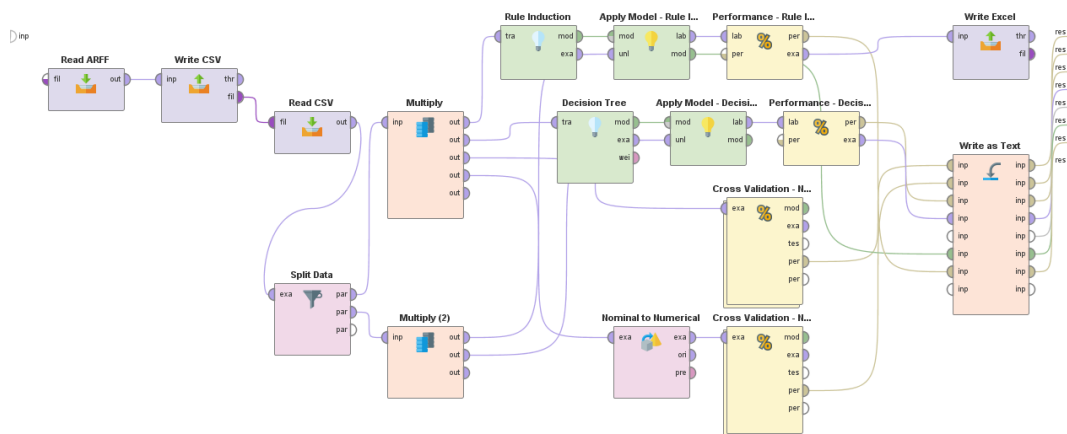
Dla algorytmów Rule Induction oraz Decision Tree zastosowano walidację na danych testowych. W tym celu wykorzystano blok Split Data, który podzielił zbiór danych bank6 na dwie części w stosunku 20%/80%. Wyniki klasyfikacji zostały zaprezentowane w Tabeli 5.

Algorytm	% poprawność klasyfikacji	Wartość Kappa
Rule Induction	68,12%	0,302
Decision Tree	69,19%	0,281

Tabela 5 - Wyniki klasyfikacji algorytmów przy zastosowaniu zbioru testowego i uczącego

Porównując wyniki, do tych otrzymanych w Tabeli 1 widzimy, że poprawność klasyfikacji oraz wartość wskaźnika Kappa nieznacznie spadła. Klasyfikacja w dalszym ciągu jest niedostateczna.

Cały proces otrzymany w wyniku zadania 3.2 został zaprezentowany na Rysunku 2.



Rysunek 2 - Proces otrzymany w wyniku zadania 3.2