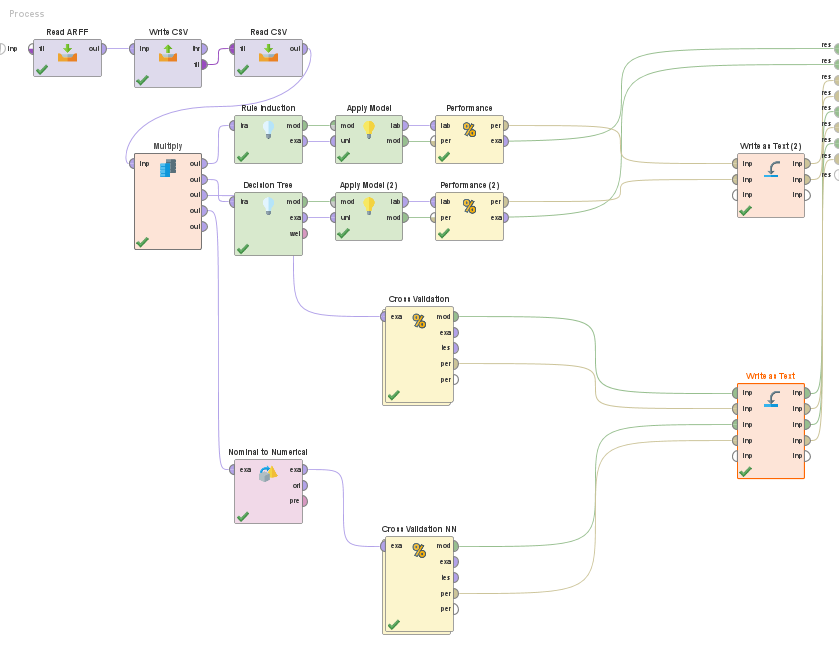
Mateusz Guściora, 228884, zadanie 3.2

3.2.1

Zostały stworzone procesy dla zbiorów klienci6 oraz bank6.

Zostały stworzone dwa procesy dla zbioru klienci6. Proces z pkt. a-f oraz proces z pkt. g czyli dodano split data w stosunku (0.2,0.8). Pierwszy proces jest pokazany poniżej.



Po odczytaniu z performance vector(1,2,3,4) uzyskaliśmy następujące accuracy:

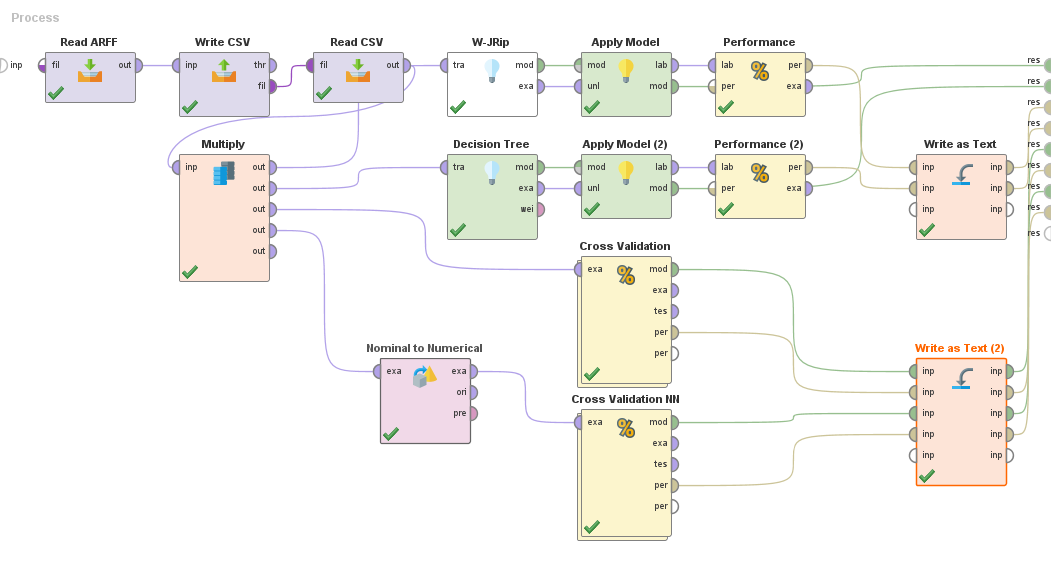
Dla rule induction 53,50%

Dla decission tree 47,60%

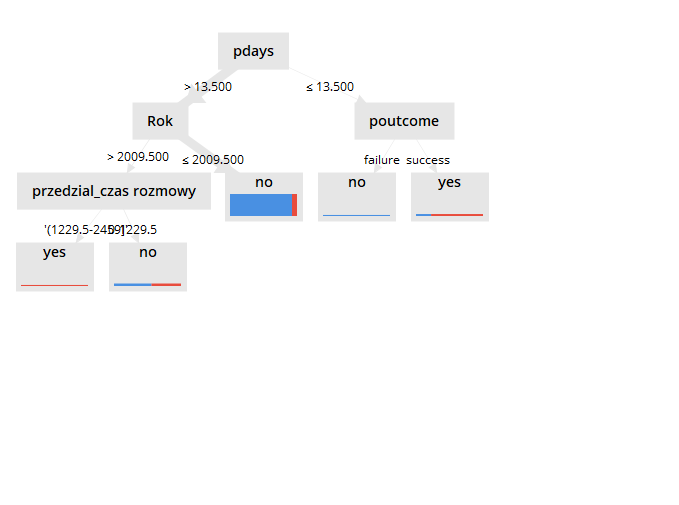
Dla Naive Bayes 45,70%

Dla Neutral Net 41,80%

Przy tworzeniu procesu dla banku6 zamieniliśmy algorytm rule induction algorytmem W-JRip, ponieważ dawał lepszą predykcje i jaśniej przedstawioną. Stworzył osiem reguł.



Algorytm decission tree stworzył nam poniższe drzewo(po zmianie parametrów).



Dokładność była tu znacznie lepsza niż w przypadku danych o klientach.

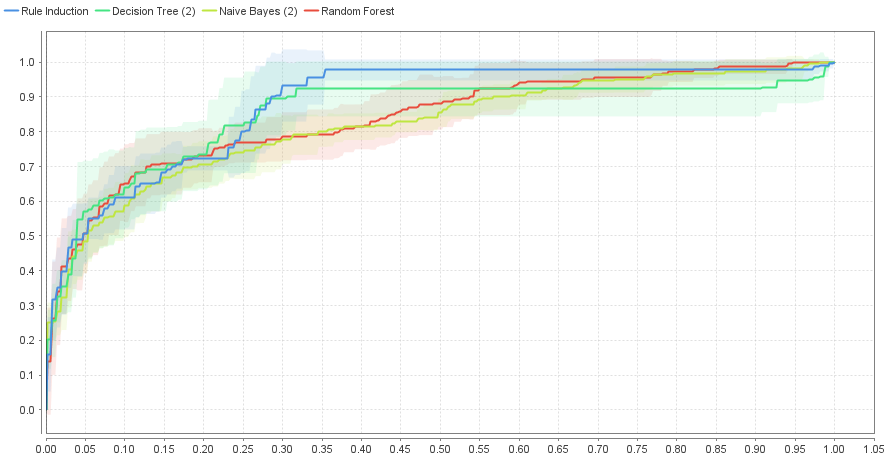
Dla W-Jrip 88,25%

Dla decission tree 88,89%

Dla naive Bayes 84,72%

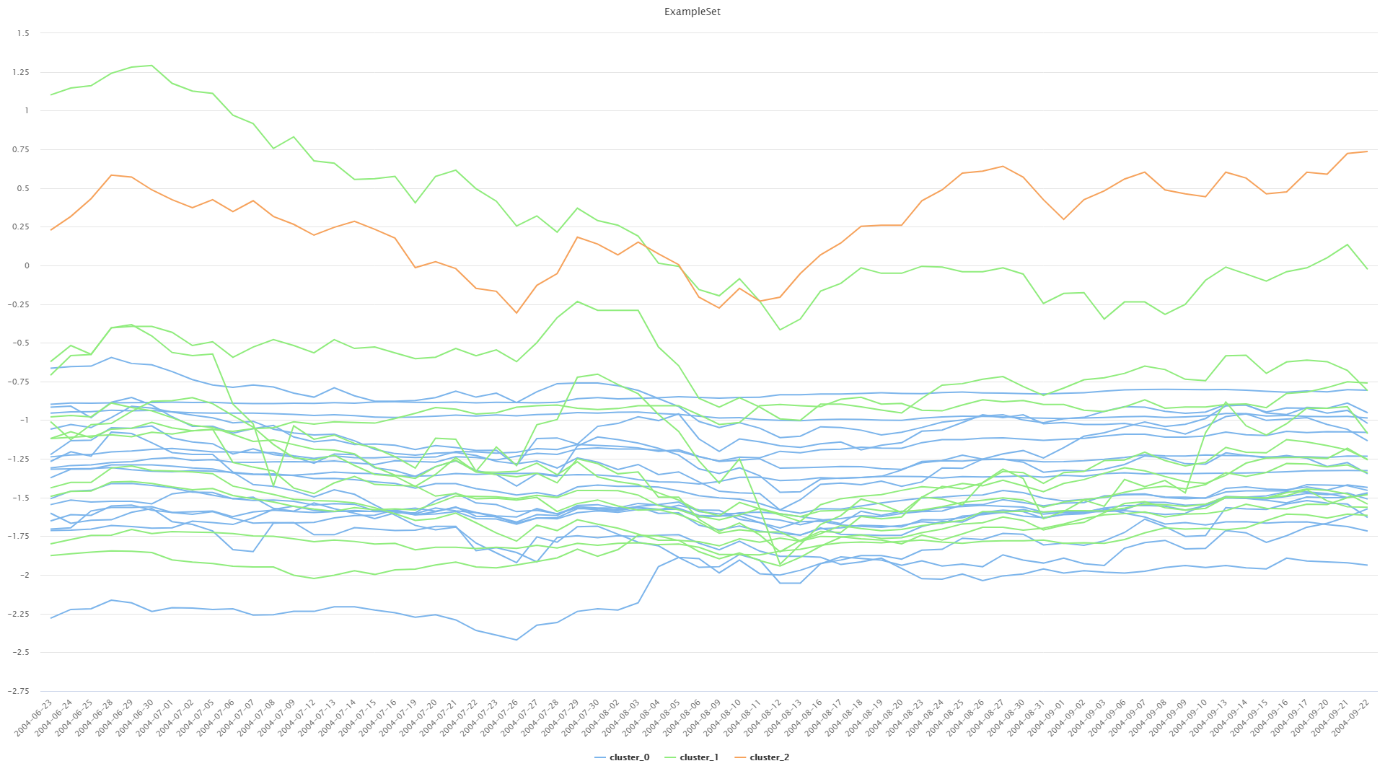
Dla Neutral Net 89,35%

W punkcie h) porównaliśmy krzywe ROC dla algorytmów :rule inducion, decission tree, naive baes i random forest.



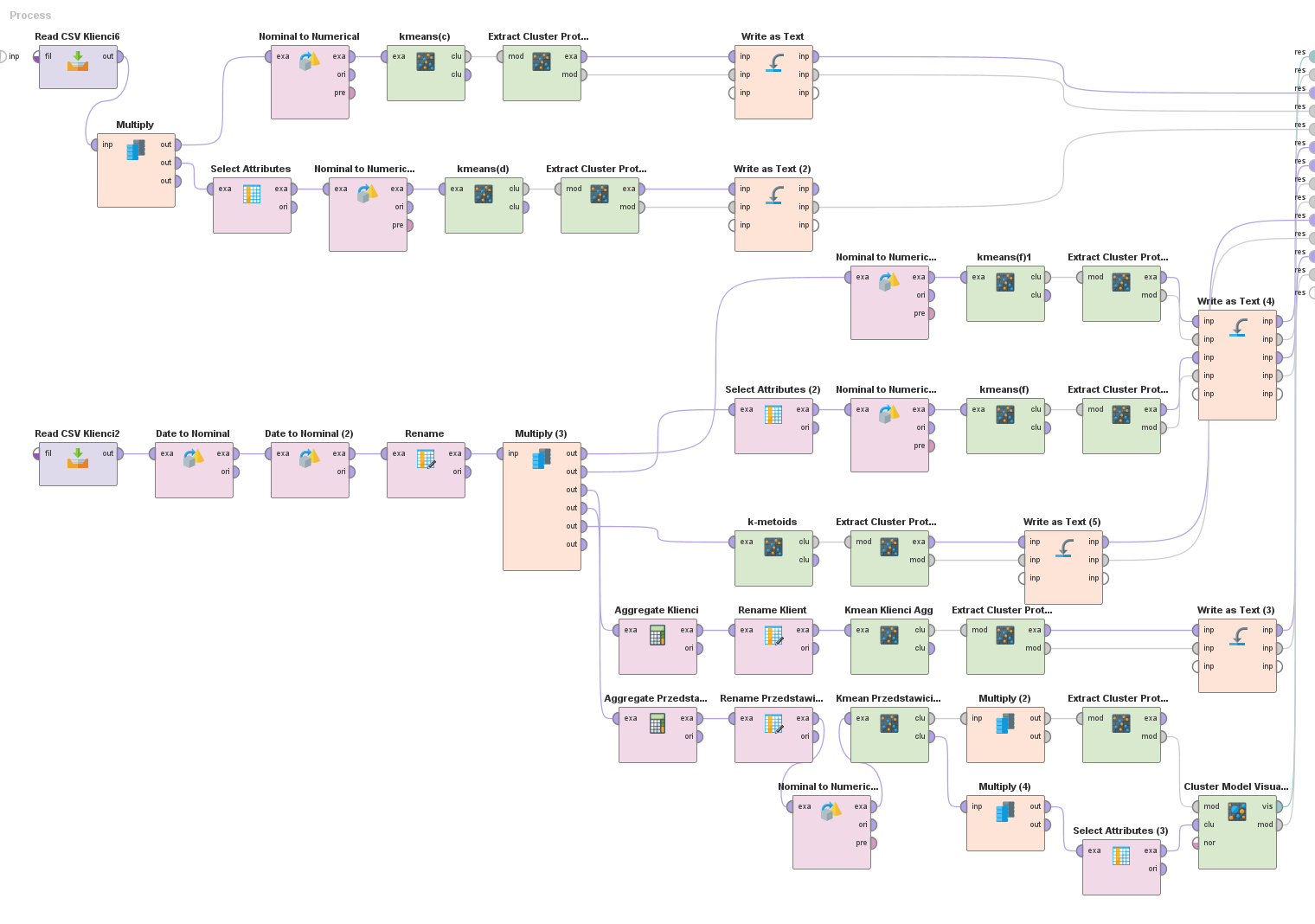
3.2.2

Na przykładzie procesu Price Risk Clustering mogliśmy zobaczyć działanie metody grupowania X-means. Oraz można było stworzyć wizualizacje do tego przykładu.

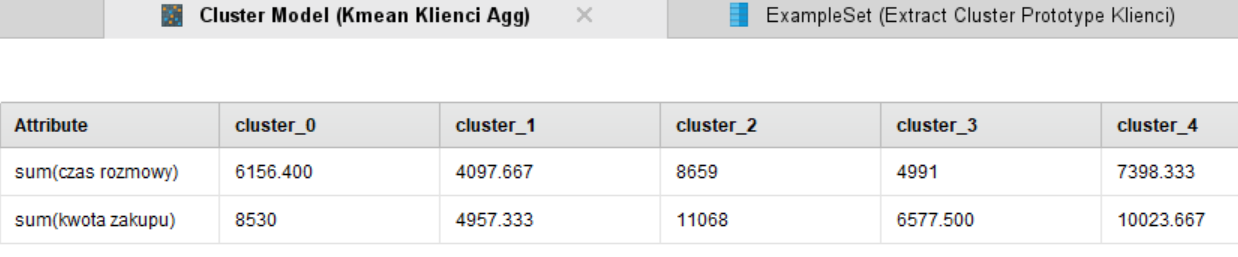


3.2.3

W tym zadaniu tworzyliśmy proces składający się z kilku strumieni i dwóch zbiorów danych csv: klienci2 oraz klienci6 z danymi pogrupowanymi. To głównie na przykładzie klienci2 używaliśmy metod grupowania k-means.



Wyniki zapisano do plików res. Ważniejsze strumienie są to strumienie gdzie grupowaliśmy klientów oraz przedstawicieli i jako funkcje agregująca była suma czasu rozmowy i kwoty zakupu.\



Obraz zawierający zrzut ekranu

Opis wygenerowany automatycznie