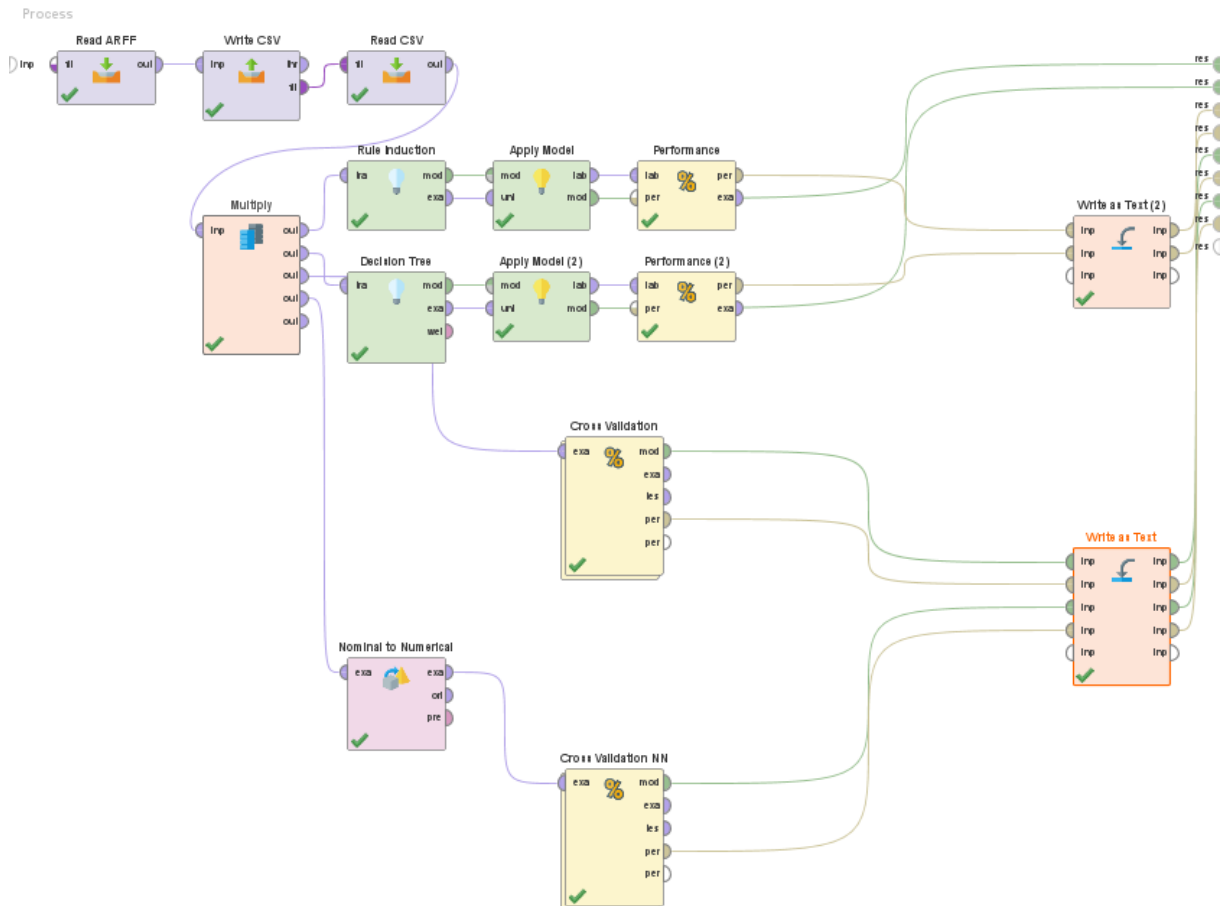


Mateusz Guściora, 228884, zadanie 3.2

### 3.2.1

Zostały stworzone procesy dla zbiorów klientów oraz banku.

Zostały stworzone dwa procesy dla zbioru klientów. 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:

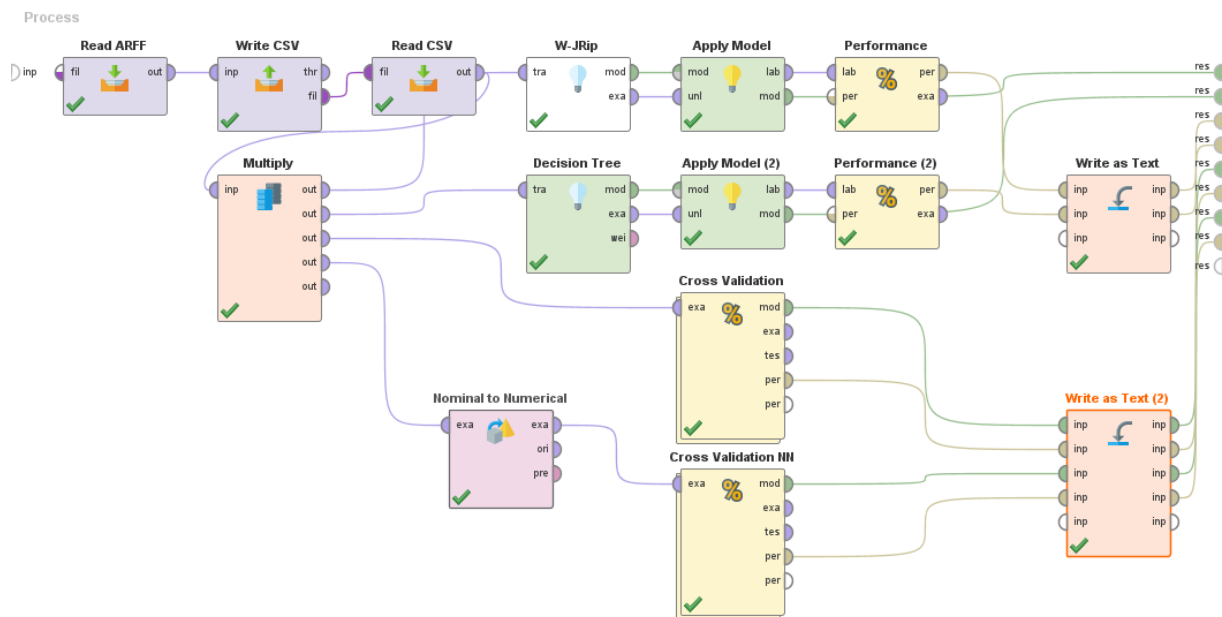
Dla rule induction 53,50%

Dla decision tree 47,60%

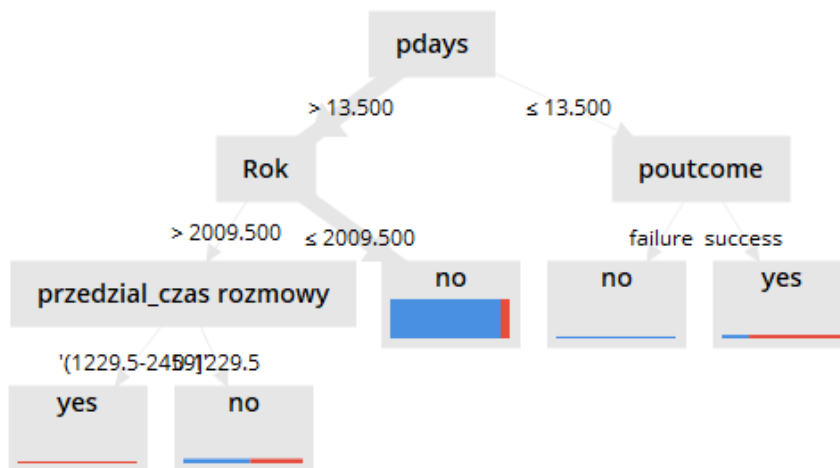
Dla Naive Bayes 45,70%

Dla Neutral Net 41,80%

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



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



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

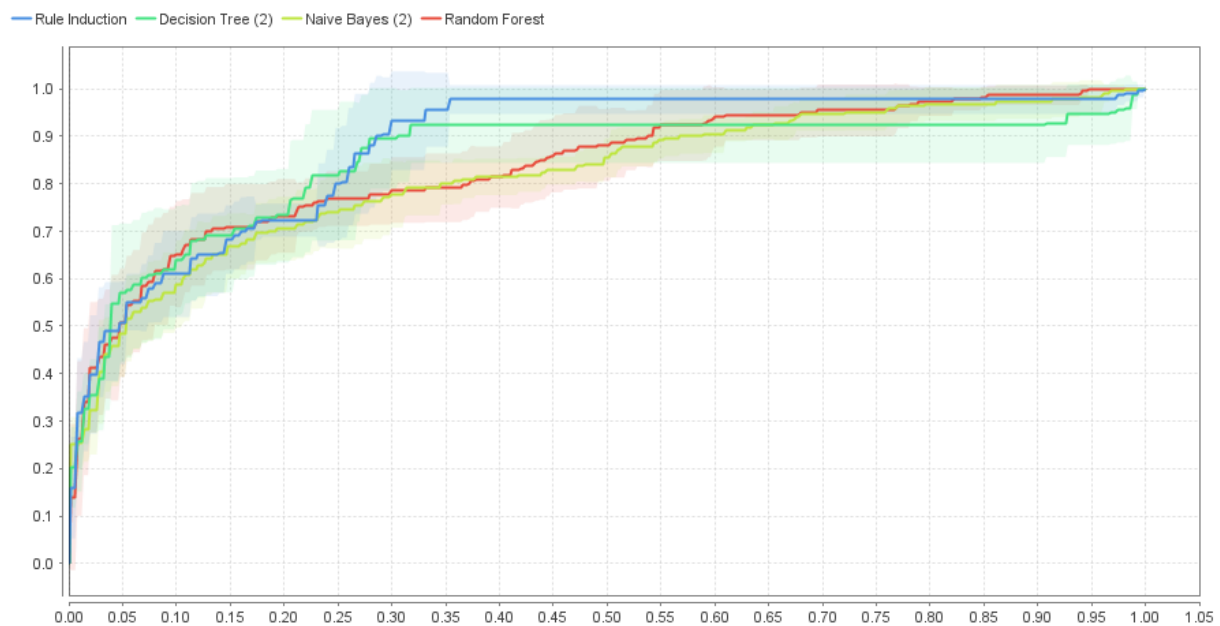
Dla W-Jrip 88,25%

Dla decision tree 88,89%

Dla naive Bayes 84,72%

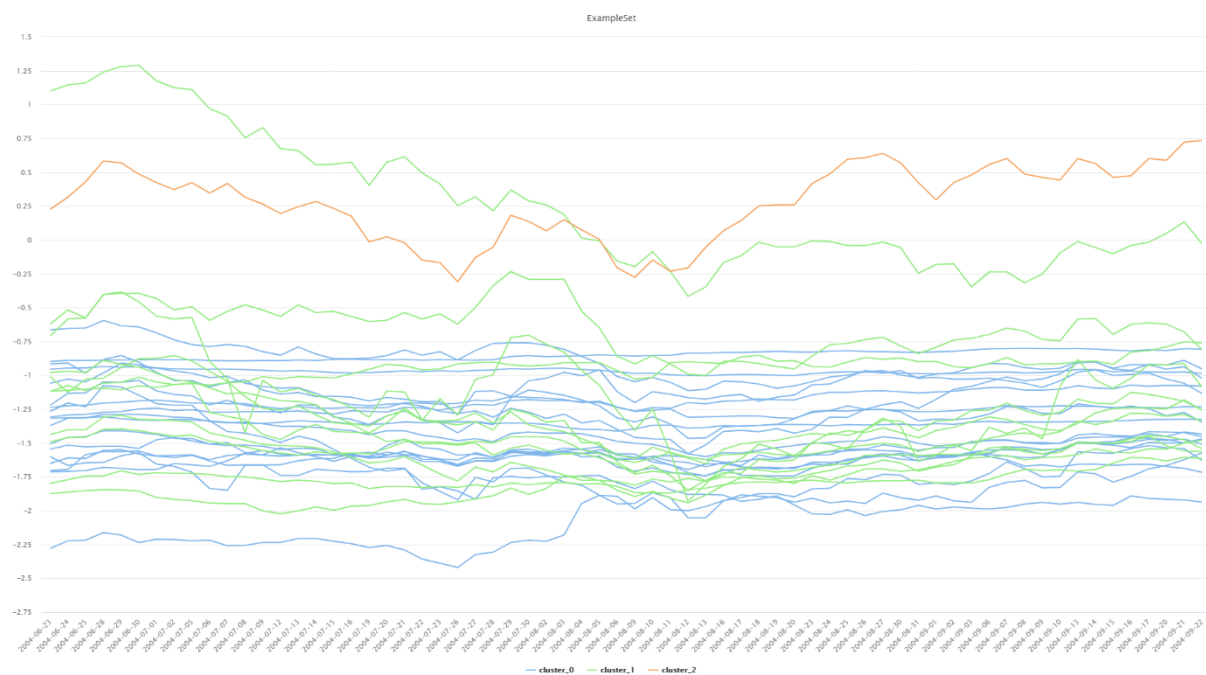
Dla Neutral Net 89,35%

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



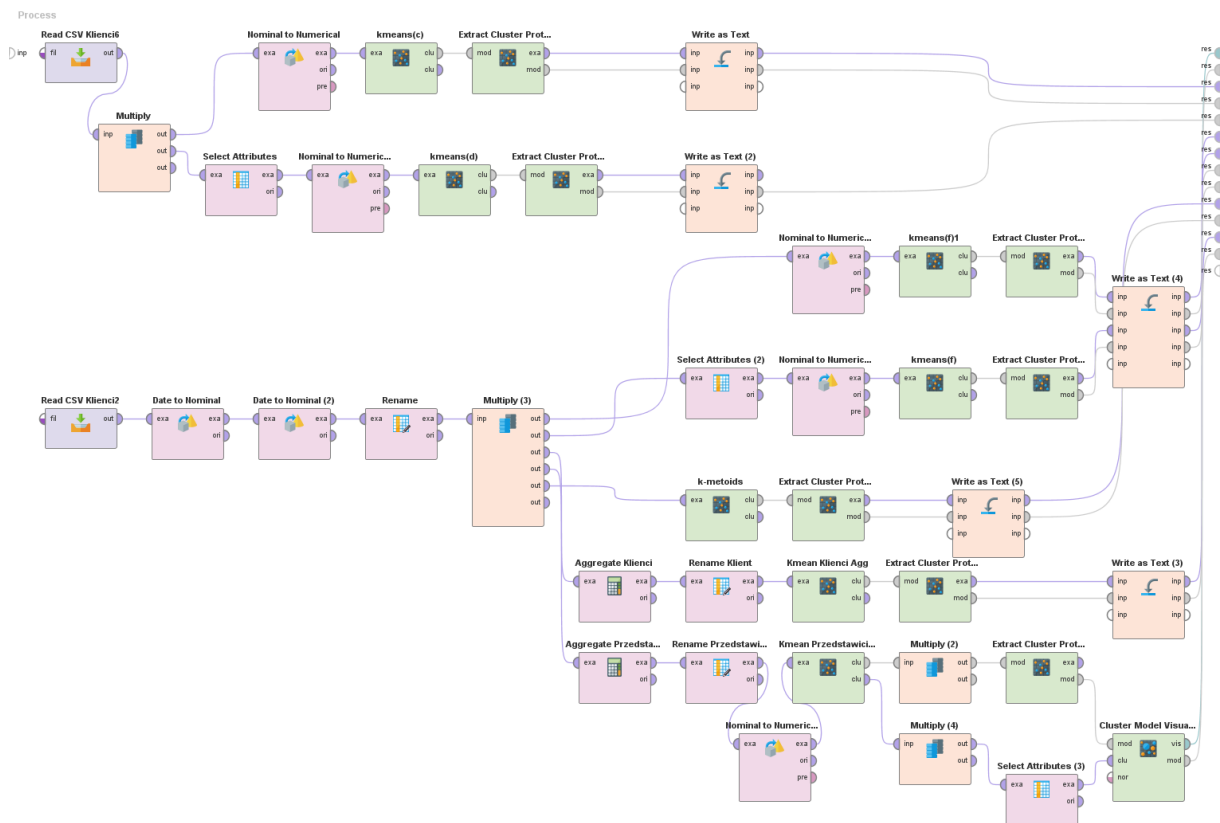
### 3.2.2

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.

	<b>Cluster Model (Kmean Klienci Agg)</b>		<b>ExampleSet (Extract Cluster Prototype Klienci)</b>
--	--	--	---

Attribute	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4
sum(czas rozmowy)	6156.400	4097.667	8659	4991	7398.333
sum(kwota zakupu)	8530	4957.333	11068	6577.500	10023.667

<b>Cluster Model (Kmean Przedstawiciele Agg)</b>		<b>ClusterModelVisualizer!OObject (Cluster Model Visualizer)</b>	
<b>Cluster Model (kmeans(d))</b>		<b>Cluster Model (kmeans(c))</b>	
<b>ExampleSet (Extract Cluster Prototypes(c))</b>		<b>ExampleSet (Extract Cluster Prototypes(f))</b>	
<b>Cluster Model (kmeans(f1))</b>		<b>ExampleSet (Extract Cluster Prototypes (f))</b>	
<b>Cluster Model (kmeans(f))</b>		<b>ExampleSet (Extract Cluster Prototypes (f) (2))</b>	
<b>Cluster Model (k-means(f))</b>		<b>ExampleSet (Extract Cluster Prototypes)</b>	
<b>Cluster Model (k-means(f))</b>		<b>Cluster Model (kmeans(f))</b>	
<b>Cluster Model (Kmean Klienci Agg)</b>		<b>ExampleSet (Extract Cluster Prototype Klienci)</b>	

Attribute	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4
id przedstawiciela = P03	0	0	1	0	0
id przedstawiciela = P02	0	1	0	0	0
id przedstawiciela = P05	1	0	0	0	0
id przedstawiciela = P01	0	0	0	0	1
id przedstawiciela = P04	0	0	0	1	0
sum(czas rozmowy)	17992	19976	19022	18192	17388
sum(kwota zakupu)	24470	24857	25155	23976	24426