

Sprawozdanie z laboratorium

Analiza Danych i Inteligencja Obliczeniowa

Rapid Miner

Bartosz Borys

Grupa 5

Pseudonim:

Koszmar

przywitajmy

1. "DANE"

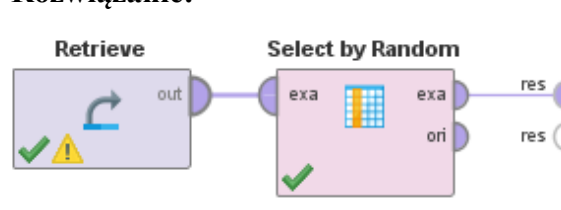
Do projektu zaimportowane zostały trzy foldery zawierające utwory artystów. Foldery te noszą nazwy: Boys, Akcent, Shazza.

Rozwiązanie:



2. "SYNONIM"

Rozwiązanie:

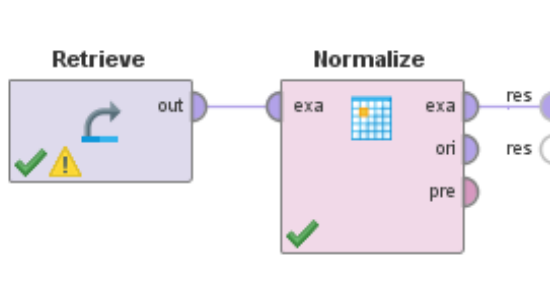


Row No.	label	metadata_file	metadata_d...	metadata_p...	Koszmar	przywitajmy
1	Boys	0.txt	Oct 23, 2018 ...	C:\Users\Bart...	0	0
2	Boys	1.txt	Oct 23, 2018 ...	C:\Users\Bart...	0	0
3	Boys	10.txt	Oct 23, 2018 ...	C:\Users\Bart...	0	0
4	Boys	100.txt	Oct 23, 2018 ...	C:\Users\Bart...	0	0

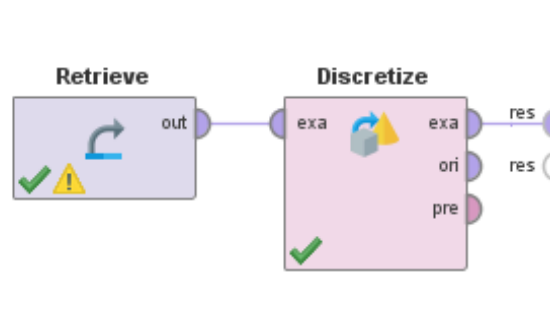
3. "PRZYGOTOWANIA"

Rozwiązanie:

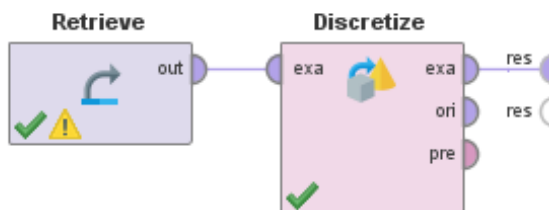
a) Normalize



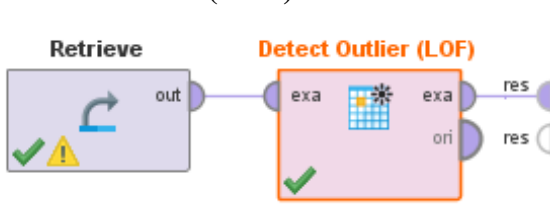
b) Discretize by Binning:



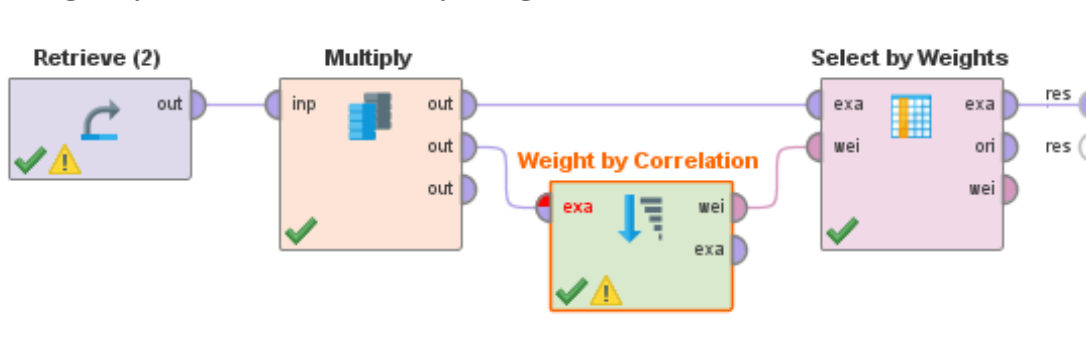
b) Discretize by Entropy:



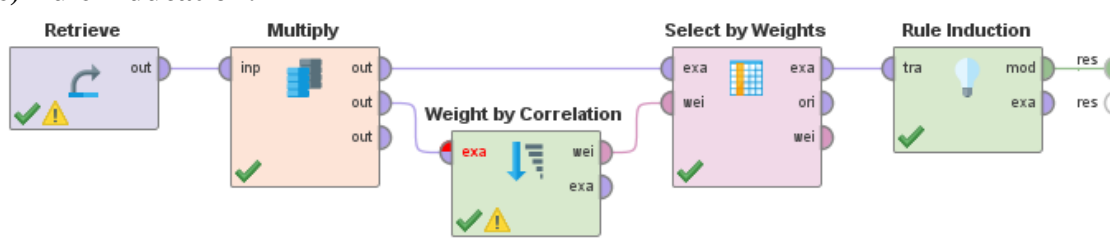
c) Detect Outlier (LOF):



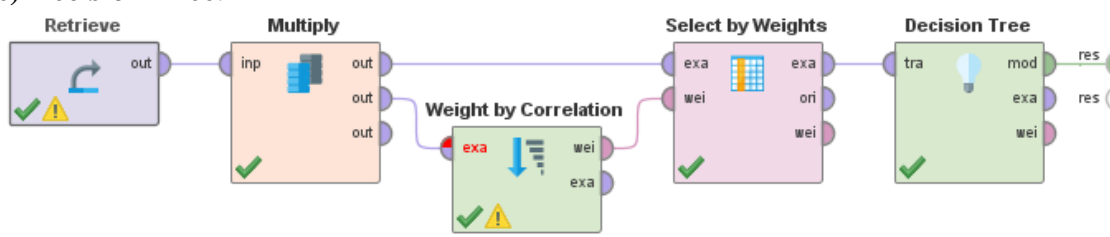
d) Weight by Correlation, Select by Weights:



e) Rule induction:



e) Decision Tree:

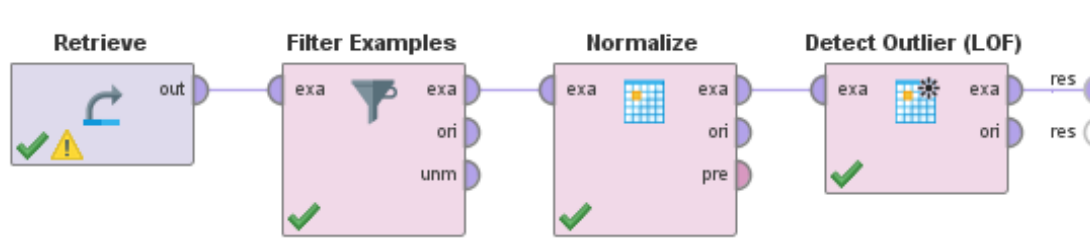


Wnioski:

Po przeprowadzeniu moich badań przy pomocy operatorów Detect Outliner, Weight by Correlation, Decision Tree i wreszcie Discretize by Entropy doszedłem do pewnych wniosków. Detect Outliner, operator ten zajmuje się przedstawieniem jak daleki od reszty zbioru jest dany element. W przypadku mojego zadania, jak bardzo odstaje od reszty dany plik. Weight by Correlation, tutaj mamy doczynienia ze sprawdzeniem wagi każdego atrybutu w kontekście danej etykiety. Decision Tree, informuje nas o najbardziej dopasowanej etykiecie do danego atrybutu. Zaś ostatni Discretize by Entropy stara dopasować granice powstałych kontenerów, aby ich nieuporządkowanie (Entropia) dążyła do zera. Dzięki temu łatwiej i lepiej jest dopasować słowa do danego zbioru.

4. "DYWERSJA"

Rozwiązanie:



Dla boys

Row No.	label	metadata_file	metadata_d...	metadata_p...	outlier ↓	A	ABY	AIMES
184	Boys	94.txt	Oct 23, 2018 ...	C:\Users\Bart...	1.978	0	0	0
37	Boys	130.txt	Oct 23, 2018 ...	C:\Users\Bart...	1.963	0.244	0	0
152	Boys	65.txt	Oct 23, 2018 ...	C:\Users\Bart...	1.661	0	0	0

Dla Akcent

Row No.	label	metadata_file	metadata_d...	metadata_p...	outlier ↓	A	ABY	AIMES	AMC
144	Akcent	73.txt	Oct 23, 2018 ...	C:\Users\Bart...	2.351	0	0	0	0
81	Akcent	170.txt	Oct 23, 2018 ...	C:\Users\Bart...	1.706	0.586	0	0	0
60	Akcent	151.txt	Oct 23, 2018 ...	C:\Users\Bart...	1.689	0	0	0	0

Dla Shazza

Row No.	label	metadata_file	metadata_d...	metadata_p...	outlier ↓	A	ABY	AIMES	AMC
79	Shazza	8.txt	Oct 23, 2018 ...	C:\Users\Bart...	1.672	0.075	0	0	0
46	Shazza	5.txt	Oct 23, 2018 ...	C:\Users\Bart...	1.667	0.757	0	0	0
68	Shazza	7.txt	Oct 23, 2018 ...	C:\Users\Bart...	1.528	0.352	0	0	0

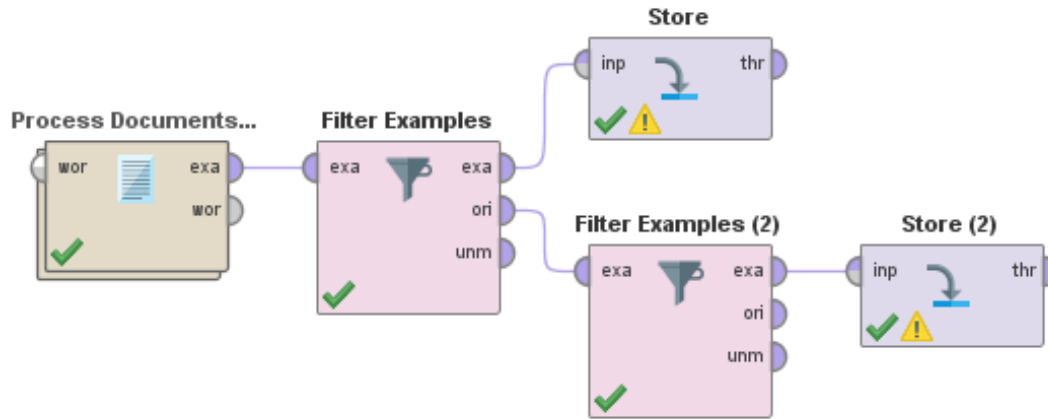
Wnioski:

Podczas przeszukiwania plików w celu znalezienia, niepasujących plików do danego zbioru. Dane wejściowe zostały znormalizowane dzięki "range transformation" w zakresie (0-1) i przeszły przez obiekt Detect Outlier, aby znaleźć wcześniej wspomniane podejrzane pliki. Rezultatem tego była lista plików ze współczynnikiem odstawania od normy. Trzy najbardziej podejrzane przedstawiłem na powyższych zrzutach.

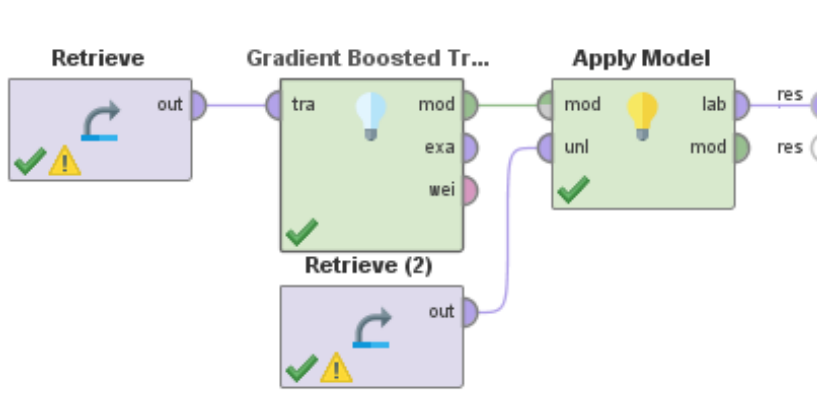
5. "NA FRONCIE"

Rozwiązanie:

Dane:



Główny:



Plik tekstowy	Przewidywany zespół	Rzeczywisty zespół
01.txt	Boys	Akcent
02.txt	Boys	Boys
03.txt	Shazza	Shazza
04.txt	Boys	Akcent
05.txt	Akcent	Akcent
06.txt	Akcent	Boys
07.txt	Shazza	Shazza
08.txt	Boys	Shazza
09.txt	Boys	Boys

Wnioski:

Klasyfikacja udało się na 5 z 9 utworów co daje nam dokładność w przybliżeniu 60%.

6. "Nowa Nadzieja"

Niestety, ze względu na chwilowy brak "Mocy" i nie uzyskaniu jej przez kompanów, nie udało się zrealizować tego zadania w którym wymagane było porównanie naszych wyników.