

Wykorzystanie analizy szeregów czasowych w praktyce

Problem

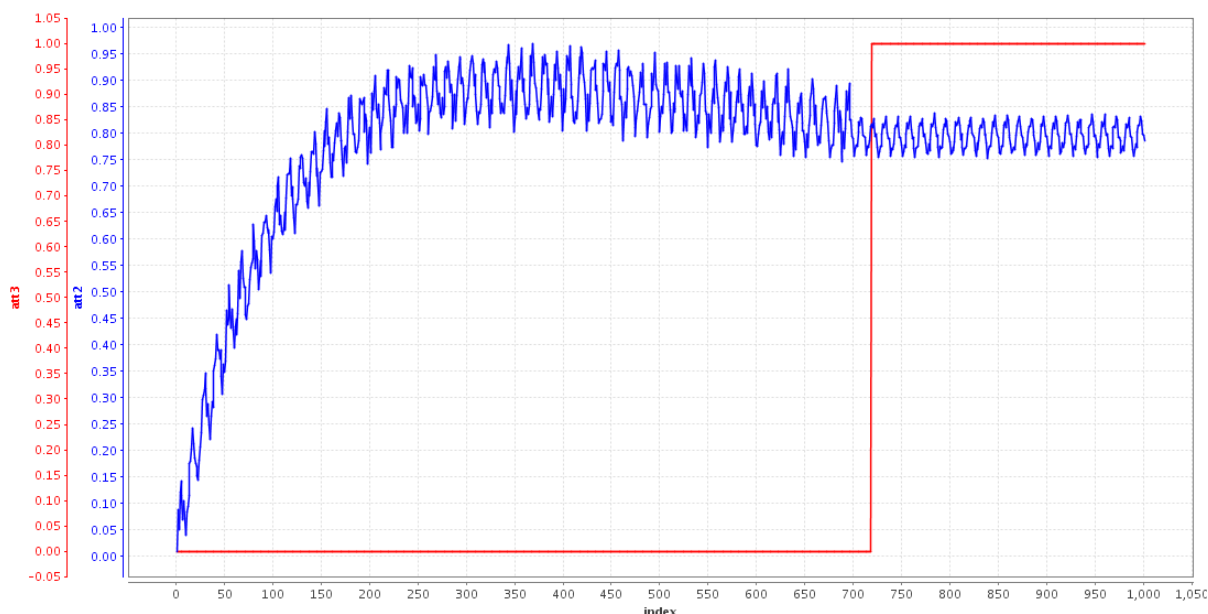
Celem niniejszego ćwiczenia jest przetwarzanie, analiza i predykcja szeregów czasowych przeprowadzona na przykładzie realnego problemu pochodzącego z przemysłu.

Jednym z przykładów problemów spotykanych w przemyśle jest przewidywanie stanu procesu technologicznego. Przykładem takiego problemu jest przewidywanie stanu uspokojonego topienia w piecu łukowym podczas topienia złomu.



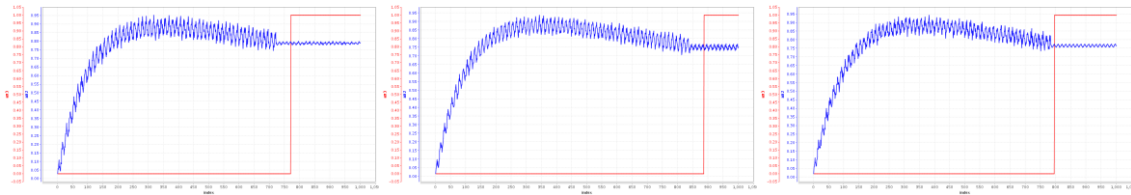
Źródłem tego problemu jest konieczność automatycznego wykrycia chwili czasowej, w której cała zawartość złomu w piecu jest stopiona, co w konsekwencji pozwala na określenie momentu wyłączenia pieca i oszczędności związanych ze zużyciem energii elektrycznej.

Aby móc zautomatyzować taki proces konieczne jest przeanalizowanie dostępnych sygnałów, których przykładem jest stan napięcia elektrod odpowiedzialnych za topienie złomu. Wykres wartości napięcia w funkcji czasu oraz stanu pieca przedstawia poniższy wykres



Gdzie kolorem niebieskim oznaczono wartość napięcia elektrod, natomiast kolorem czerwonym stan kąpieli, gdzie wartość 1 odpowiada stanowi uspokojonego topienia, czyli chwili czasowej, w której cały złom jest stopiony.

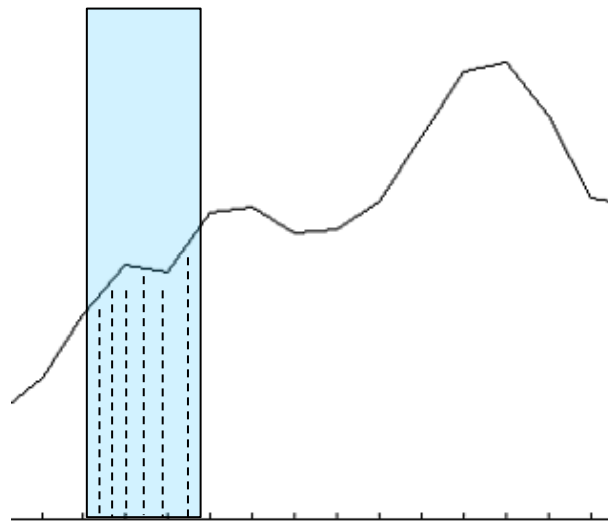
Analiza szeregu różnych wytopów:



wskazuje, iż stan uspokojonego topienia związany jest z poziomem szumu wartości rejestrowanego napięcia. Tym samym, aby móc określić stan pieca konieczna jest estymacja parametru sygnału, jakim jest jego wariancja (odchylenie standardowe).

Estymacja parametrów sygnału

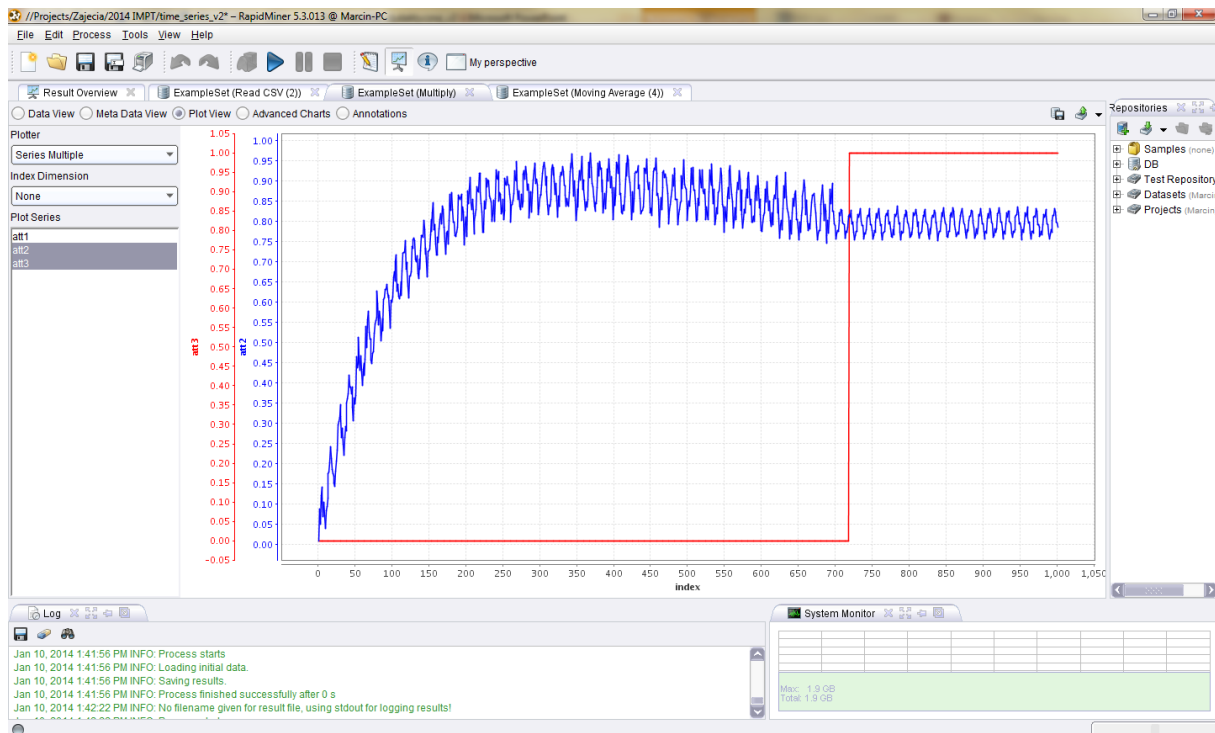
Jedną z podstawowych metod estymacji parametrów sygnału jest tzw. okienkowanie (ang. windowing) które polega na wycięciu fragmentu sygnału i estymacja wybranego parametru w danym oknie czasowym.



Dzięki takiemu rozwiązaniu możliwe jest określenie parametrów dla każdej kolejnej chwili czasowej. Przykładem estymowanego parametru może być uśrednianie (ang. moving window average), które w efekcie prowadzi do redukcji poziomu szumu sygnału.

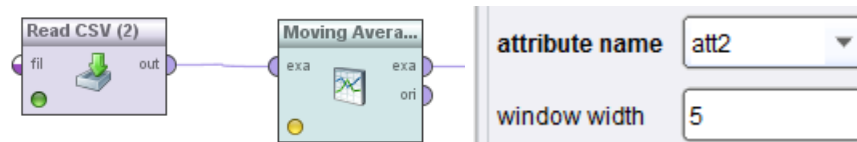
Zadanie

W RapidMinerze wczytaj zbiór danych *timeData.csv*, korzystając z operatora *Read CSV*, a następnie podłącz go do wyjścia i zaobserwuj zapisane w nim sygnały. Zbiór ten składa się z trzech kolumn, pierwsza to czas, kolejna to właściwy sygnał opisujący wartość napięcia elektrody, a ostatnia to stan pieca. Zwizualizuj wczytane sygnały korzystając z widoku *Results Perspective*, *Plot View* ustawiając typ wykresu na *Series Multiple*. Powinieneś uzyskać wykres jak poniżej



Zadanie

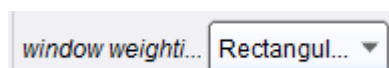
W następnym kroku z dodatku *Series* podłącz operator *Moving Average (Series->Data Transformation ->Moving Average)*. Operator ten do działania wymaga określenia atrybutu dla którego ma być zastosowany (pole opcji o nazwie *attribute name*), w naszym przypadku jest to *att2* (atrybut zawierający sygnał). Kolejnym istotnym parametrem jest *window size* który określa rozmiar okna.



Sprawdź i pokaż wpływ rozmiaru okna ustawiając różne wartości tego parametru (można stosować wartości typu kilkadziesiąt i większe). W sprawozdaniu pokaż wpływ tego parametru.

Zadanie

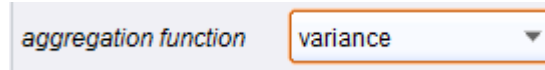
Najprostsze rozwiązanie systemu okienkowania zakłada okno prostokątne, jednakże możliwe są również inne kształty okna, np., Gaussowskie, Trójkątne. Taki kształt okna powoduje, że kolejne wartości analizowane w ramach okna pobierane są z coraz mniejszą wagą, czyli im dalej oddalimy się od prawej krawędzi okna tym dane mają mniejszy wpływ na obecną jego wartość.



Kształt okna określa się za pomocą opcji *window weight* operatora *Moving Average*. Pokaż w sprawozdaniu wpływ kształtu okna na wynik redukcji szumu, czyli filtracji sygnału.

Zadanie

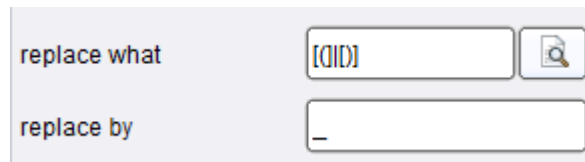
Jak wspomniano we wstępie w celu identyfikacji stanu pieca konieczna jest analiza poziomu szumu sygnału, dlatego też inną możliwością oferowaną przez operator *Moving Average* jest zmiana typu estymowanego parametru z średniej na variancję lub odchylenie standardowe.



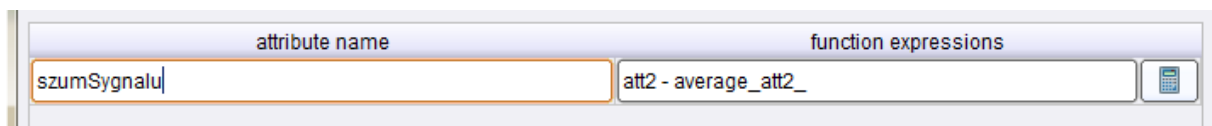
W tym celu zmień typ estymatora na variancję i zaobserwuj wynik pokazując na wykresie sygnał oryginalny oraz wartość estymowanego poziomu variancji.

Zadanie

Innym sposobem estymacji poziomu szumu jest odjęcie od oryginalnego sygnału jego postaci odfiltrowanej czyli sygnału po uśrednieniu. W tym celu wykorzystaj operator *Moving Average* z opcją estymacji wartości średniej arytmetycznej, w efekcie zbiór danych będzie zawierał zarówno sygnał oryginalny jak i sygnał odfiltrowany (po uśrednieniu). Teraz należy odjąć od siebie obydwa atrybuty. Aby to zrealizować konieczne jest zastosowanie operatora *Generate Attributes* dostępnego w *Data Transformation -> Attribute Set Reduction and Transformation -> Generation*. Jednak aby go zastosować konieczne jest aby nazwy atrybutów nie zawierały znaków specjalnych takich jak nawiasy np. "(", „ Aby zmienić nazwy atrybutów wykorzystaj operator *Rename by Replacing* dostępny w *Data Transformation -> Name and Role Modification*. W operator ten pozwala na automatyczną zmianę nazw atrybutów poprzez zastąpienie fragmentów nazwy innym napisem. Konieczne jest więc stworzenie odpowiedniego wyrażenia regularnego którego celem jest zastąpienie symboli nawiasów symbolami podkreślenia. Odpowiednie wyrażenie regularne to: `[() | D]` które oznacza: znajdź symbol specjalny (lub symbol specjalny) (symbole specjalne muszą być umieszczone w nawiasach kwadratowych).



Po takiej zamianie symboli w nazwach atrybutów można dokonać odjęcia dwóch sygnałów. Aby tego dokonać w operatorze *Generate Attributes* ustaw parametr *function description* jak na rysunku



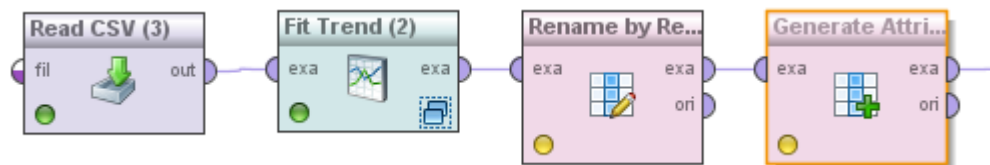
Zapis ten oznacza, że powstanie nowy atrybut *szumSygnału* który powstanie jako różnica wartości obliczona dla każdego wektora jako *att2 - average_att2_*. W następnym kroku podłącz operator estymacji variancji krokowej (patrz poprzednie zadania – *Moving average* z odpowiednią opcją).

Pokaz na wykresie uzyskany wynik operacji i skomentuj uzyskane rezultaty. Poniżej dla ułatwienia pokazana schemat połączeń

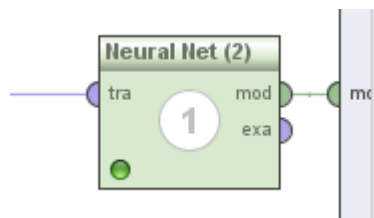


Zadanie

Istnieją również inne metody pozwalające na uzyskanie podobnego rezultatu. Główna idea pozostaje jednak taka sama czyli odjęcie od oryginalnego sygnału jego postaci pozbawionej szumu. Powyższe zadanie można zrealizować korzystając z operatora estymacji trendu. Zwykle w tym miejscu jesteśmy przyzwyczajeni do estymacji trendu w postaci trendu liniowego, jednakże może to być dowolna krzywa uzyskana za pomocą dowolnego algorytmu regresji. Poniżej zostanie przedstawiany przykład z wykorzystaniem trendu wyznaczonego za pomocą sieci neuronowej. W tym celu po wczytaniu danych podłącz operator *Fit Trend*, którego znajdziesz w *Series -> Data Trensformation*. Następnie dokonaj zamiany nazw atrybutów pozbywając się znaków specjalnych (jak w poprzednim zadaniu), oraz wyznacz różnicę pomiędzy sygnałem oryginalnym a estymowanym jako trend (patrz poprzednie zadanie). Na koniec wstaw ten sam estymator wariancji kroczącej ustawiając go na nowo wygenerowany atrybut. Schemat połączeń przedstawia rys.



Pozostaje jeszcze konfiguracja operatora *Fit Trend*. Aby jej dokonać należy wejść do podprocesu z nim związanego i ustawić odpowiedni algorytm regresji. Przykładem może tutaj być sieć neuronowa.



Jednym z istotnych parametrów sieci neuronowej jest jej układ tzn. liczba warstw i liczba neuronów w warstwach. Tych ustawień konfiguracyjnych dokonujemy w ustawieniach sieci neuronowej klikając na



I po otwarciu okna określamy liczbę neuronów osobno dla każdej warstwy oraz jej nazwę:

hidden layer name	hidden layer sizes
w1	2
w2	2

Sprawdź różne ustawienia sieci neuronowych w tym z jedną warstwą i różną liczbą neuronów (1,2,5 neuronów) oraz z dwoma warstwami (2 neurony w pierwszej warstwie i 1, 2, 5 neuronów w drugiej) i obserwuj jak wpływają parametry sieci na wyznaczoną krzywą trendu oraz estymowany poziom szumu. Przykładowe wykresy zamieść w sprawozdaniu.

Zadanie

Ponieważ sygnał ulega tutaj ciągłym zmianom, a naszym celem jest estymacja poziomu szumu, więc kolejną możliwością to wyznaczenie pochodnej sygnału. Pochodna danej funkcji pokazuje szybkość zmian, tym samym po jej wyznaczeniu otrzymujemy sygnał stanowiący składową będącą szumem. Aby wyznaczyć pochodną sygnału, należy posłużyć się operatorem *Differentiate* dostępnym w *Series -> Data Transformation*. Do poprawnego funkcjonowania tego operatora potrzebne jest jedynie wybranie odpowiedniego atrybutu.

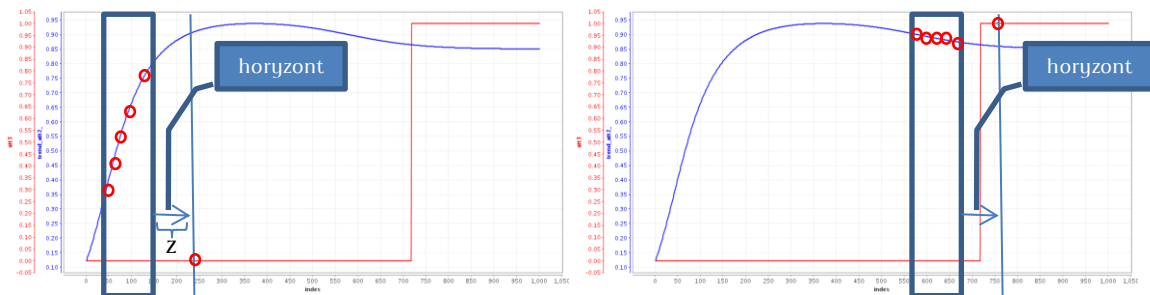
Zaobserwuj wyznaczony wykres pochodnej (i pokaż go w sprawozdaniu) oraz wyznacz poziom szumu korzystając z operatorów z poprzednich zadań. Pokaż wykres pokazujący poziom estymowanej wartości szumu.

Budowa modelu predykcyjnego

Dysponując wcześniej przygotowanym zbiorem zawierającym przetworzony sygnał można przejść do kolejnego etapu, jakim jest zbudowanie modelu predykcyjnego. Idea budowy takiego modelu bazuje ponownie na okienkowaniu, gdyż algorytmy uczące się do swojego działania wymagają danych zapisanych w postaci tabeli o stałej liczbie kolumn. Odpowiedni system buduje się więc tak, iż po określeniu rozmiaru okna oznaczonego symbolem k , dysponując dwoma (lub jednym sygnałem), naszym celem jest znalezienie funkcji

$$C_{n+z}=f(x_n,x_{n-1},\dots,x_{n-k})$$

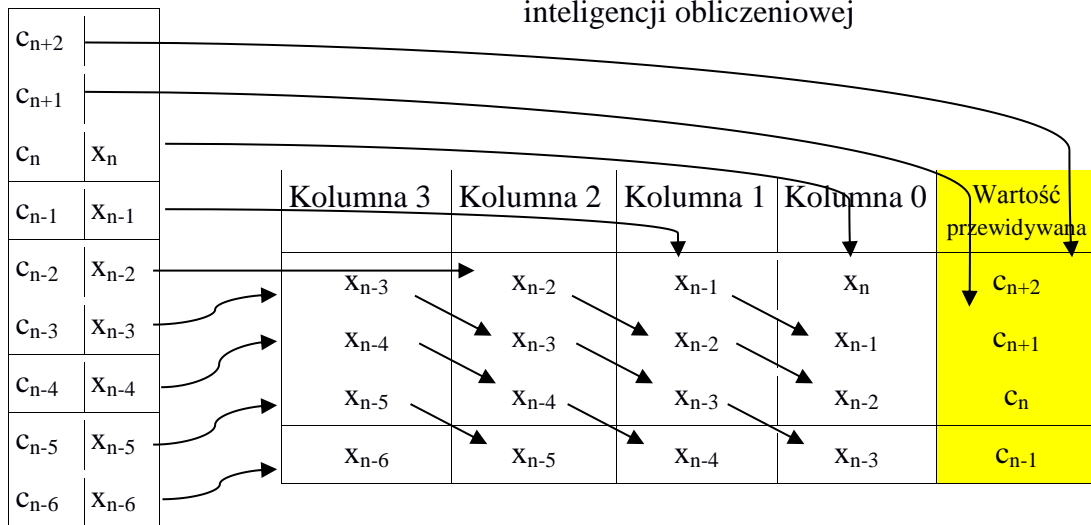
Która stara się przewidzieć wartość C w chwili $n+z$, gdzie z nazywane jest horyzontem predykcji, czyli określa na ile w przód chcemy dokonać predykcji. Całość pokazuje poniższy wykres



Gdzie na podstawie sygnału niebieskiego próbuje się przewidzieć wartość sygnału czerwonego. Na przedstawionym przykładzie wartość historii, czyli wartość parametru k określona jest na $k=5$, natomiast wartość przewidywana jest wyznaczana z określonym przesunięciem wprzód oznaczonym jako *horyzont*

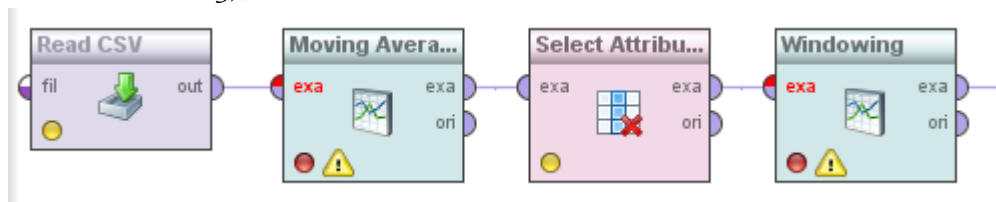
Innymi słowy naszym celem jest dokonanie takiej transformacji danych, który można pokazać w postaci tabeli jak poniżej:

Seria danych Seria danych po przygotowaniu do uczenia z wykorzystaniem metod inteligencji obliczeniowej



Aby dokonać powyższej transformacji konieczne jest:

- 1) Wczytanie danych z pliku
- 2) Zastosowanie jednej z wyżej opisanych metod estymacji poziomu szumu w sygnale
- 3) Zastosowanie selekcji atrybutów tak, aby pozostawić jedynie konieczne atrybuty – atrybut zawierający podstawową informację czyli w naszym przypadku jeden atrybut zawierający estymowany poziom szumu oraz atrybut zawierający wartość, którą chcemy przewidywać -> selekcji atrybutów dokonujemy za pomocą operatora *Select Attributes* i wybieramy atrybuty *att3* (atrybut wyjściowy) oraz atrybut powstały jako estymata poziomu szumu (patrz zajęcia z podstaw RapidMinera)
- 4) Na koniec, tak przygotowane dane podajemy na wejście operatora *Windowing* (*Series -> Windowing*),

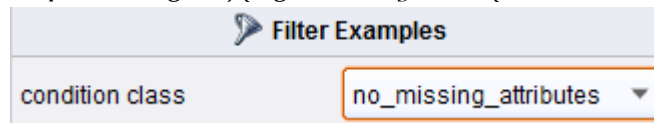


gdzie konieczne jest w procesie jego konfiguracji wybranie odpowiedniego atrybutu stanowiącego wyjście systemu oraz określenie wielkości horyzontu predykcji

label attribute	att3
horizon	10

- 5) W kolejnym kroku na wyjście tak przygotowanego układu należy podłączyć do model predykcyjnego pozwalającego nam na przewidywanie pożądanej wartości. W naszym przypadku modelem predykcyjnym powinien być model klasyfikacyjny konieczne jest więc dokonanie małej transformacji zamieniającej kolumnę przewidywaną na zbiór wartości symbolicznych. Realizujemy to za pomocą operatora *Numerical to Polynomial* lub *Numerical to Binomial* dostępnych w *Data transformation -> Type Conversion*. (Uwaga należy zaznaczyć opcję *include special*

attributes). Należy zwrócić uwagę iż tak powstały zbiór danych będzie zawierał wartości brakujące z którymi zwykle modele predykcyjne nie potrafią sobie poradzić, dlatego też wskazany jest aby je usunąć. W tym celu należy wykorzystać operator *Filter examples* konfigurując go tak aby usunął wartości brakujące:



- Na koniec takiego układu podłączamy model predykcyjny np. sieć neuronową
- 6) Ostatnim krokiem jest sprawdzenie czego system się nauczył, w tym celu dokonujemy analizy poprawności nauczania sieci neuronowej dla uproszczenia testując ją na tym samym zbiorze danych. W tym celu wykorzystujemy operator *Apply Model* oraz operator *Performance* w celu estymacji dokładności predykcji

