

Wczytywanie danych z plików

1. Dane w pliku gdzie pola oddzielonym przecinkiem, a w pierwszym wierszu są nazwy atrybutów
2. Plik zawiera dane demograficzne, takie jak wiek, dochód, płeć, stan cywilny, liczbę dzieci oraz informację o grupie ryzyka kredytowego (good risk, bad profit and bad loss).
 - a. w pliku są pola zawierające wartości ciągłe (np: id, wiek, dochód) jak i katagoryczne (np: płeć, stan cywilny, ryzyko).
3. Wstaw węzeł *Plik zmiennych* i wybierz plik o nazwie *SmallSampleComma.txt*
4. Pojawiają się dane z kilku pierwszych linijek
5. Nazwy atrybutów będą brane z pierwszego wiersza, bo jest zaznaczona opcja: *Odczytaj nazwy zmiennych z pliku*:
6. Kliknij zakładkę *Dane* (czyta pierwszych 50 linii i przedstawia raport):
 - a. *Składowanie* przedstawia, w jaki sposób dane są przechowywane i jak będą traktowane przez system (np. całkowita lub rzeczywista będą traktowane jako ciągłe wartości przez węzły modelujące) – jeżeli chcemy żeby były inaczej traktowane to musimy to zmienić zaznaczając *Nadpisz* i zmieniając sposób przechowywania na *String*
 - b. Jako *Składowanie* może być: *łańcuch, liczba całkowita, liczba rzeczywista, czas, data, znacznik czasu, lista, nieznane*.
7. Kliknij zakładkę *Filtrowanie* aby zmienić nazwę zmiennej albo w ogóle ją wyeliminować
 - a. W *Zmienna* z lewej strony są nazwy zmiennych odczytane z pliku a po prawej można dać im nowe nazwy (zmień ID na IDNumber, kliknij strzałkę przy Marital i ją zaznacz)
 - b. Można sortować, itd.
8. Kliknij zakładkę *Typy*
 - a. *Poziom pomiaru* wskazuje jak dane będą traktowane w czasie przekształceń i modelowania (np. liczba całkowita będzie ilościowa, łańcuchy – jakościowe)
 - b. Wciśnij *Odczytaj wartości*
9. Pierwsze sprawdzenie danych
 - a. Dodaj węzeł *Tabela* aby sprawdzić czy mamy dobre dane i uruchom
 - b. Zobacz, że w nagłówku są informacje o ilości danych, zwróć uwagę a wartość \$null\$ (sprawdź w danych, że tam jest faktycznie "")
10. Wczytywanie innych typów tych samych danych (.sav – plik „statistics”, czy Excel – ustawić „typ pliku” = „Excel 97-2003” i bez zaznaczania „pierwszy wiersz zawiera nazwy zmiennych”)