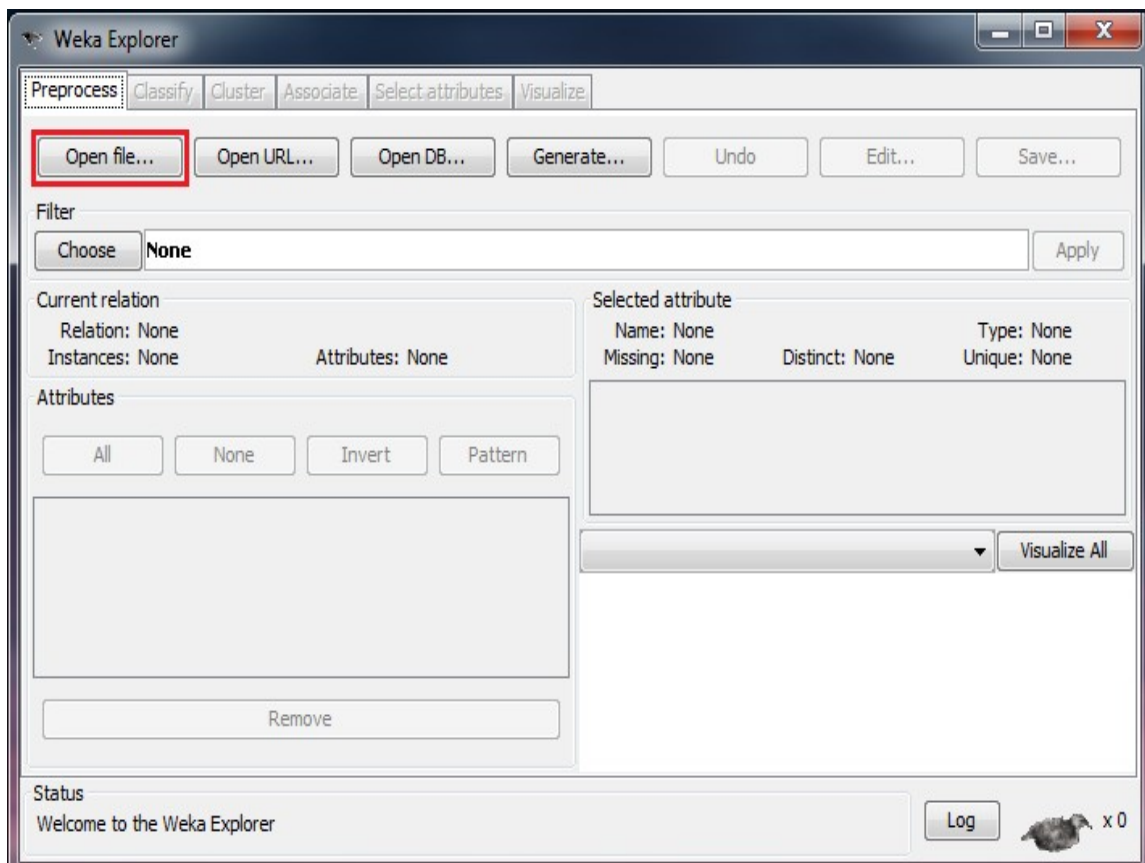


Instrukcja obsługi programu WEKA, czyli jak przestałem się martwić i pokochałem uczenie maszynowe

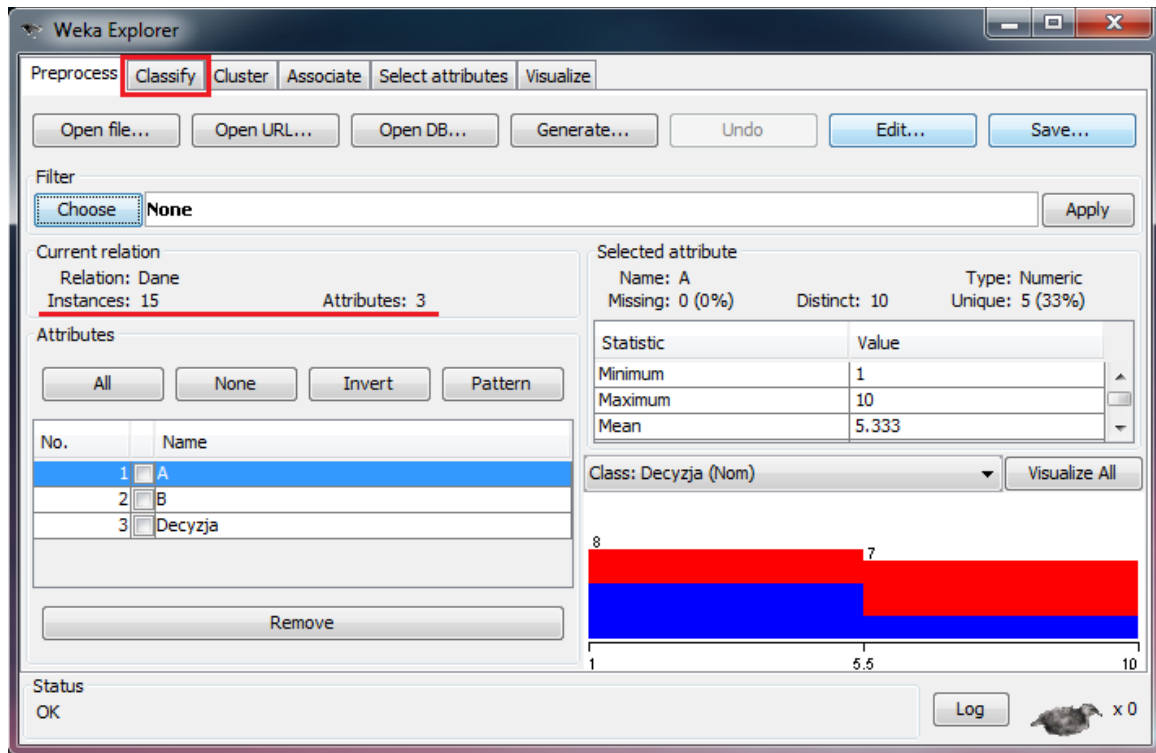
1. Sprawdź, czy posiadasz już w systemie zainstalowany program WEKA. Jeżeli tak, uruchom go. Jeżeli nie, to pobierz go z: <https://www.cs.waikato.ac.nz/ml/weka/downloading.html>, rozpakuj archiwum i uruchom *weka.jar*.
2. Powinno pokazać się okno powitalne. Kliknij w przycisk 'Explorer'.



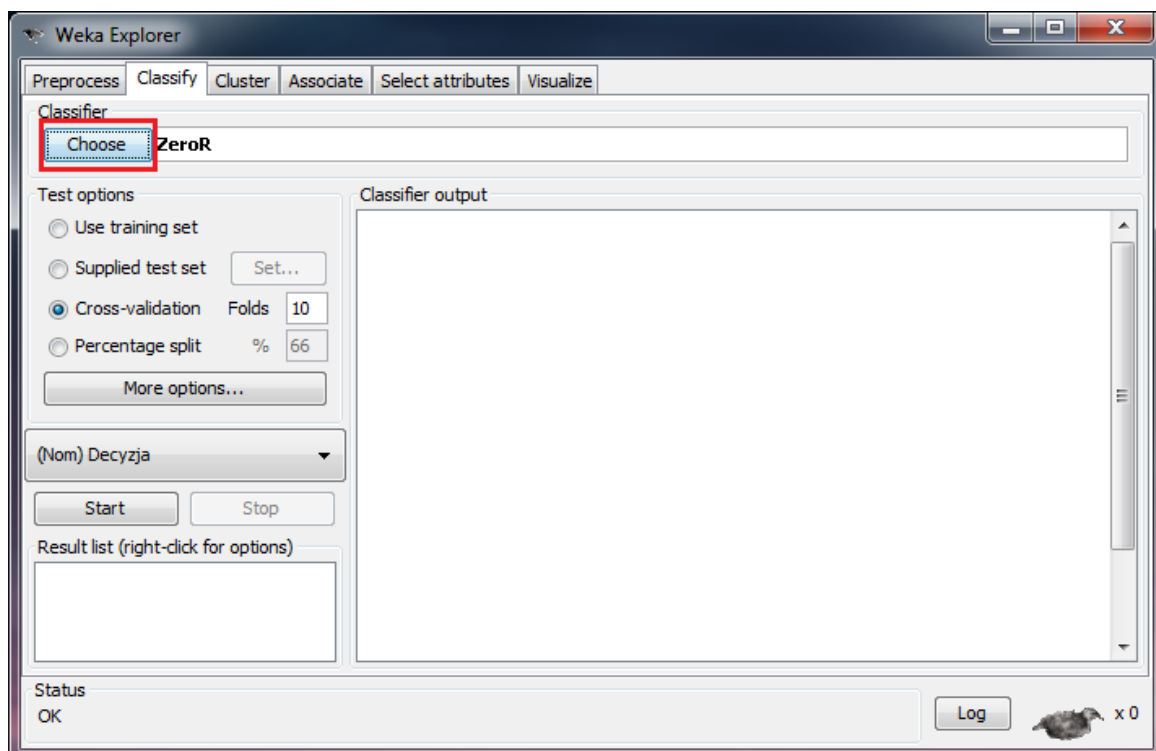
3. Wyświetli się panel zawierający informacje o zbiorze danych. Aktualnie żaden zbiór nie został jeszcze wczytany. Kliknij na 'Open file...' i wczytaj plik *zad3.arff* z katalogu *dane* ([link](#)).



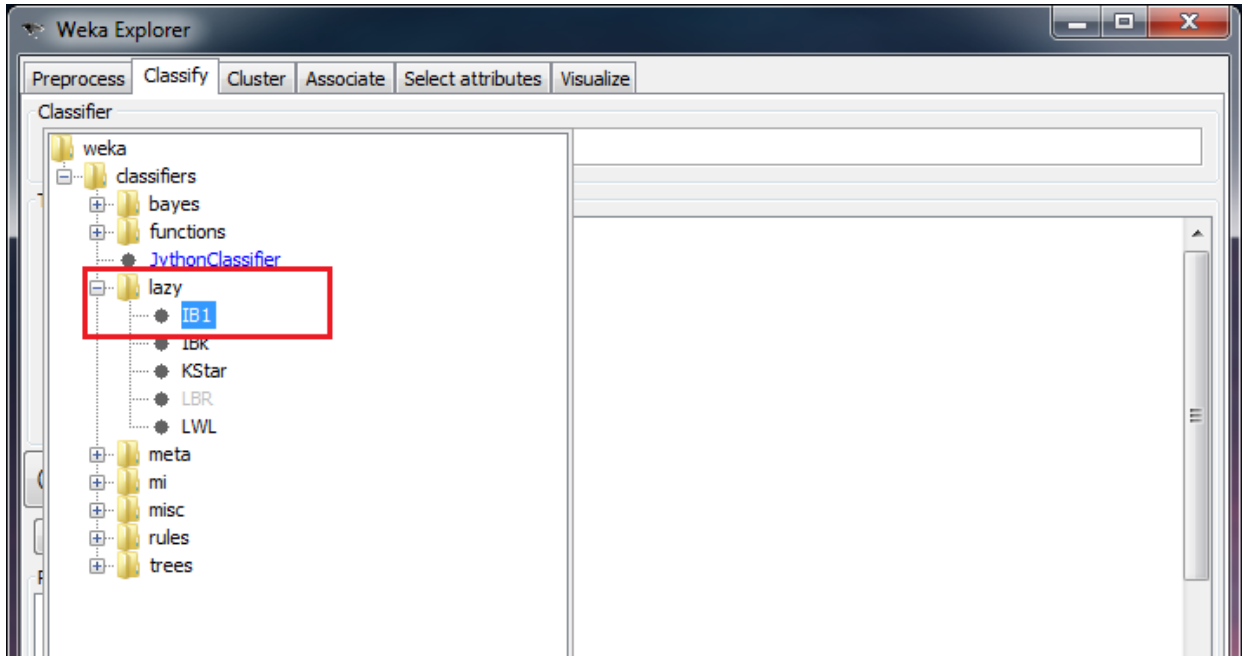
- Wyświetlą się podstawowe informacje i statystyki dotyczące wczytanego zbioru danych. W tym wypadku dane to 15 przykładów opisanych trzema atrybutami, z których ostatni (*Decyzja*) WEKA domyślnie traktuje jako atrybut decyzyjny.



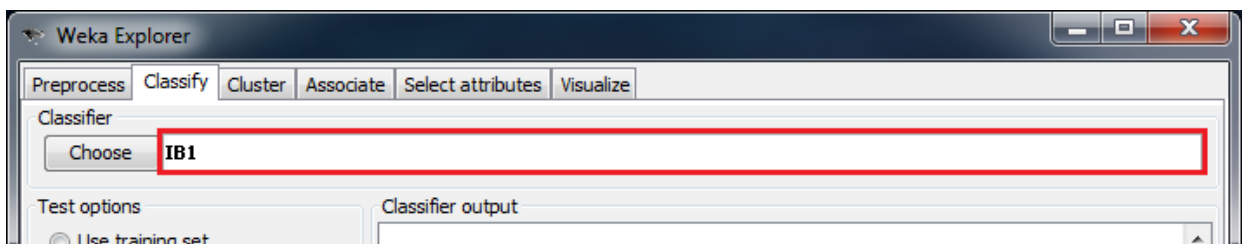
- Spróbujmy zastosować na tych danych klasyfikator **k-NN**. W tym celu przechodzimy do zakładki 'Classify'.
- Zmieniamy algorytm uczenia klikając na 'Choose'.



- Wybieramy *classifiers/lazy/IBk*, czyli implementację algorytmu **k-NN** w WEKA. **k-NN** jest leniwym algorytmem uczenia się w tym sensie, że „uczenie się” jest odroczone do momentu klasyfikacji konkretnego przypadku testowego. Pozostałe algorytmy uczenia (np. drzewa decyzyjne) najpierw na podstawie zbioru uczącego tworzą hipotezę jak najtrafniej tłumaczącą dane z tego zbioru, a następnie wykorzystują ją do klasyfikacji nowych danych. Uczenie się jest więc w ich przypadku dokonywane przed rozpoczęciem klasyfikacji jakiegokolwiek przykładu.



- Jednym z mniej intuicyjnych elementów interfejsu WEKA jest ustalanie parametrów algorytmu. W tym celu trzeba kliknąć na pole z nazwą. Można tam ustawić m.in. parametr k algorytmu **k-NN**, który określa liczbę sąsiadów branych pod uwagę.



- Spróbujmy użyć IBk do klasyfikacji naszych danych. Nie zmieniając nic w opcjach klikamy na 'Start'. W polu tekstowym opatrzonym etykietą 'classifier output' możemy zobaczyć wyniki działania algorytmu. Najważniejsze informacje dotyczące ich interpretacji przedstawione są na następnej stronie.

Interpretacja wyniku klasyfikacji:

=== Run information ===

Scheme: weka.classifiers.lazy.IB1
Relation: Dane
Instances: 15
Attributes: 3
A
B
Decyzja

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

IB1 classifier

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	5	33.3333 %
Incorrectly Classified Instances	10	66.6667 %
Kappa statistic	-0.3393	
Mean absolute error	0.6667	
Root mean squared error	0.8165	
Relative absolute error	131.1475 %	
Root relative squared error	160.253 %	
Total Number of Instances	15	

Widzimy, że algorytm poprawnie przydzielił klasę decyzyjną w **33,3333 %** przypadków. Liczbę tę nazywamy **trafnością klasyfikacji**.

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.286	0.625	0.286	0.286	0.286	0.33	0
	0.375	0.714	0.375	0.375	0.375	0.33	1
Weighted Avg.	0.333	0.673	0.333	0.333	0.333	0.33	

=== Confusion Matrix ===

a b <-- classified as

2 5 | a = 0

5 3 | b = 1

Macierz pomyłek. Możemy z niej wyczytać, że przykłady klasy **0** zostały poprawnie rozpoznane **2 razy**, a klasy **1** zostały poprawnie rozpoznane **3 razy**.