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: <u>https://www.cs.waikato.ac.nz/ml/weka/downloading.html</u>, 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* (<u>link</u>).

🐑 Weka Explorer		
Preprocess Classify Cluster Associate Select attributes Visualiz	2	
Open file Open URL Open DB Gene	undo	Edit Save
Choose None		Apply
Current relation Relation: None Instances: None Attributes: None	Selected attribute Name: None Missing: None Disti	Type: None inct: None Unique: None
Attributes All None Invert Pattern Remove		▼ Visualize All
Status Welcome to the Weka Explorer		Log 💉 x 0

4. 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.

🐨 Weka Explorer				
Preprocess Classify Cluster Associate Select attributes Visualize				
Open file Open URL Open DB Generate Undo Edit Save				
Filter				
Choose None Apply				
Current relation Relation: Dane Instances: 15 Attributes: 3	Selected attribute Name: A Missing: 0 (0%) Distinct: 10	Type: Numeric Unique: 5 (33%)		
Attributes	Statistic Value			
All None Invert Pattern	Minimum 1	A		
	Maximum 10			
No. Name	Mean 5.333			
1 🗖 A	Class: Decyzja (Nom)	▼ Visualize All		
2 B	<u></u>			
3 Decyzja	87			
	· · · · · · · · · · · · · · · · · · ·			
Remove				
	1 5.5	10		
OK		Log 💉 x 0		

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

🕆 Weka Explorer		x
Preprocess Classify Cluster Associate Select attributes Visualize		
Classifier		
Choose ZeroR		
Test options Classifier output		
O Use training set		<u>^</u>
Supplied test set Set		
Cross-validation Folds 10		
Percentage split % 66		
More options		=
(Nom) Decyzja		
Start Stop		
Result list (right-click for options)		
OK Log		x0
	_ ~	P

7. 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.

💎 Weka Explorer	
Preprocess Classify Cluster Associate Select attributes	/isualize
Classifier	
📔 weka	
JythonClassifier	
🚊 🔒 lazy	
• IB1	
IBK	
• LWL	E
📄 🖶 🛄 meta	
🛛 🌔 🌐 🕌 mi	

8. 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.

🐑 Weka Explorer		
Preprocess Classify Cluster	Associate Select attributes Visualize	
Classifier		
Choose IB1		
Test options	Classifier output	
 Use training set 		

9. 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 Α В Decyzja Widzimy, że algorytm poprawnie przydzielił klasę decyzyjną w 33,3333 % Test mode: 10-fold cross-validation przypadków. Liczbę tę nazywamy trafnością klasyfikacji. === Classifier model (full training set) === IB1 classifier Time taken to build model: 0 seconds === Stratified cross-validation === === Summary === 33.3333 % **Correctly Classified Instances** 5 **Incorrectly Classified Instances** 66.6667 % 10 Kappa statistic -0.3393Mean 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 === Detailed Accuracy By Class === TP Rate FP Rate Precision Recall F-Measure ROC Area Class 0.625 0.286 0.33 0.286 0.286 0.286 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 25 | 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**.