

WEKA

Learning Outcomes

- Upon successful completion of this lab, you will have demonstrated the abilities to:
 - Build and compare different models using WEKA

Instructions:

1. Install and take a look at the WEKA tutorial [here](https://www.cs.waikato.ac.nz/ml/weka/) [\(https://www.cs.waikato.ac.nz/ml/weka/\)](https://www.cs.waikato.ac.nz/ml/weka/).
2. Download the following data sets [here](https://learn.ontariotechu.ca/courses/19275/files/2375504?wrap=1). <https://learn.ontariotechu.ca/courses/19275/files/2375504?wrap=1> ↓ https://learn.ontariotechu.ca/courses/19275/files/2375504/download?download_frd=1
 - The zip file contains three datasets in [.arff](https://www.cs.waikato.ac.nz/ml/weka/arff.html#:~:text=An%20ARFF%20(Attribute%2DRelation%20File,the%20Weka%20machine%20learning%20software.)) [\(https://www.cs.waikato.ac.nz/ml/weka/arff.html#:~:text=An%20ARFF%20\(Attribute%2DRelation%20File,the%20Weka%20machine%20learning%20software.\)](https://www.cs.waikato.ac.nz/ml/weka/arff.html#:~:text=An%20ARFF%20(Attribute%2DRelation%20File,the%20Weka%20machine%20learning%20software.)) format. They are ready to be used with WEKA.

Part I:

This first task is to conduct classification tasks with WEKA and compare the performance of several common methods.

1. Use all three datasets in the zip file (letter.arff, segment.arff, and waveform-5000.arff).
2. Run the classification task using IBk (k-NN), J48, AdaboostM1(J48) and NaiveBayes.
3. Set most parameters to their default values, except for the following key parameters.
 1. IBk: K=1/3/5;
 2. J48: M=2/4;
 3. AdaboostM1(J48): M=2;
 4. NaiveBayes: all default.
4. In your report, record the prediction accuracy of the classification task, as in Table 1, 2, and 3.

Table 1. Performance evaluation of IBk Classification.

Table 1. Accuracy of IBk Classification.

Dataset	K =1	K=3	K=5
letter			
segment			
waveform-5000			

Table 2. Accuracy J48 Classification.

Dataset	J48 (M=2)	J48 (M=4)	AdaboostM1+J48
letter			
segment			
waveform-5000			

Table 3. Accuracy of NaiveBayes Classification.

Dataset	Accuracy (%)
letter	
segment	
waveform-5000	

Part II:

This second task is to conduct clustering tasks with WEKA and compare the performance k-means clustering algorithm:

1. Use all three datasets in the zip file (letter.arff, segment.arff, and waveform-5000.arff).
2. Run the clustering using SimpleKMeans algorithm; Choose different Ks as follows:
 - **letter** (K1 = 11, K2 = 24 , K3= 38)
 - **segment** (K1 = 3, K2 = 5 , K3 = 10)
 - **waveform-5000** (K1 = 2, K2 = 3, K3= 5)
3. In your report, record the SSE for doing the clustering task, as in Table 4.

Table 4. SSE of SimpleKMeans Clustering.

Dataset	K = K1	K = K2	K = K3
letter			
segment			
waveform-5000			

Report:

1. Your report should have a cover letter including the group member names
2. Organize all your *tables* and *interpretations* in your lab report (*PDF format*)