

**WQD7006**  
**Machine Learning for Data Science**

**Introduction to Weka**  
**Introduction to Naïve Bayes**

**File/Data loading and conversion:**

1. Converting .csv file to .arff (Attribute-Relation File Format) – ASCII format for text
2. Open Weka – Tools – ArrfViewer – open HeartRate.csv
3. Attribute values can be changed by clicking on each field
4. Save as – arff file

**OR**

1. Open .csv file in notepad/text editor
2. Define attributes:

HeartRate.csv

@relation HeartRate ----Name of dataset

@attribute hrv numeric (or real)

@attribute hr numeric

@attribute dayID numeric

@attribute sequenceID { 1of4,2of4,3of4,4of4 }

@attribute patternSleep { Irregular,Regular }

@attribute hoursAwake numeric

@attribute healthGrade nominal

@data

70.4,75.2,1,1of4,Irregular,0,1

65,84.1,1,2of4,Irregular,3,0

74.4,73.8,1,3of4,Irregular,4,1

75.1,72.7,1,4of4,Irregular,15,1

65.2,91.2,2,1of4,Irregular,0,0

52.9,98.9,2,2of4,Irregular,4,0

3. Save as .arff
4. Open in Weka

### **Descriptive Statistics**

1. Get to know the data
2. Filtering attributes, relation, editing etc.

### **Visualization**

1. Check each variables
2. Click on Visualize to check plots
3. Change size of plots etc.

### **Sample Weka Datasets**

**Binary classification** – Breast Cancer, and PIMA diabetes (Class must be nominal)

**Multiclass** – Iris, Soybean (Class must be nominal)

### **Classification**

#### **Naïve Bayes**

Answer the following questions using HeartRate data.

1. Describe the dataset?
2. How many variables are there? 7
3. Describe Pattern of Sleep and Health Grade
4. Perform Naïve Bayes using the default setting (Tab – Classify – Locate Naïve Bayes).  
Describe the relevant results – report accuracy, precision, recall and F-measure.
5. We would like to see if Health grade can be classified using only hrv and pattern sleep. (Hint: remove the rest of the attributes). Perform Naïve Bayes using default setting. Explain your results.

