# DWM Practical 7 (Part 2)

|  |
| --- |
| Use WEKA to implement classification |

**Title - Classification using the Weka toolkit – Part 2**

**Aim - To perform classification on data sets using the Weka machine learning toolkit**

**Requirements**

1. **Load the ‘weather.nominal.arff’ dataset into Weka and run Id3 classification algorithm.**

* **List the attributes of the given relation along with the type details**
* **Create a table of the weather.nominal.arff data**
* **Study the classifier output and answer the following questions**

**Answer the following questions**

1. **Draw the decision tree generated by the classifier**
2. **Compute the entropy values for each of the attributes**
3. **What is the relationship between the attribute entropy values and the nodes of the decision tree?**
   1. **Draw the confusion matrix? What information does the confusion matrix provide?**
   2. **Describe the Kappa statistic?**
   3. **Describe the following quantities:** 
      * **TP Rate**
      * **FP Rate**
      * **Precision**
      * **Recall**
4. **Load the ‘weather.arff’ dataset in Weka and run the Id3 classification algorithm. What problem do you have and what is the solution?**
5. **Load the ‘weather.arff’ dataset in Weka and run the OneR rule generation algorithm. Write the rules that were generated.**
6. **Load the ‘weather.arff’ dataset in Weka and run the PRISM rule generation algorithm. Write down the rules that are generated.**
7. **Load the glass.arff dataset and perform the following tasks?**
   * **How many items are there in the dataset?**
   * **List the attributes are there in the dataset.**
   * **List the classes in the dataset along with the count of instances in the class.**
   * **How will you determine the color assigned to each class?**
   * **By examining the histogram, how will you determine which attributes should be the most important in classifying the types of glass?**
8. **Perform the following classification tasks:**
   * **Run the 1Bk classifier for various values of K?**
   * **What is the accuracy of this classifier for each value of K?**
   * **What type of classifier is the 1Bk classifier?**
9. **Perform the following classification tasks:**
   * **Run the J48 classifier**
   * **What is the accuracy of this classifier?**
   * **What type of classifier is the J48 classifier?**
10. **Compare the results of the 1Bk and the J48 classifiers. Which is better?**
11. **Run the J48 and 1Bk classifiers using** 
    * **the cross-validation strategy with various fold levels. Compare the accuracy results.**
    * **holdout strategy with three percentage levels. Compare the accuracy results.**
12. **Perform following tasks:**
    1. **Remove instances belonging to the following classes:**
       * **build wind float**
       * **build wind non-float**
    2. **Perform classification using the 1Bk and J48 classifiers. What is the effect of this filter on the accuracy of the classifiers?**
13. **Perform the following tasks:**
    1. **Run the J48 and the NaiveBayes classifiers on the following datasets and determine the accuracy:**
       1. **vehicle.arff**
       2. **kr-vs-kp.arff**
       3. **glass.arff**
       4. **wave-form-5000.arff**

**On which datasets does the NaiveBayes perform better? Why?**

1. **Perform the following tasks**
   * **Use the results of the J48 classifier to determine the most important attributes**
   * **Remove the least important attributes**
   * **Run the J48 and 1Bk classifiers and determine the effect of this change on the accuracy of these classifiers. What will you conclude from the results?**