ASSIGNMENT: #2

MACHINE LEARNING

SUBMITTED BY:

Tayyaba Amber (SP24-RCS-011)

SUBMITTED TO:

DR. MUHAMMAD SHARJEEL

DATE OF SUBMISSION:

6th OCTOBER 2024

COMSATS University
Islamabad
Lahore Campus

Task no 1&2:

- ## Hand-crafted features & ML experiments in WEKA:

Hand-crafted features can be manually engineered from the raw data. These features are designed to capture important data characteristics that might help the classifier make more accurate predictions.

Here is a table that summarizes the classifier output:

Feature:

- Name of length
- First letter of name
- Last letter of the name
- Total no of vowels in a name

classifier	Test Mode	Classifier Output		
Tree.j48	Split 80 %	Correctly Classified Instances Incorrectly Classified Instances	14 2	87.5 % 12.5 %
Tree RandomForest	split 80 %	Correctly Classified Instances Incorrectly Classified Instances	13 3	81.25 % 18.75 %
Tree REPTree	Split 80%	Correctly Classified Instances Incorrectly Classified Instances	14 2	87.5 % 12.5 %
Tree.REPTree	Training data	Correctly Classified Instances Incorrectly Classified Instances	73 7	91.25 % 8.75 %
Lazy.IBK	Training data	Correctly Classified Instances Incorrectly Classified Instances	76 4	95 % 5 %
Tree. RamdomForest	Training data	Correctly Classified Instances Incorrectly Classified Instances	76 4	95 % 5 %

Feature:

- First letter of name
- Last letter of the name
- Total no of vowels in a name
- Total length of name
- Total no of consonant
- Name length even or odd
- Total no of A's
- Length is up to 4 or not
- Start with a vowel or not

classifier	Test Mode	Classifier Output		
Tree RandomForest	Split 66%	Correctly Classified Instances Incorrectly Classified Instances	25 2	92.592 % 7.4074 %
Tree RandomForest	Training dara	Correctly Classified Instances Incorrectly Classified Instances	79 1	98.75 % 1.25 %
Lazy.IBK	Training data	Correctly Classified Instances Incorrectly Classified Instances	79 1	98.75 % 1.25 %
Lazy.IBK	Split 66%	Correctly Classified Instances Incorrectly Classified Instances	25 2	92.592 % 7.4074 %
Tree j48	Split 66%	Correctly Classified Instances Incorrectly Classified Instances	26 6	81.75 % 18.25 %

Task no 3:

Write a paragraph about your experience of working with the standard ML pipeline in your own words.

Step 1: Data preparation

Firstly, I made an ARFF file, which are text file that describes data set, including attributes and instances.

Step 2: Build a classifier

Once the data is ready, next I select the classifier and build the model.

Weka provides some classifiers like decision tree (j48, Random forest) and lazy (IBK).

Step 3: Evaluating the model

It describes the performance of your model:

- Train data split
- Cross-validation
- Training data

Conclusion:

I faced from the analysis is that **Random Forest** consistently outperformed the other classifiers, achieving the highest accuracy (92%) in the test. This indicates that Random Forest is better suited for this particular dataset, likely due to its ability to handle feature variability and avoid overfitting compared to simpler models like J48 and REPTree.