ASSIGNMENT 02

Machine Learning

Junaid Ali Bhatti
FA22-RCS-005

Characteristics of the dataset

A notable attribute is "*starts_with_vowel*", which has been found to be a **prominent feature** for classification in the data. This feature alone is enough to exactly classify 86 instances correctly.

J48 Classification Results

Correctly Classified Instances: 94 out of 100 (94%)

Incorrectly Classified Instances: 6 out of 100 (6%)

Precision: For class 1, precision is 0.978; for class 0, precision is 0.907.

F-Measure: Both classes have high F-measure values (0.938 for 1 and 0.942 for 0).

Confusion Matrix:

- Class 1: 45 correct, 5 incorrect.
- Class 0: 49 correct, 1 incorrect.

Run Information of J48 in Weka

```
Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     names-weka.filters.unsupervised.attribute.Remove-R1
Instances:    100
Attributes:   10
              length
              vowel_count
              consonant_count
              starts_with_vowel
              first_alphabet
              second_alphabet
              last_alphabet
              contains_double_letters
              contains_repeated_letters
              label
Test mode:    10-fold cross-validation
```

Experience with the ML Pipeline

It was really a very smooth experience working with the standard pipeline of the machine learning model in WEKA. Loading the dataset was easy, and using the J48 classifier was insightful and easy to understand. Probably, one strength of WEKA is the facility to allow users to train a model without necessarily having to code; hence, it makes it much easier. Built-in Evaluation metrics provided a sense of the model's performance as it enabled me to identify features and observe how they contribute to classification. The number of features that were critical in the outcome was determined by the J48 algorithm, hence showing which features,

for example, if whether a name started with a vowel mattered for the predicted outcomes. Overall, the attributes within WEKA proved easy, even for beginners, to understand, given the friendly interface and clarity of the results.