

Experiment 6: Feature Selection to Improve Model Performance

Aim:

To select **prominent feature subsets** from the Iris dataset to improve **classification model performance**.

Theory:

- **Feature selection** identifies the **most relevant attributes** for predicting the class.
 - Reduces **overfitting**, improves **accuracy**, and decreases **training time**.
 - WEKA provides **attribute selection tools** like **InfoGainAttributeEval + Ranker**.
-

Dataset (iris.arff)

```
@relation iris
```

```
@attribute SepalLength numeric  
@attribute SepalWidth numeric  
@attribute PetalLength numeric  
@attribute PetalWidth numeric  
@attribute Species {Setosa, Versicolor, Virginica}
```

```
@data
```

```
5.1,3.5,1.4,0.2,Setosa  
4.9,3.0,1.4,0.2,Setosa  
6.2,3.4,5.4,2.3,Virginica  
5.9,3.0,5.1,1.8,Virginica  
6.0,2.2,4.0,1.0,Versicolor  
5.5,2.3,4.0,1.3,Versicolor  
...
```

Procedure (Using WEKA):

1. Open **WEKA → Explorer**.

2. Click **Open File** → select **iris.arff**.
 3. Go to **Select Attributes tab**.
 4. Choose **Attribute Evaluator** → **InfoGainAttributeEval**.
 5. Choose **Search Method** → **Ranker**.
 6. Click **Start** → WEKA ranks attributes based on importance.
 7. Identify the **most prominent features** for classification.
-

Result (Sample / Expected):

Attribute Importance (Info Gain)

PetalLength 0.9

PetalWidth 0.88

SepalLength 0.5

SepalWidth 0.3

- **Prominent Features:** PetalLength and PetalWidth → most useful for predicting Species.
-

Conclusion:

- Selecting **PetalLength and PetalWidth** improves model accuracy and reduces complexity.
- Feature selection helps in **better performance and faster training**.
- WEKA provides a simple interface to **rank and select attributes**