

Experiment 4: Prediction Using k-Nearest Neighbor (k-NN) Classification

Aim:

To predict the **Result** of a new student record using **k-Nearest Neighbor (k-NN)** and evaluate the model accuracy.

Theory:

- k-NN is a **lazy supervised learning algorithm**.
 - Predicts the class of a new instance based on the **majority class of its k nearest neighbors**.
 - Distance metric (e.g., **Euclidean distance**) is used to find nearest neighbors.
 - Accuracy can be evaluated using **cross-validation**.
-

Dataset (student.arff)

```
@relation student
```

```
@attribute Attendance numeric  
@attribute InternalMarks numeric  
@attribute AssignmentScore numeric  
@attribute SemesterMarks numeric  
@attribute Result {Pass, Fail}
```

```
@data
```

```
80,75,70,85,Pass  
60,65,60,70,Pass  
50,55,50,45,Fail  
90,80,85,90,Pass  
70,60,65,60,Pass  
45,50,55,50,Fail  
85,75,80,88,Pass
```

Class Attribute: Result (Pass/Fail)

New Test Instance: 65, 70, 60, 68, ?

Procedure (Using WEKA):

1. Open **WEKA** → **Explorer**.
 2. Click **Open File** → select **student.arff**.
 3. Go to **Classify tab**.
 4. Choose **Classifier** → **lazy** → **IBk (k-NN)**.
 5. Set **k = 3** (or any suitable value).
 6. Click **Start** to train the model and evaluate accuracy.
 7. Use the trained model to **predict Result** for the new test instance.
-

Result (Sample / Expected):

- **Predicted Result:** Pass
 - **Model Accuracy:** 100% (all instances correctly classified in this dataset)
-

Conclusion:

- k-NN accurately predicted the Result of a new student based on **nearest neighbors**.
- WEKA provides a **quick way** to implement and test k-NN classification.
- Accuracy depends on **k value** and dataset distribution.