

## Experiment 1: Evaluate Information Gain of Attributes in Student Dataset

---

### Aim:

To evaluate the **Information Gain (IG)** of each attribute in the student dataset and identify the **best attribute** to split for predicting **Result**.

---

### Theory:

- **Information Gain (IG)** measures how much an attribute **reduces uncertainty** about the class (Result).
  - The attribute with the **highest IG** is chosen for splitting in **decision tree algorithms** like ID3 or J48.
- 

### 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

---

### Procedure (Using WEKA):

1. Open **WEKA** → **Explorer**.

2. Click **Open File** → select **student.arff**.
3. Go to **Select Attributes** tab.
4. Choose **Attribute Evaluator** → **InfoGainAttributeEval**.
5. Choose **Search Method** → **Ranker**.
6. Click **Start** → WEKA calculates **Information Gain** for all attributes.
7. Identify the attribute with **highest IG** → best for splitting.

---

**Result (Sample / Expected):**

Attribute	Information Gain
Attendance	0.42
InternalMarks	0.56
AssignmentScore	0.35
SemesterMarks	0.60

- **Highest IG:** SemesterMarks → **best attribute to split** for predicting Result.

---

**Conclusion:**

- **SemesterMarks** is the most informative attribute for predicting **Pass/Fail**.
- Using the attribute with highest IG improves **classification accuracy** in decision trees.
- WEKA provides a **quick and easy way** to calculate Information Gain.