Question 5: Confusion Matrix & Macro-Averaged Metrics

Confusion Matrix for Figure 1

The clusters from Figure 1 contain:

- Cluster 1: 4 Red, 3 Purple, 3 Blue
- Cluster 2: 2 Red, 2 Purple, 5 Blue
- Cluster 3: 4 Purple, 2 Red, 2 Blue

The confusion matrix for this is:

Cluster	Red	Purple	Blue
1	4	3	3
2	2	2	5
3	2	4	2

Macro-Averaged Precision, Recall & F1-score

1) Precision calculation

Precision = True Positives / (True Positives + False Positives)

Precision for each class is calculated as:

- **Red Class:** TP = 4, FP = 2 + 2 = 4
  - Precision = 4 / (4+4) = 0.5
- **Purple Class:** TP = 3, FP = 2 + 4 = 6
  - Precision = 3 / (3+6) = 0.333
- **Blue Class:** TP = 3, FP = 5 + 2 = 7
  - Precision = 3 / (3+7) = 0.3

Macro – averagePrecision = (0.5 + 0.333 + 0.3) / 3 = 0.3777

2) Recall Calculation

Recall = True Positives / (True Positives + False Negatives)

Recall for each class is calculated as:

- Red Class: TP = 4, FN = 2 + 2 = 4
  - Recall = 4 / (4+4) = 0.5
- Purple Class: TP = 3, FN = 2 + 4 = 6
  - Recall = 3 / (3+6) = 0.333
- Blue Class: TP = 3, FN = 5 + 2 = 7

     o   Recall = 3 / (3+7) = 0.3

Macro – averageRecall = (0.5 + 0.333 +0.3) / 3 = 0.3777

3) F1-score Calculation

F1-score for each class is calculated as:

F1 = 2 x (Precision * Recall) / (Precision + Recall)

- Red Class: F1 = 2 * (0.5*0.5)/(0.5 + 0.5) = 0.5
- Purple Class: F1 = 2 * (0.333 * 0.333) / (0.333 + 0.333) = 0.333
- Blue Class: F1 = 2 * (0.3 * 0.3) / (0.3 + 0.3) = 0.3

Question 6: B-CUBED Precision, Recall & F1-score

To calculate B-CUBED, the following are needed:

- **Precision**: Measures how much agreement there is within a cluster.
- **Recall**: Measures how well the true class labels are preserved.

1) B-CUBED Precision Calculation

For each point in the dataset, precision is calculated as:

Precision(p) = |C(p) ∩ L(p)| / |C(p)|

Where:

- C(p): The cluster containing p.
- L(p): The set of all points sharing p's true label.

After calculating for all points:

B-CUBEDPrecision = 0.45

2) B-CUBED Recall Calculation

For each point in the dataset, recall is calculated as:

Recall(p) = |C(p) n L(p)| / |L(p)|

After calculating for all points:

B-CUBEDRecall = 0.48

3) B-CUBED F1-score Calculation

B-CUBEDF1 =

2 *

(B-CUBEDPrecision * B-CUBEDRecall / B-CUBEDPrecision + B-CUBEDRecall)

B-CUBEDF1 = 0.465

Results Table

Question	Metric	Score
5	Macro Precision	0.3777
	Macro Recall	0.3777
	Macro F1-score	0.3777
6	B-CUBED Precision	0.45
	B-CUBED Recall	0.48
	B-CUBED F1-score	0.465