# CS 620–Introduction to Data Science and Analytics, HW5
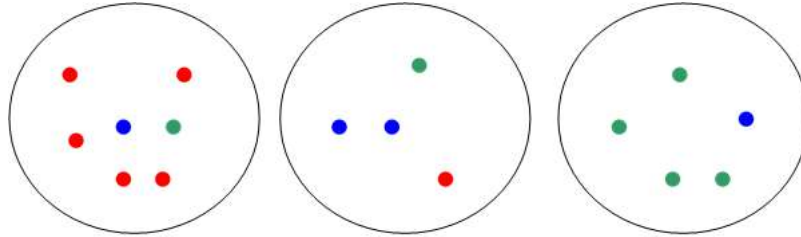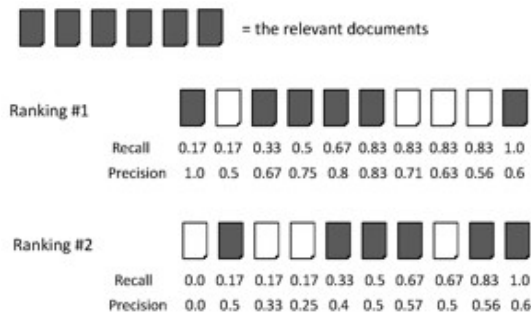
1) **(35 points)** Consider the following 3 clusters.
   a. (5 pts) Calculate the Purity
   b. Create the contingency table (confusion matrix) and using the contingency table,
      i. (20 pts) Calculate the Rand index
      ii. (10 pts) Calculate the Balanced F measure



2) **(45 points)** Consider the 2 ranking algorithms in the figure below.
   a. (10 pts) Calculate the confusion matrix values (tp, fp, tn, fn) for position 9 ($1^{st}$ position is in the left most) in each ranking method.
   b. (10 pts) Using the confusion matrix calculated in part a.), compute the Accuracy and Harmonic Mean (F measure) at position 9 for both ranking methods.
   c. (10 pts) Calculate the Mean Average Precision (MAP) for both ranking methods (for all the retrieved documents).
   d. (15 pts) Interpolation defines precision at any recall level as the maximum precision observed in any recall-precision point at a higher recall level. Calculate the Interpolated Precision for each standard recall values (0.0,0.1….1.0) and generate the Recall-Precision graph for Ranking algorithms #1 and #2.



$\blacksquare\blacksquare\blacksquare\blacksquare\blacksquare$ = the relevant documents

Ranking #1

| Recall | 0.17 | 0.17 | 0.33 | 0.5 | 0.67 | 0.83 | 0.83 | 0.83 | 0.83 | 1.0 |
|--------|------|------|------|-----|------|------|------|------|------|-----|
| Precision | 1.0 | 0.5 | 0.67 | 0.75 | 0.8 | 0.83 | 0.71 | 0.63 | 0.56 | 0.6 |

Ranking #2

| Recall | 0.0 | 0.17 | 0.17 | 0.17 | 0.33 | 0.5 | 0.67 | 0.67 | 0.83 | 1.0 |
|--------|-----|------|------|------|------|-----|------|------|------|-----|
| Precision | 0.0 | 0.5 | 0.33 | 0.25 | 0.4 | 0.5 | 0.57 | 0.5 | 0.56 | 0.6 |

3) **(20 points)** Describe your thoughts about what you think it means to work as a data scientist. You may therefore – if you like – be very personal and describe your own plans and fears for your future career, criticism (or appreciation) for your education, skills you need to develop further, and soon. This question is intended to encourage you to reflect about yourself and your future career, and will therefore be graded generously!

**What to turn in:** PDF should contain the following information at the top. **Lastname-hw5.pdf**
> CS620
> HW5
> @author:
Submit your pdf file to Blackboard. Due: Sunday, Dec. 05, 2021, 11.59pm