

## MARS Charts and MARS Metrics User Instructions:

To produce the results:

Upload a CSV (comma-separated-value) text file with the classification decisions for the classifiers you are comparing, in the following format:

1. Column 1: OBSERVATION\_ID: The observation ID's for the observation that was classified by the classifier. Duplicate values permitted in this column since multiple classifiers will make classification decisions for each observationID.
2. Column 2: CLASSIFIER\_NAME: Name of the classifier. Duplicate values permitted, as the classifier was applied (by you) to each observationID
3. Column 3: PREDICTED\_CLASS: The class predicted by the respective classifier for respective observation ID. This should be a 1 or 0 value, where 1 indicates that classifier predicted the observation was in the target class, and 0 indicates the classifier predicted the observation was not in the target class.
4. Column 4: ACTUAL\_CLASS: Actual (true) class respective to that observation and classifier. This should be a 1 or 0 value, where 1 indicates the observation was in the target class, and 0 indicates the observation was not in the target class.

## How to interpret the results:

1. MARS Shine Through Metrics table (Proportions):
  - Each cell in the Shine-Through Metrics Table consists of two values (For E.g., 0.11 | 0.44)
    - a. The first term (the term of the left of the pipe symbol) indicates the proportion of true positives found exclusively by the classifier on the y-axis, relative to total unique (non-duplicated) true positives found across all classifiers. Non-duplicated means each observation is counted once only, and each observation is counted as a true positive if any classifier was able to identify that observation as a true positive. For example, 0.11 indicates that 11% of the true positive observations identified by the classifier named on the y-axis, were exclusive to that classifier, and were not found (discovered) by any other classifier (“shined-through”).
    - b. The second term indicates the proportion of true positives found exclusively by the combination of the classifier on the y-axis and the classifier on the x-axis, together, relative to total unique (non-duplicated) true positives found across all classifiers. Non-duplicated means each observation is counted once only, and each observation is counted as a true positive if any classifier was able to identify that observation as a true positive. For example, 0.44 indicates that 44% of the true positive observations identified by the classifiers named on the y-axis and x-axis, were exclusive to those two classifiers, and were not found (discovered) by any other classifier (“shined through”).
2. MARS Shine Through Bubble Chart:
  - The x-axis and y-axis denote all the classifiers being compared.
  - The orange-colored circle corresponds to the pair of classifiers on the X-axis and Y-axis and indicates the number of True Positives found exclusively by the combination of the two

classifiers named on the x-axis and the y-axis (“exclusive” means true positives which are not identified by any other classifier).

- The yellow-colored circle represents True Positives found exclusively by the classifier named on the y-axis (“exclusive” means true positives which are not identified by any other classifier).
- The radius of each bubble represents the count of unique observations. The area of each bubble should not be interpreted as area does not scale linearly with radius: instead, interpret the radiuses alone, as the difference in radiuses clearly shows the magnitude differences of the observation counts.

3. MARS Occlusion Metrics Table(proportions):

- Each cell in the Occlusion Metrics Table consists of two values (E.g., 0.05 | 0.33)
  - a. The first term (the term on the left of the pipe symbol) indicates the proportion of false negatives exclusively labelled by the classifier on the y-axis which were correctly labelled as true positives by any of the other classifiers, relative to total unique (non-duplicated) true positives found across all classifiers. Non-duplicated means each observation is counted once only, and each observation is counted as a true positive if any classifier was able to identify that observation as a true positive. For example, 0.05 indicates that 5% of the false negative observations identified by the classifier named on the y-axis, were labelled correctly as true positives by any of the other classifiers. More simply: 0.05 indicates the y-axis classifier occluded (hid / missed) 5% of the available true positives seen by any other classifier.
  - b. The second term indicates the proportion of false negatives exclusively labelled by the combination of classifier on the y-axis and the classifier on the x-axis which are labelled correctly as true positives by any of the remaining classifiers, relative to total unique (non-duplicated) true positives found across all classifiers (if its labelled correctly by any of the other classifiers, the observation is included in the denominator count). For example, 0.33 indicates that 33% of the false negative observations identified by the classifiers named on the y-axis and x-axis, were found (discovered) by any of the other classifiers. More simply: the combination of x- and y-axis classifiers occluded (hid / missed) 33% of the available true positives seen by any other classifier.

4. MARS Occlusion Bubble Chart:

- The x-axis and y-axis denote all the classifiers being compared.
- The orange-colored circle represents false negatives of the classifier named on the y-axis.
- The red circle represents false negatives of the classifier named on the x-axis and y-axis (i.e., observations in the target class that were missed by the classifier on the x-axis and y-axis) which were identified correctly by any of the other classifiers.
- The radius of each bubble represents the count of unique observations. The area of each circle should not be interpreted as area does not scale linearly with radius: instead, interpret the radiuses alone, as the difference in radiuses clearly shows the magnitude differences of the observation counts.