An **ROC curve** (**receiver operating characteristic curve**) is a graph which describes the performance of a classification model at all classification thresholds. This curve plots has two parameters:

- **True Positive Rate (TPR) =TP/(TP+FN)**
- **False Positive Rate (FPR) =FP/(FP+TN)**

An ROC curve plots TPR vs. FPR at different classification thresholds. Lowering the classification threshold classifies more items as positive, thus increasing both False Positives and True Positives. Computed ROC and AUC for given data of Autor1 and Author 2 are given below.

Author 1

ROC curve together with the AUC
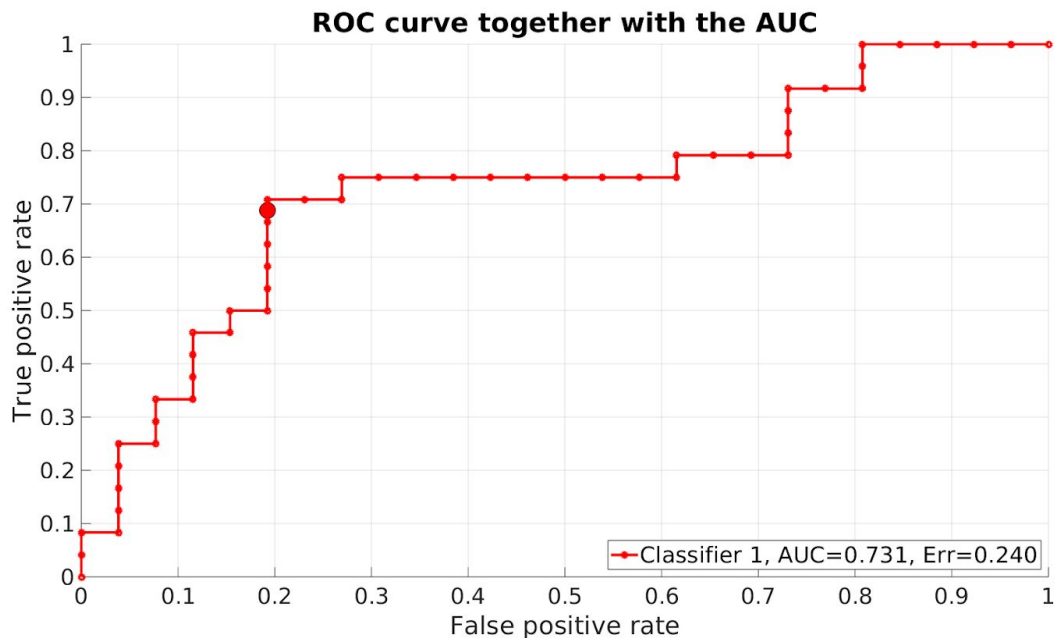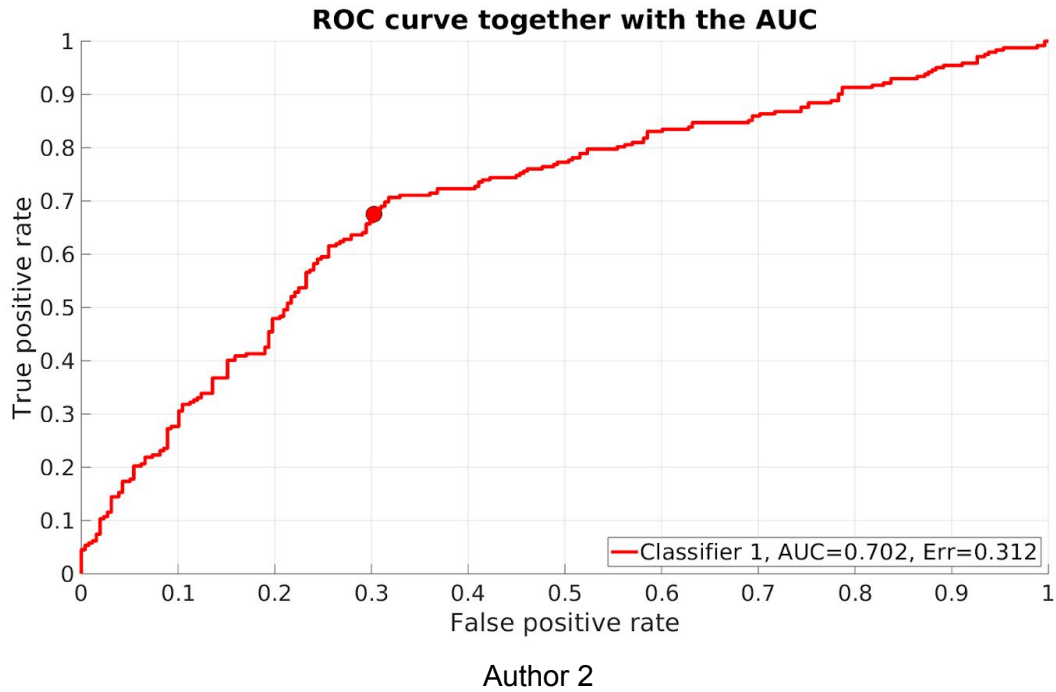
Author 2

As shown in figures, the author-1 obtained an AUC of 0.73 and author-2 obtained an AUC of 0.702. Clearly, as reported the performance of author-1 classifier is better than the one obtained by author-2's classifier, but the size of evaluated points/data differ in length.
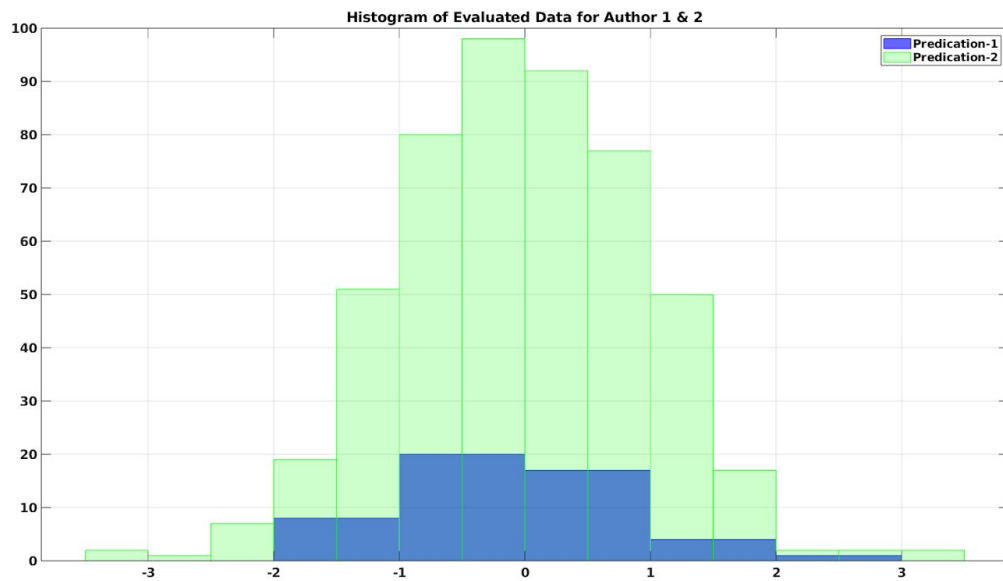

Histogram analysis of predictions of author 1 and 2

Histogram analysis of predictions of author 1 and 2 provides the following results

- The size of the dataset used by author 1 is 10 times smaller than the size of dataset of author 2.
- Both reported results follows Gaussian distribution
- Provided information is insufficient to tell the performance of classifier 1

We can increase the size of evaluation dataset and check ROC & AUC on random values from a Gaussian distribution by assigning the labels based on the sign of drawn sample using f=randn(50,1) for dataset-1 and f=randn(500,1) for dataset-2. AUC values obtained in the 1000 runs are shown below.
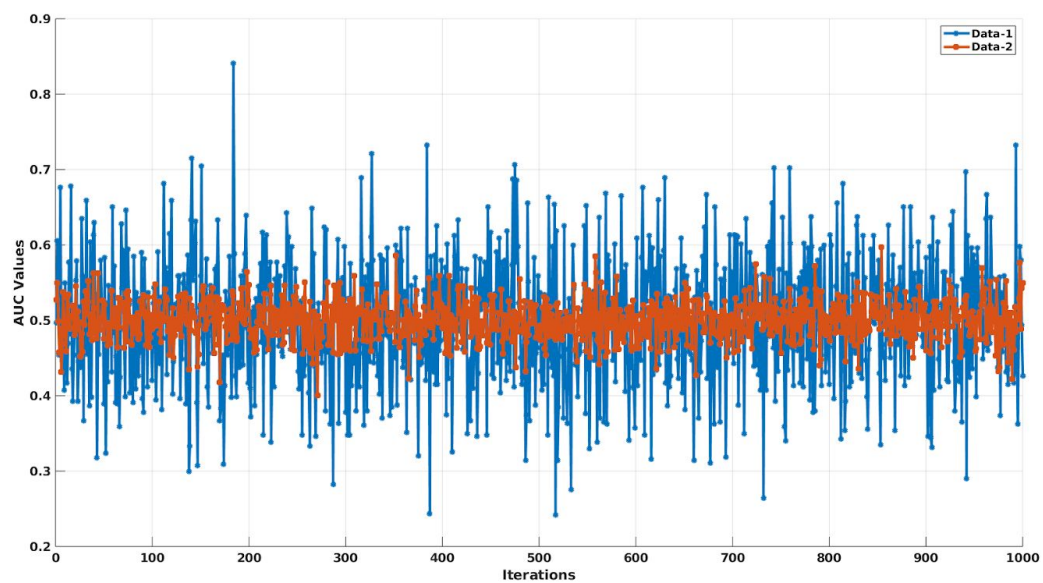
AUC values for both datasets for 1000 iterations using random classifier (Gaussian Distribution).
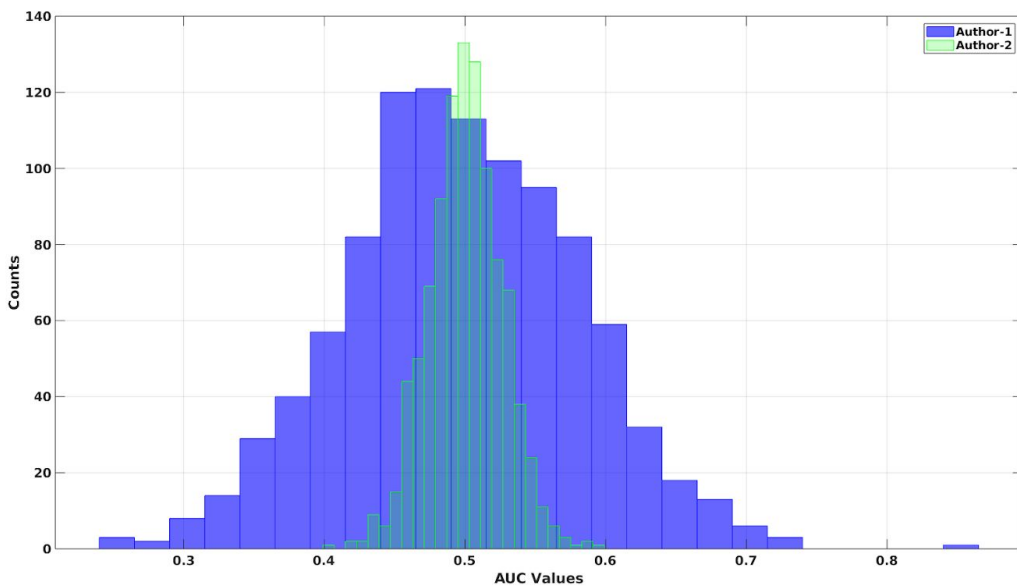
Histogram analysis for both dataset on random classifier.

Key Findings:
Variance of the AUC values obtained by a random classifier shows significant change, the random classifier has a much higher variance around 0.5 on dataset-1 than in the dataset-2 which is 10 times larger dataset. In terms of AUC values, we can see the maximum value at 0.84 for dataset-1, and 0.6 for dataset-2.

With the help of this experiment we can witness that, significantly on dataset-2 the AUC value reported by the author is much better than the one obtained by a random classifier, but this is not the case for 1st author. _Although AUC of author 1 is better than author 2, this does not prove vital information about the actual performance of classifier. Author 1 should provide further results to prove the quality of classifier._