Area under ROC curve (AUC score)

We compute the area under ROC curve from the prediction/anomaly/confidence scores. This can be calculated in a straight forward way for algorithms operating on single data point. For algorithms operating on a window, we assign the anomaly score of window to the center-most point in the window and thus we get a score for each point in the time series.

Table 1: Model performances in terms of AUC Score

Models	**Dataset 1**
Local Outlier Factor	0.816
Isolation Forest	0.796
Elliptic Envelope	0.736
One Class SVM	0.714
Streaming Least Squares	**0.918**

Mean Average Precision

Table 2: Model performances in terms of Mean Average Precision

Models	**Dataset 1**
Local Outlier Factor	0.241
Isolation Forest	0.111
Elliptic Envelope	0.123
One Class SVM	0.160
Streaming Least Squares	**0.403**

- AP (average precision) score is defined as the mean precision at the set of 11 equally spaced recall values, $R_i = [0, 0.1, 0.2, \cdots, 1.0]$

$$\text{AP} = \frac{1}{11} \sum_{\text{R}_i} \text{Precision}(\text{R}_i)$$

  where $\text{Precision}(\text{R}_i) = \mathbf{max}_{\text{R}_{i'}:R_{i'} \geq R_i} \text{Precision}(\text{R}_{i'})$
  Recall can be varied by varying the threshold for anomaly score.

- Anomaly detections are determined to be true or false depending upon the Intersection over Union (IoU) threshold.

$$\text{IoU} = \frac{\text{Intersection between ground truth and predicted anomaly}}{\text{Union of the ground truth and predicted anomaly}}$$

- Mean Average Precision score is calculated by taking the mean AP over all IoU thresholds. Averaging over multiple IoU thresholds rather than only considering one generous threshold of IoU tends to reward models that are better at precise localization.

In case of algorithms that operate on single data points in the time series and generate an anomaly score for each point, we create a window around each point and assign the anomaly score of the point to that window. Window is created in such a way that point lies at the center of the window. Now, we start selecting windows with the maximum anomaly score and remove all other overlapping windows with lower anomaly score. At the end, we are left with a set of non-overlapping windows and their scores. These windows are then used for the calculation of mean average precision as described above.