**Evaluation**

Accuracy is measured using the Area-Under the Precision-Recall Curve (AUPRC) to assert whether or not the accuracy has improved.

**Precision:** is a ratio of the number of true positives divided by the sum of the true positives and false positives. It describes how good a model is at predicting the positive class. Precision is referred to as the positive predictive value. Precision = TP / (TP + FP).

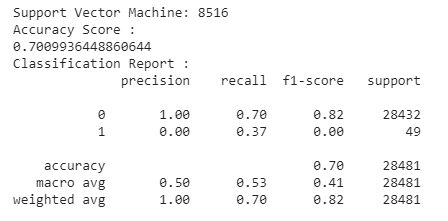
**Recall:** is calculated as the number of true positives divided by the sum of the number of true positives and the number of false negatives. It describes how good the model is at predicting the positive class when the actual outcome is positive. The true positive rate is also referred to as sensitivity. Recall = TP / (TP + FN).

**F-Measure or F1 score:** is a composite score that attempt to summarize the precision and recall. F1 score calculates the harmonic mean of the precision and recall (harmonic mean because the precision and recall are ratios).

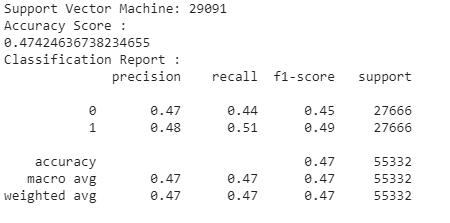
**Observations:**

In my proposal, I stated that I will use logistic regression, decision tree, and support vector machines (SVM).

Before (Without Data-Point Approach):



After (With Data-Point Approach):



For this iteration I will discuss about the SVM findings.

**Precision** improved from **0% to 48%** - This data shows that if the data is highly imbalance, the model struggles to detect fraudulent transactions. After using under sampling which is a technique of the data-point approach. There was a significant improvement in the ability to predict positive classes.

**Recall** improved from **37% to 51%** - Recall describes how good the model is at predicting the positive class when the actual outcome is positive, therefore this show the strength of True Positive which are actual fraudulent cases. We saw a 14% improvement.

NOTE: The research is on-going and there is still plenty of research required to identify the most suitable technique to handle highly imbalance data.