

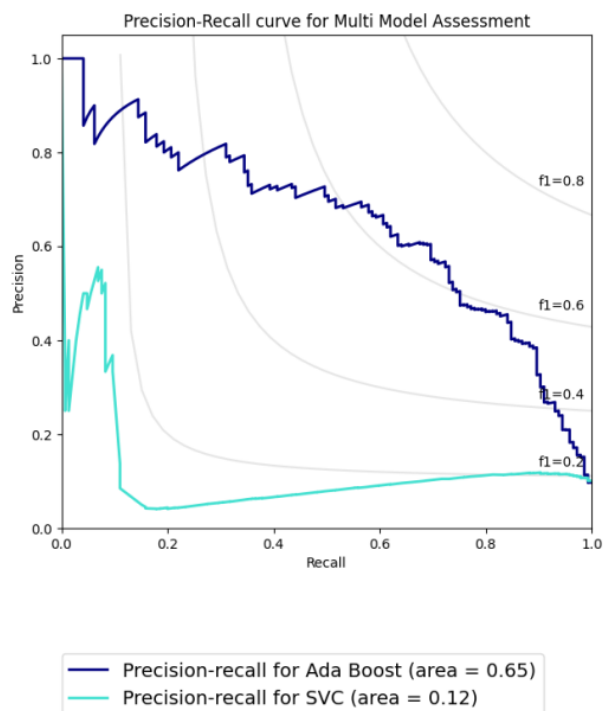
# ASSIGNMENT 5 REPORT

Elliott Barnes | Marty Whelan

- 1. short and clear description and justification of any data pre-processing done**
  1. We cleaned the data by detecting and replacing null values with zero
- 2. short and clear description and justification of the ML methods evaluated and the grid used for grid search CV**
  1. We chose two reasonably different style classifiers and tested them against each other to see the results, and created a grid that would increased the speed of the evaluation process [\[classifier comparison\]](#)
- 3. short and clear justification of the performance metric used to select the best model**
  1. We used f1 scoring because it helps with class imbalance by balancing the metric across positive/negative samples
- 4. the best hyper-parameter setting found per method**
  1. Best params: {'algorithm': 'SAMME.R', 'learning\_rate': 0.1, 'n\_estimators': 200}
  2. Best params: {'C': 0.1, 'gamma': 1, 'kernel': 'sigmoid'}
- 5. a table with the CV results for the best model per ML method with mean and standard deviation**

Method	Mean	Std	CV
AdaBoost	0.92	0.02	[0.94633508 0.92922674 0.90432503 0.94495413 0.89121887]
SVC	0.90	0.01	[0.90052356 0.90039318 0.90694626 0.90694626 0.87942333]

6. a figure showing a graphical representation of the cross-validation performance of the best models (one model per ML method)



7. an acknowledgement section listing your collaborations and online sources

1. <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.AdaBoostClassifier.html>
2. <https://scikit-learn.org/stable/modules/svm.html#svm-classification>
3. [https://scikit-learn.org/stable/auto\\_examples/classification/plot\\_classifier\\_comparison.html](https://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html)

8. a program specification section listing the Python version and libraries you used.

1. Python 3.8.1

2. Sys, NumPy, Sklearn, Pandas, Matplotlib