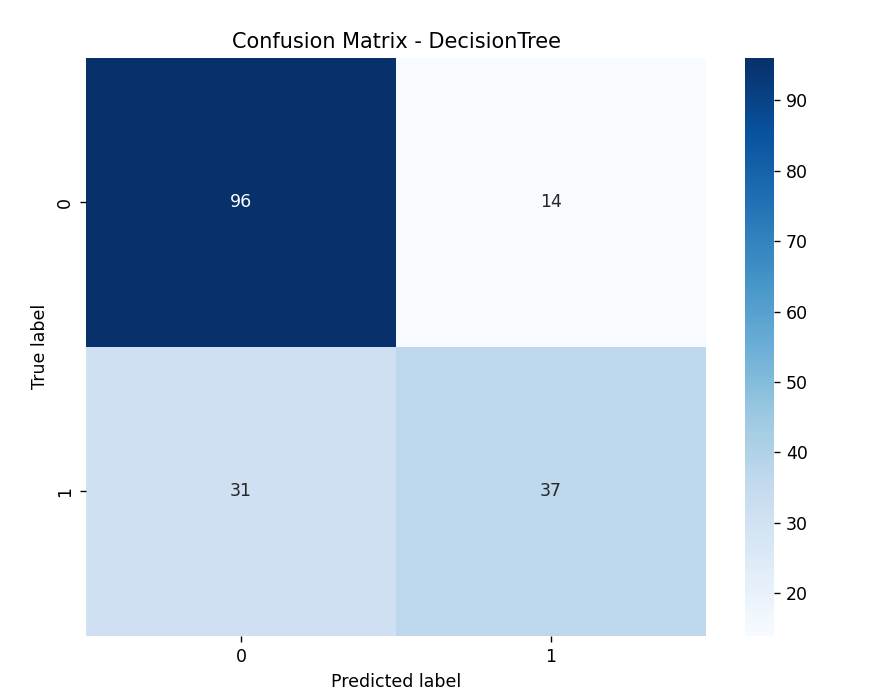
**Assignment – 3**

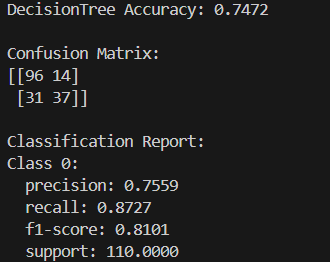
**Trees and Boosting**

Kaushik Budur

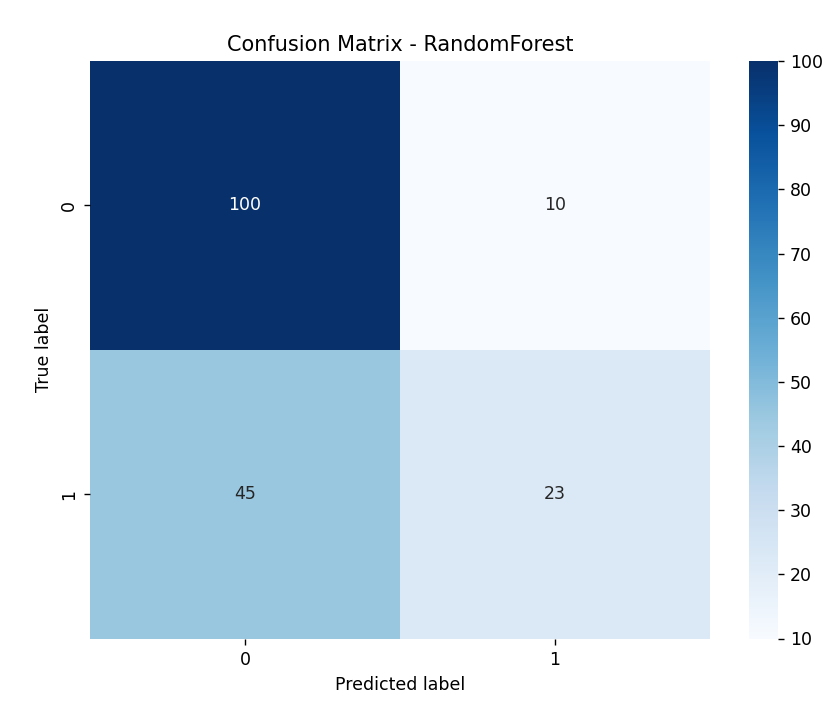
Mymav ID: 1002224112

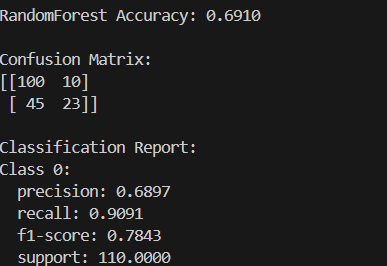
In our implementation, we use a Decision Tree as the base model, with the predict and fit functions, and parameters that include either misclassification rate, Gini impurity, or entropy. The parameters also include max\_depth (which limits how deep the tree can grow), min\_samples\_split (the minimum number of samples needed to split a node), and min\_samples\_leaf (the minimum number of samples required for a leaf node). With a maximum depth of 10 and selecting Gini impurity as the splitting criterion, our model achieves an accuracy of 74.72%.



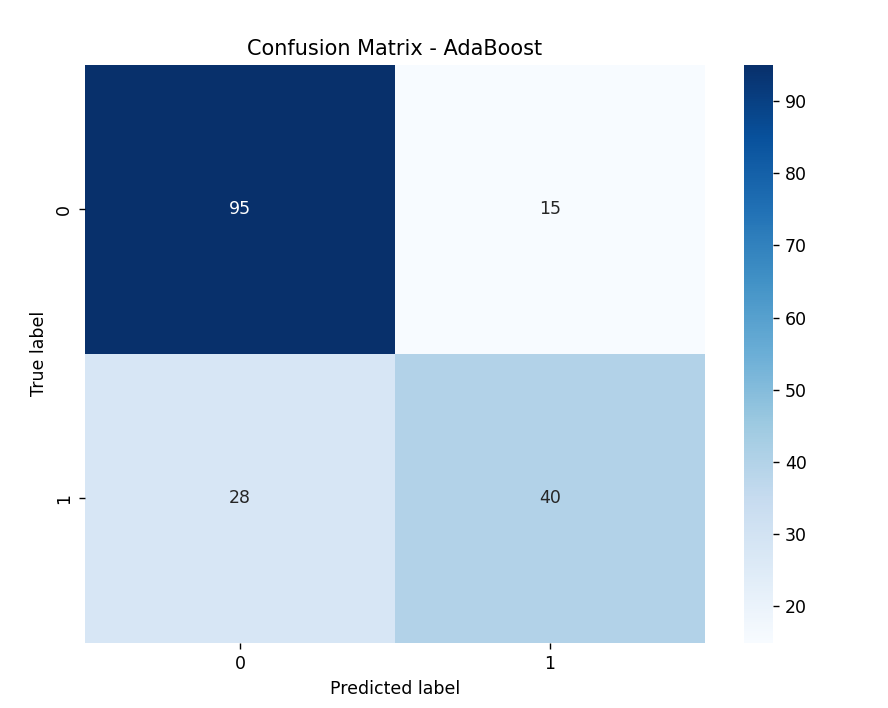


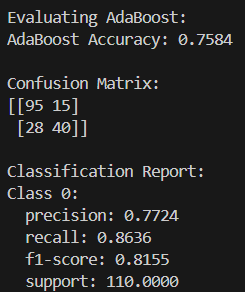
Next, we use Random Forest as our first ensemble method, with the Decision Tree as the base classifier. The model is configured with parameters such as num\_trees, which specifies the number of trees in the forest, and min\_features, which determines the minimum number of features to consider when selecting a subset of features for each tree. We also utilize the predict and fit functions in our Random Forest classifier. In the predict function, the model returns the prediction with the most votes. During training, sampling is done with replacement, and features are selected based on min\_features and num\_features, where num\_features is a hyperparameter that is typically smaller than the total number of features. By setting the minimum number of features to 2, we achieve an accuracy of 69.10%.





Finally, we use AdaBoost as our second ensemble method, with the Decision Tree serving as our weak learner. The model is configured with parameters such as num\_learners, which specifies the maximum number of learners to use when fitting the ensemble, and learning\_rate, which controls the weight applied to each weak learner during each iteration. The AdaBoost model includes both predict and fit methods. With a learning rate of 0.01, our model achieves an accuracy of 75.84%.





In our application of ensemble learning techniques to the Titanic dataset, we utilized the train.csv and test.csv files from the Kaggle competition. We performed data preprocessing steps, including handling missing values, encoding categorical variables, and remove outliers for model training, we employed AdaBoost with Decision Trees as weak learners and Random Forest classifiers. These ensemble methods are effective for classification tasks.

Incorporating AdaBoost with Decision Trees as weak learners for our Titanic Dataset has been shown to enhance model performance, resulting in improved recall and F1-scores. Conversely, the Random Forest model, despite being an ensemble method, may underperform compared to AdaBoost and Decision Trees in certain metrics.