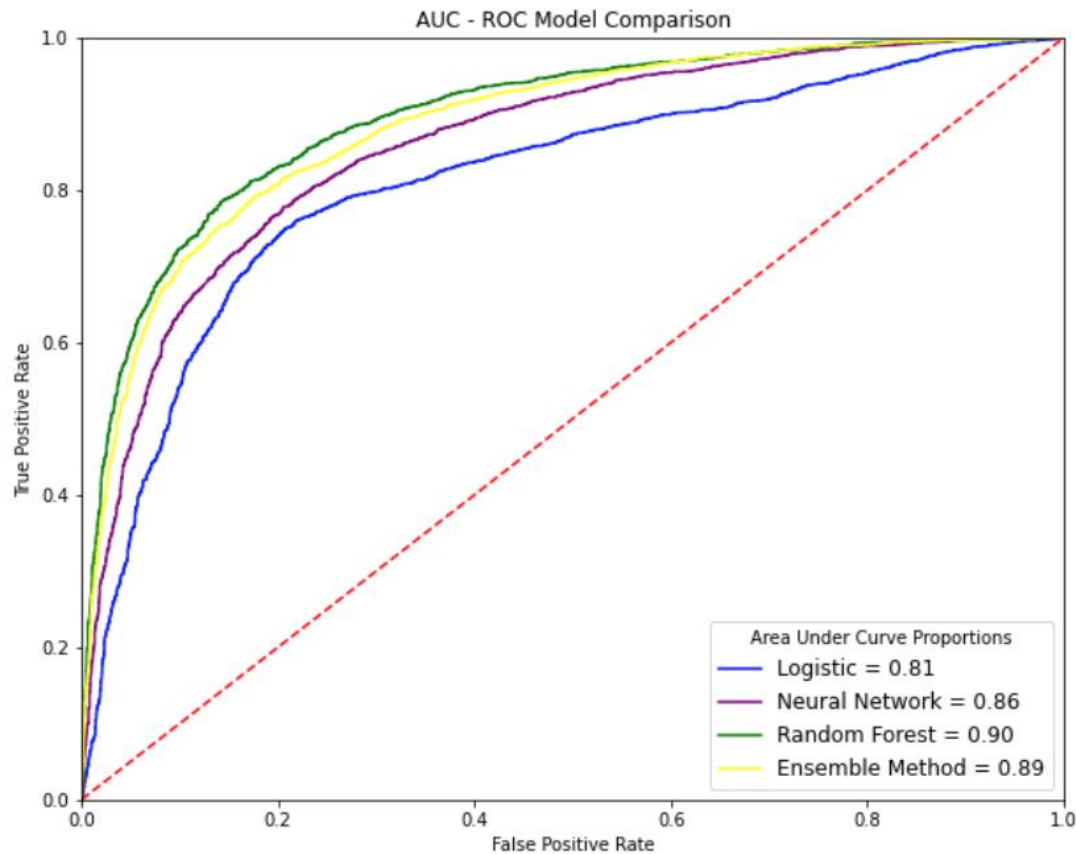


Model Comparison Analysis: Affiliate Action Probability



[Graph Interpretation] An Optimal Score occurs when X-value reaches 0.0 and Y-value reaches 1.0, therefore correctly categorizing the highest proportion of entries.

Executive Summary:

Across multiple performance metric comparisons, the Random Forest Model outperforms the Neural Network, Logistic Regression, and Ensemble Method Models. With an average accuracy of 83% and an average recall of 80%, the highest-value proportion on AUC-ROC, and a moderate runtime, it is clear the Random Forest Model is the leading option.

Detailed Analysis:

In reference to the above AUC-ROC graph, the top performing model holds the most overall area, with 1.0 area of the graph representing 100% of the total data entries. Therefore, with a Random Forest score of 0.90, we know this model has the higher overall performance in relation to both True-Positive rate and False-Positive rates.

5-fold cross validation:

Logistic Regression

Accuracy : 76.01%
Precision : 74.41%
Recall : 79.29%
Total Runtime:0.3907 seconds

Neural Network

Accuracy : 77.55%
Precision : 76.90%
Recall : 78.96%
Total Runtime:30.1546 seconds

Random Forest

Accuracy : 83.05%
Precision : 82.68%
Recall : 80.29%
Total Runtime:279.5386 seconds

Voting Classifier

Accuracy : 78.73%
Precision : 78.65%
Recall : 79.12%
Total Runtime:307.8373 seconds

Applying 5-Fold Cross Validation as a secondary evaluator confirms similar findings. While Random Forest and Ensemble (Voting Classifier) accuracy and precision have the least amount of difference, Random Forest runtime is notably faster than Ensemble Methods. This is understandable, as the Ensemble Model is built from Logistic, Neural Network, and Random Forest and therefore must train each model independently.

While it is true the Ensemble Method performs similarly to Random Forest, it is important to keep in mind the Ensemble Method runtime is +10% slower. This would surely compound when training a larger dataset. Considering how Random Forest has the highest AUC score, while also performing with a moderate runtime, it is the most logical choice for a final model.

It is also worth noting that during testing the Random Forest Model accuracy values did at times reach 84% without Cross-Validation. Additionally, I found Ensemble Methods with hard voting created a more accurate classifier (83% accuracy), however this parameter does not allow for a probability prediction output. The best model evaluator is classification accuracy, however the most useful data-tool for the Commerce Team is one with probability prediction capabilities.