

Group Name: GIG group

Name: Rupert Tawiah-Quashie

Email: rupertquash@gmail.com

Country: USA

College: Hampshire College

Specialization: Data Science

To determine the best model for ABC Bank's objective of improving marketing efficiency by targeting customers more likely to subscribe to term deposits, let's analyze the results of the Logistic Regression, Random Forest, and Gradient Boosting models.

Performance Analysis

1. Logistic Regression Model

- Precision (Yes): 0.58 - This 58% precision means that of those predicted to subscribe to term deposits, 58% of them actually did. So over half of positive predictions were correct.
- Recall (Yes): 0.21 - A low recall of 21% means that only 21% of actual subscribers were correctly identified. In other words, 79% of true subscribers were missed.
- F1-Score (Yes): 0.30 - The low 0.30 F1 score for the 'yes' class of subscribers demonstrates the model's poor balance between precision and recall for this group.
- Accuracy: 89% - The relatively high 89% overall accuracy is influenced largely by the 'no' class, since those who don't subscribe are more prevalent in the data.

2. Random Forest Model

- Precision (Yes): 0.63 - At 63% precision for the 'yes' class, this is better than the Logistic Regression Model. So 63% of those predicted to be subscribers did indeed subscribe.
- Recall (Yes): 0.39 - The recall for subscribers here is almost double that of the Logistic Regression, at 39%. So 39% of actual subscribers were correctly identified.
- F1-Score (Yes): 0.48 - The higher 0.48 F1 score for subscribers shows this model's much better balance of precision and recall for the 'yes' class.
- Accuracy: 90% - The overall 90% accuracy is slightly better than the Logistic Regression. However, it remains skewed by the majority 'no' class.

3. Gradient Boosting Model

- Precision (Yes): 0.62 - With 62% precision for subscribers, this is similar to the Random Forest Model's performance and better than Logistic Regression. 62% of subscriber predictions were accurate.
- Recall (Yes): 0.38 - At 38% recall for predicting subscribers, Gradient Boosting is very close to Random Forest's 39% recall. So 38% of actual subscribers were correctly recognized.
- F1-Score (Yes): 0.47 - The 'yes' class F1 score of 0.47 is comparable to Random Forest, indicating that Gradient Boosting also balances precision and recall reasonably well.
- Accuracy: 90% - The overall model accuracy of 90% is consistent with that of the Random Forest Model and better than Logistic Regression.

Model Suitability Evaluation

- Logistic Regression: While it has reasonably high overall accuracy, the critically low recall of 21% for the 'yes' subscriber class means this model misses much of the target group that the bank aims to identify. Therefore, it is less suitable for this objective.
- Random Forest and Gradient Boosting: These two models have very similar performance across the board. Both substantially outperform Logistic Regression in terms of recall for identifying subscribers, while maintaining precision. They are much better suited for recognizing actual subscribers without too great an increase in false positives.

Key Decision Factors

1. Importance of Recall: With the goal being to target potential subscribers for term deposits, high recall for the 'yes' class is particularly important. Both Random Forest and Gradient Boosting do significantly better than Logistic Regression here.
2. Computational Resource Constraints: As ensemble methods, Random Forest and Gradient Boosting have higher computational demands than Logistic Regression. Of the two, Random Forest is usually less intensive, so may be preferred if resources are limited.
3. Need for Model Interpretability: Logistic Regression offers the most model transparency and interpretability. This can provide valuable insights into the key drivers and predictors of customer propensity to subscribe. However, its low performance diminishes its suitability for the bank's objectives.
4. Model Tuning Requirements: There is ample opportunity to further tune and optimize the Random Forest and Gradient Boosting models by adjusting hyperparameters. This has the potential to improve their performance even more for this application. Tuning takes more resources, however.

Recommendation

Considering the objective of accurately targeting potential term deposit subscribers, while balancing precision, Random Forest stands out as the most suitable model. It significantly outperforms Logistic Regression in identifying likely subscribers, while retaining strong precision and not requiring as many computational resources as Gradient Boosting. It also enables further tuning to potentially improve performance. However, Gradient Boosting remains a strong candidate if more resources can be utilized for optimization.