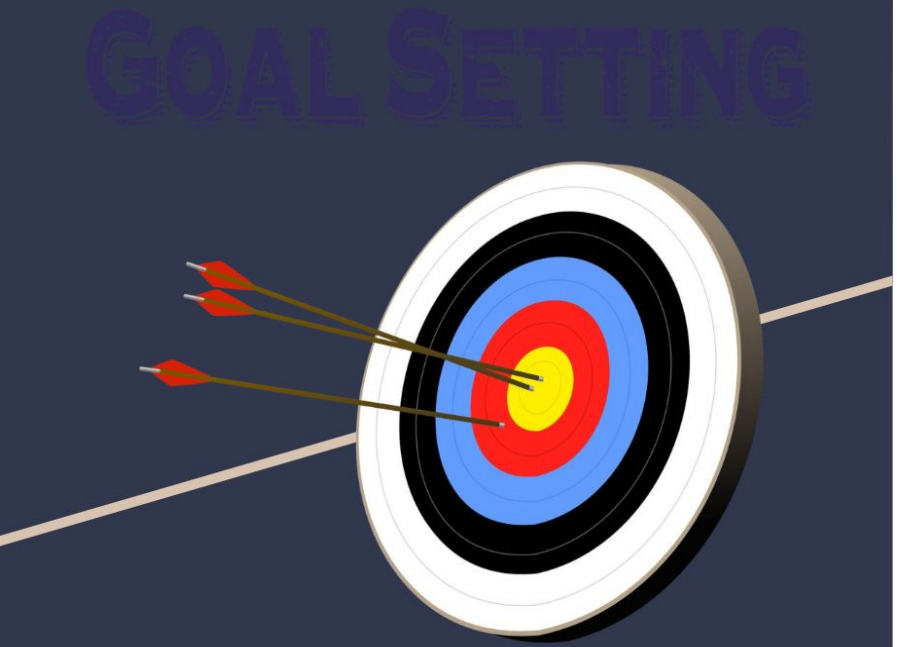
Bank Marketing Classification

by Jonathan Bardey
Flatiron School Data Science

Viewing the Data

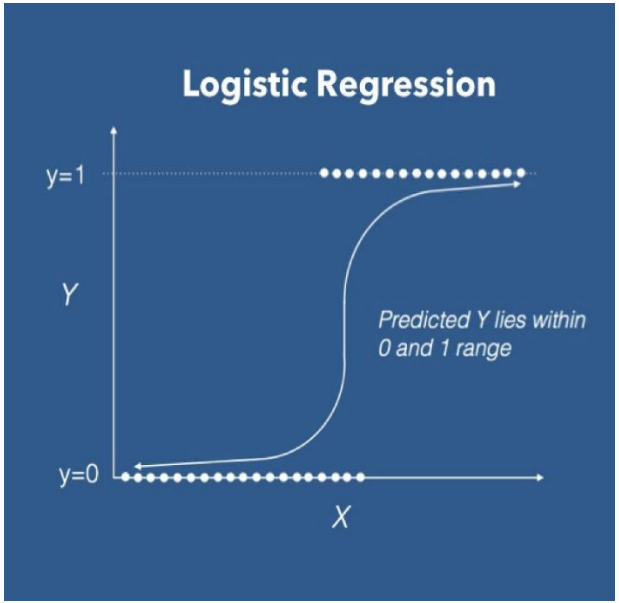
- 41888 phone calls to potential customers of bank term deposit
- 20 features: Client data (age, marital status, education) and Social/Economic attributes. Mix of categorical and numerical data.
- Target = yes deposit (1) or no deposit(0)
- Unbalanced data: 36548 no, 4640 yes

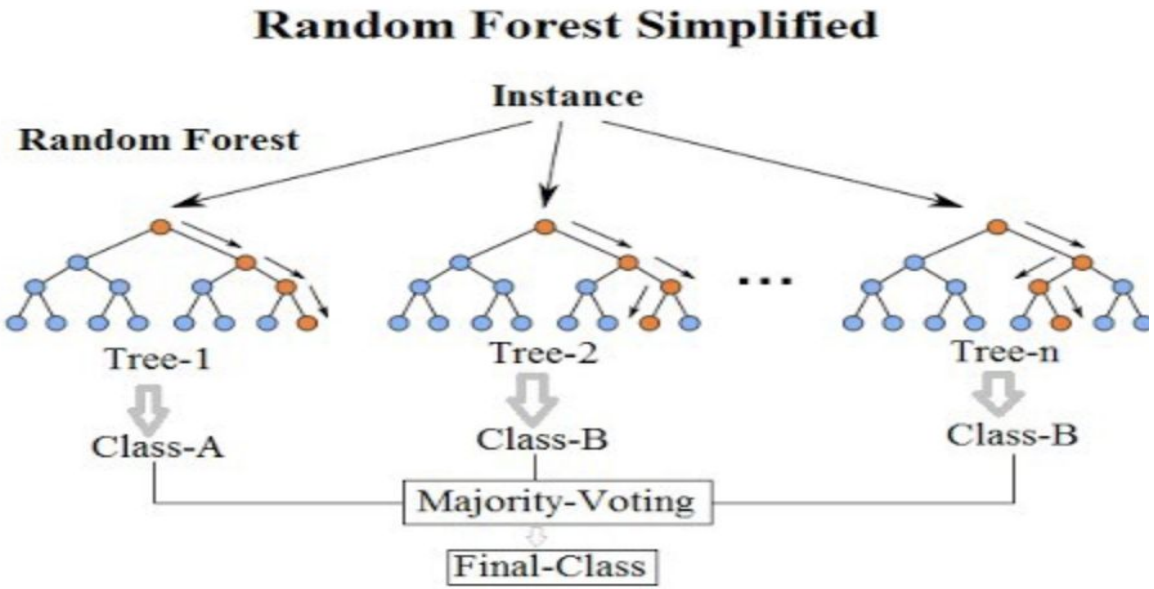
The Goal?



- Create a model to accurately predict success of the bank marketing campaign - will customers subscribe to the term deposit?

Binary Classification Models



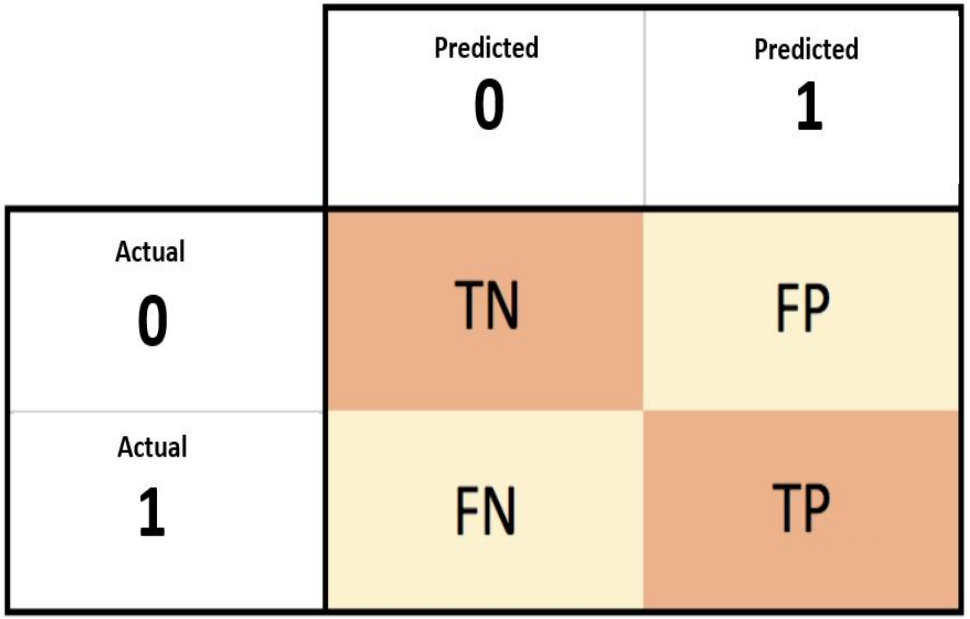


- Ran Logistic Regression, Decision Tree and Random Forest models on original and upsampled data.

Model Performance Measures

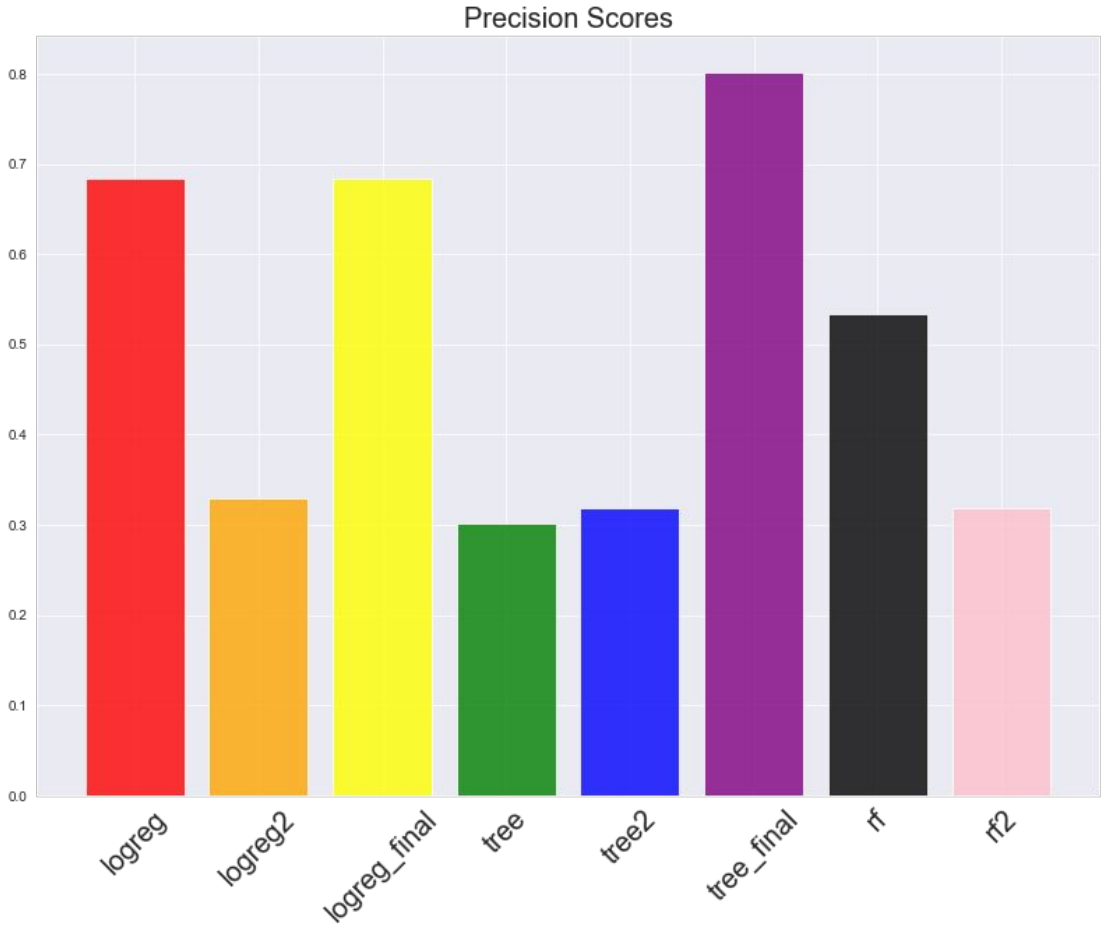
- Accuracy: (TP+TN)/(TP+TN+FP+FN)
- Precision: TP / (TP + FP)
- Recall: TP / (TP + FN)

We want to prioritize Precision to limit False Positives (predicting a customer subscribed when they did not).



Maximized Precision Model

Non-upsampled Decision Tree with tuned hyperparameters (tree_final) achieved the highest Precision score (80%) and Accuracy around 90% in line with other models.



Most Important Features:

- nr.employed
- euribor3m
- emp.var.rate
- age

“It’s the Economy, Stupid.”

The weaker the economic indicators, the more likely customers are to subscribe.

Recommendations and Future Work

- The bank should aggressively push the term deposit when economy turns downward, as a safe play for their customers and pursue other strategies in times of economic strength.

- Experiment with other classification algorithms and ensemble methods to improve predictive ability and performance metrics.

- Model based only on the client data, without the socioeconomic attributes. May glean insights to apply targeted marketing regardless of the economic environment.

Any Questions?

Thank you!