

Predicting Credit Card Default Using Payment History

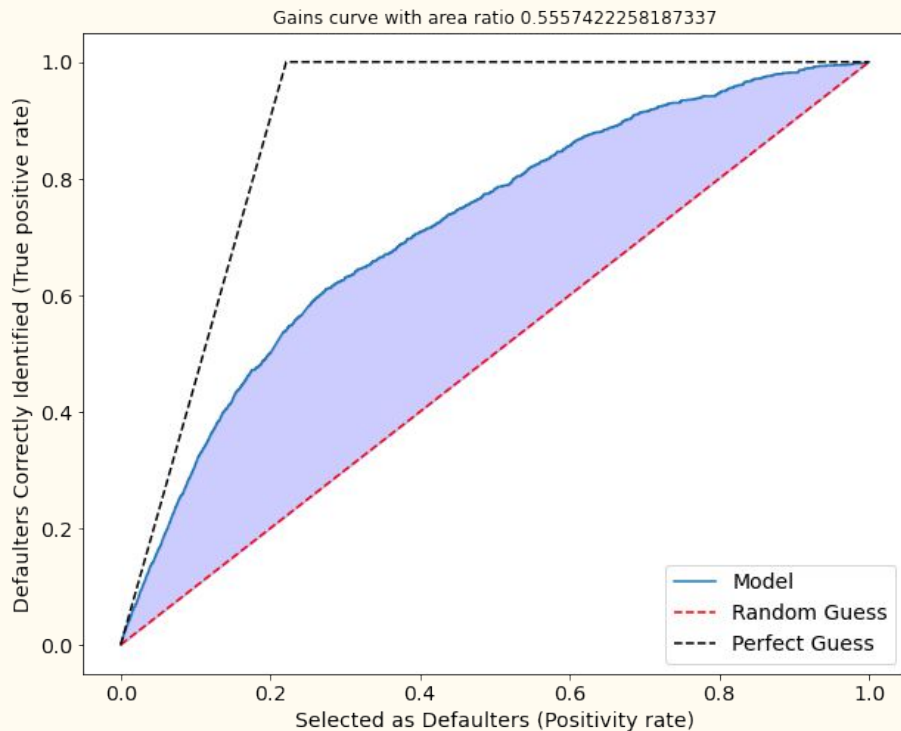
Team Rosetta: Hannah Alpert, Amin Idelhaj, and Soumya Sankar

Key Performance Indicators

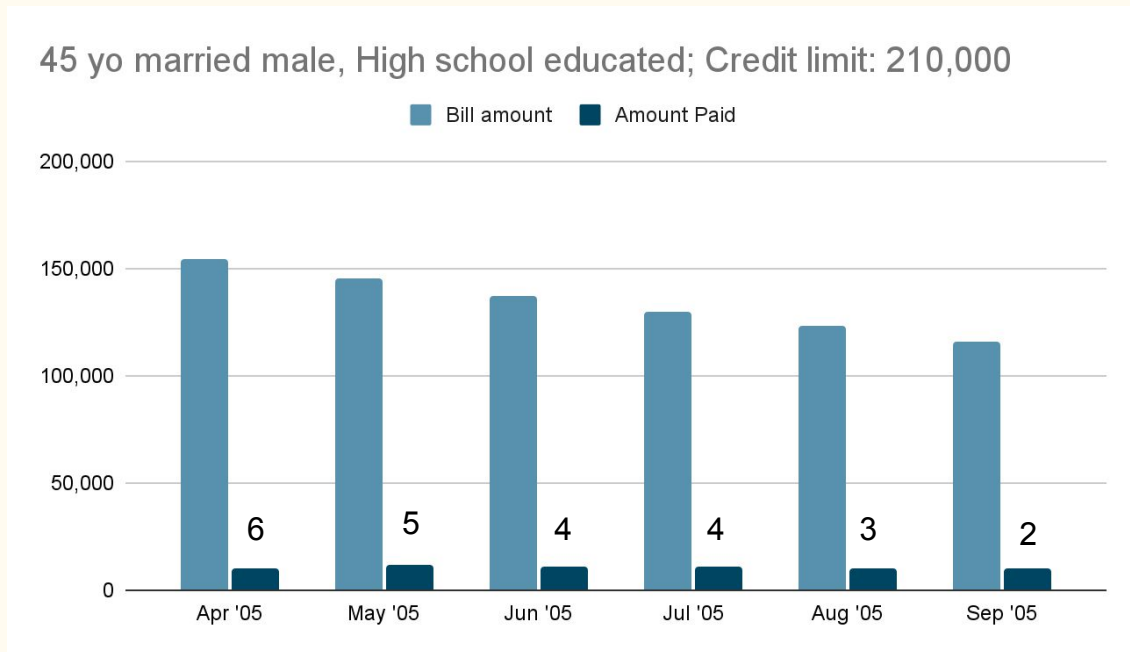
- Area ratio of a gains curve:

$$\frac{\text{Area between model and random guess}}{\text{Area between perfect and random guesses}}$$

- F1 score: .545
- Accuracy score: .818
- Weighted accuracy: .715



Will this client default?

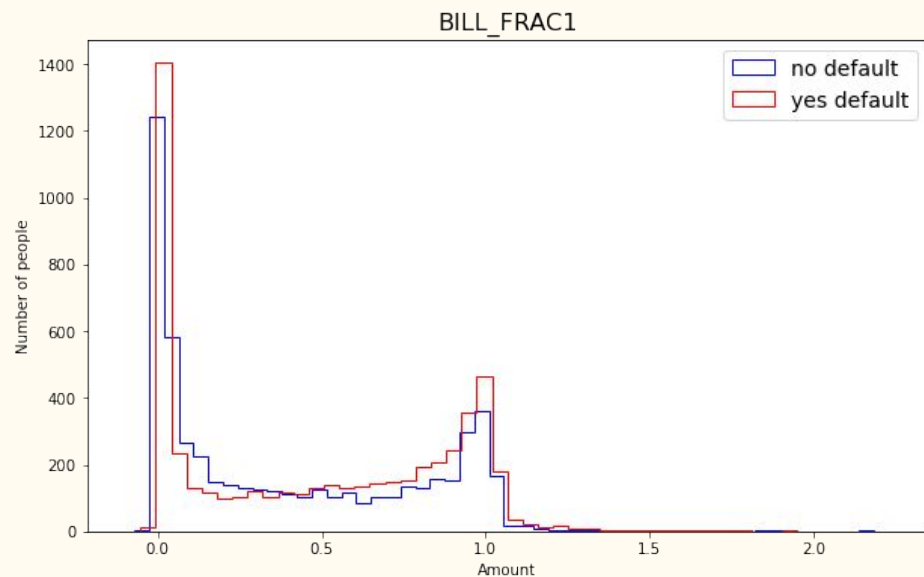
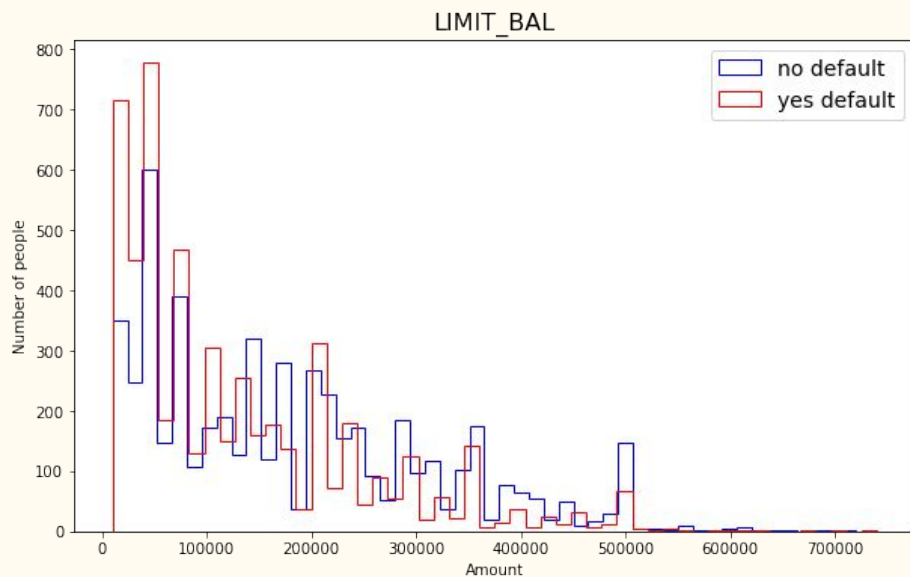


Defaulted!

Goal:
To find the best
model to predict
whether clients will
default

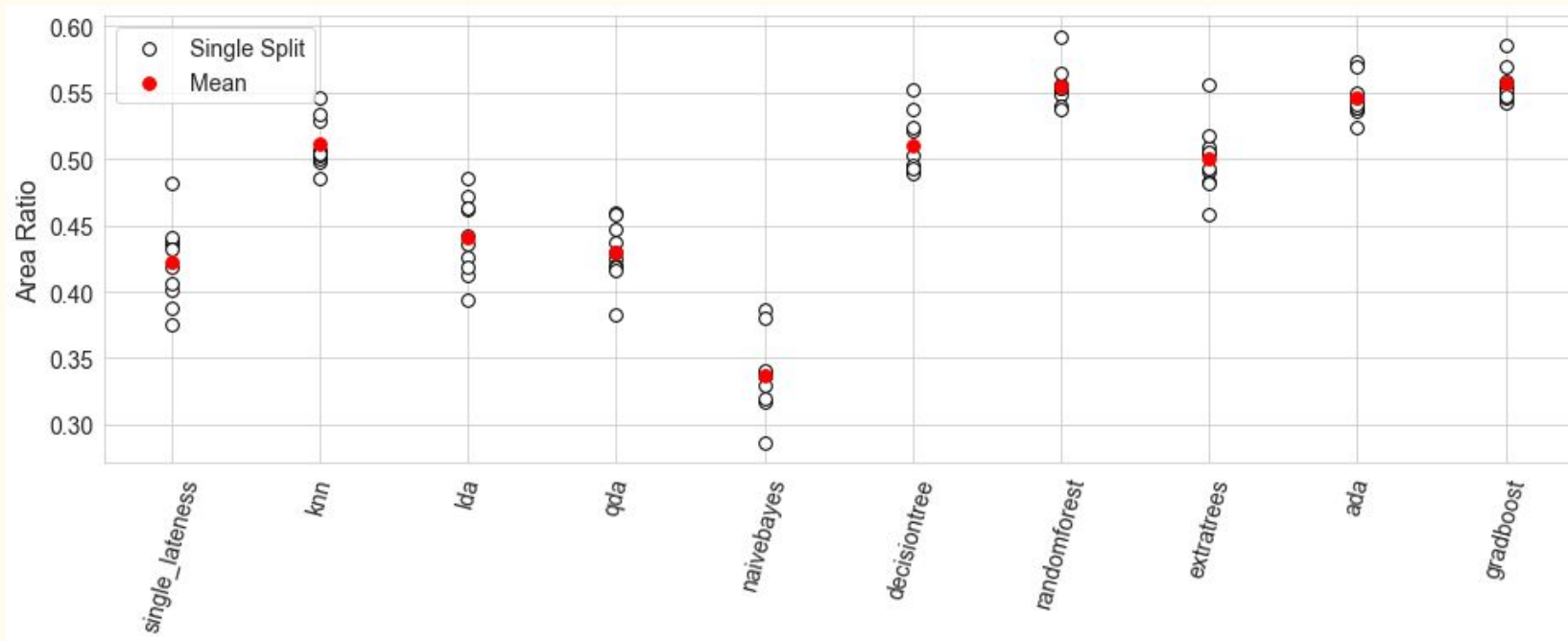
Data source: <https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients>

Feature engineering



Upshot: We can refine our variables to get more out of the data

Model comparison: Area ratio



Conclusions: which models are better?

- Linear Discriminant Analysis
- Quadratic Discriminant Analysis
- Naïve Bayesian Model
- K Nearest Neighbors
- Decision Tree Algorithm
- Extra Trees Algorithm
- Random Forest
- Adaptive Boosting
- Gradient Boosting

BAD

INTERMEDIATE

GOOD

Additional considerations for stakeholders

- Scores can be used to predict probability cutoffs
- Modification of variables
- Better data collection needed (e.g. income, dependents, other credit lines, etc.)