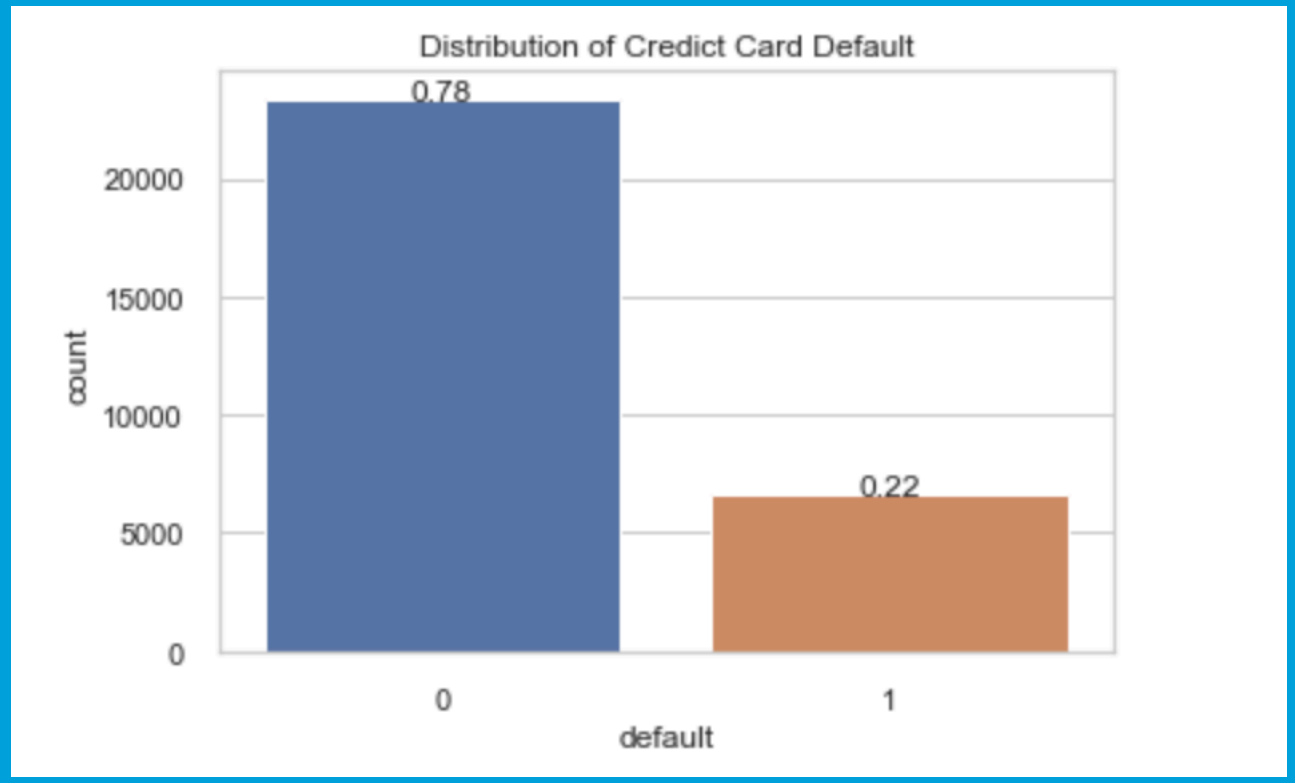
Identifying Credit Card Default

Luyuan Kong
Metis SF

Potential Clients

DATA



- Imbalanced Target
- Use SMOTE

Methodology

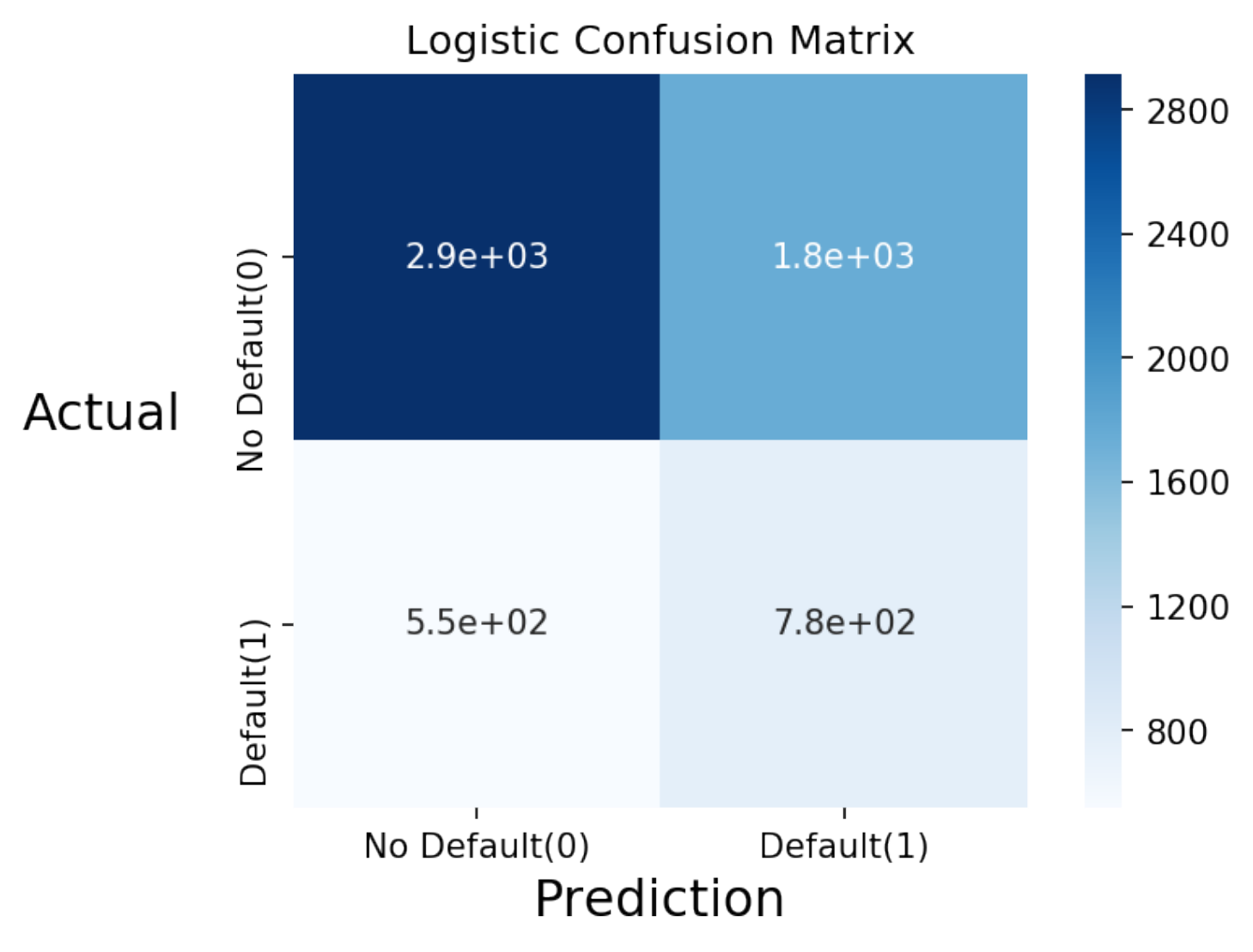
Data Loading &Data Cleaning & EDA → Feature Engineering → Hyperparameters Tuning → Model Selection → Predict on Test

Results

Compare the model with Fbeta(beta =2)

Model	Fbeta Score (in sample w/o scaling)	Fbeta Score (out sample w/o scaling)	Fbeta Score (in sample w/ scaling)	Fbeta Score (out sample w/ scaling)
XGBoost	0.35	0.29	0.48	0.48
Gaussian Naïve Bayes	0.58	0.59	0.58	0.59
Logistic Regression	0.39	0.37	0.59	0.60
Random Forest	0.28	0.29	0.53	0.52
LinearSVC	0.55	0.56	0.55	0.50
KNN	0.45	0.43	0.35	0.35

Results



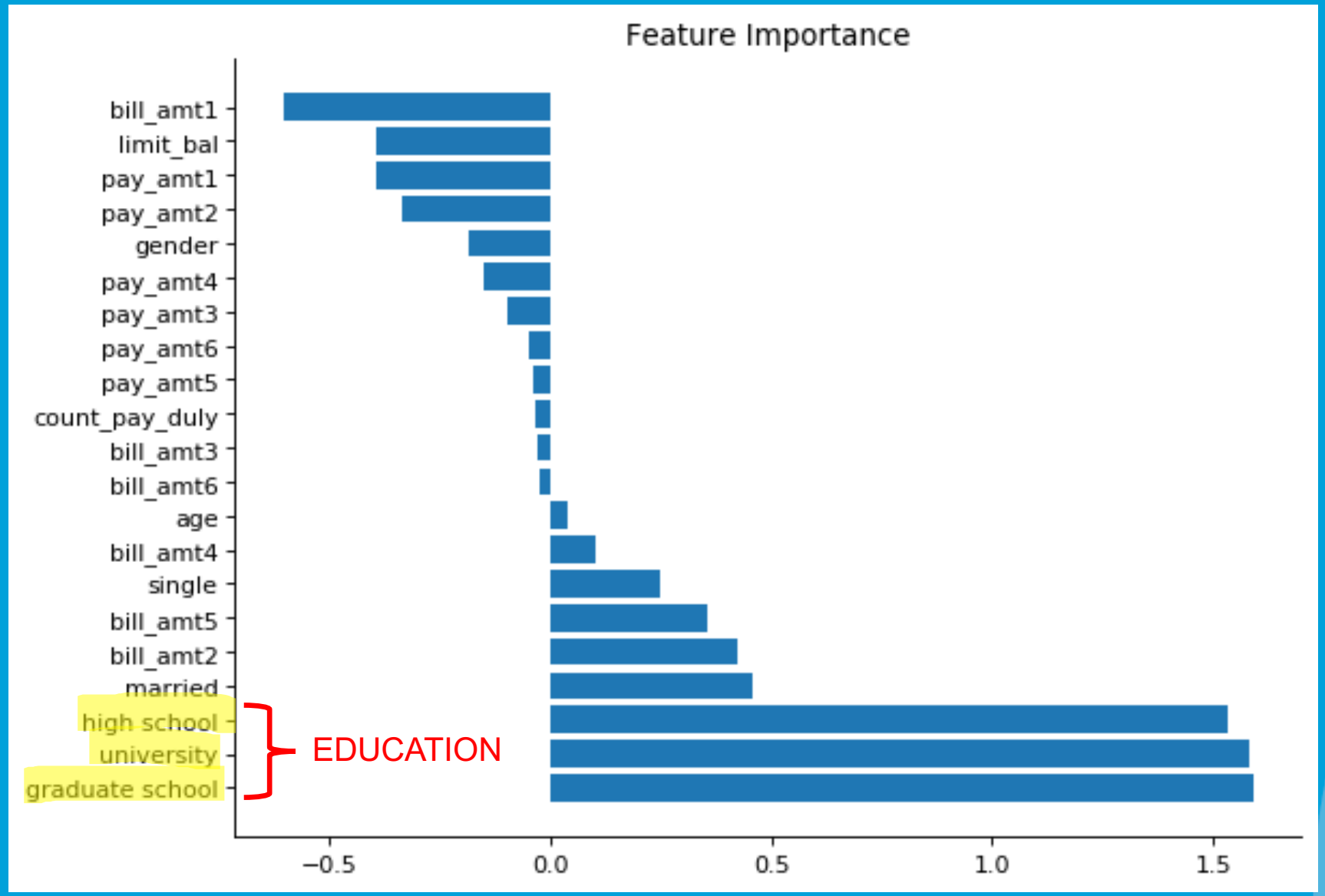
Threshold: 0.73

Recall: 0.58

Precision: 0.30

Fbeta: 0.50

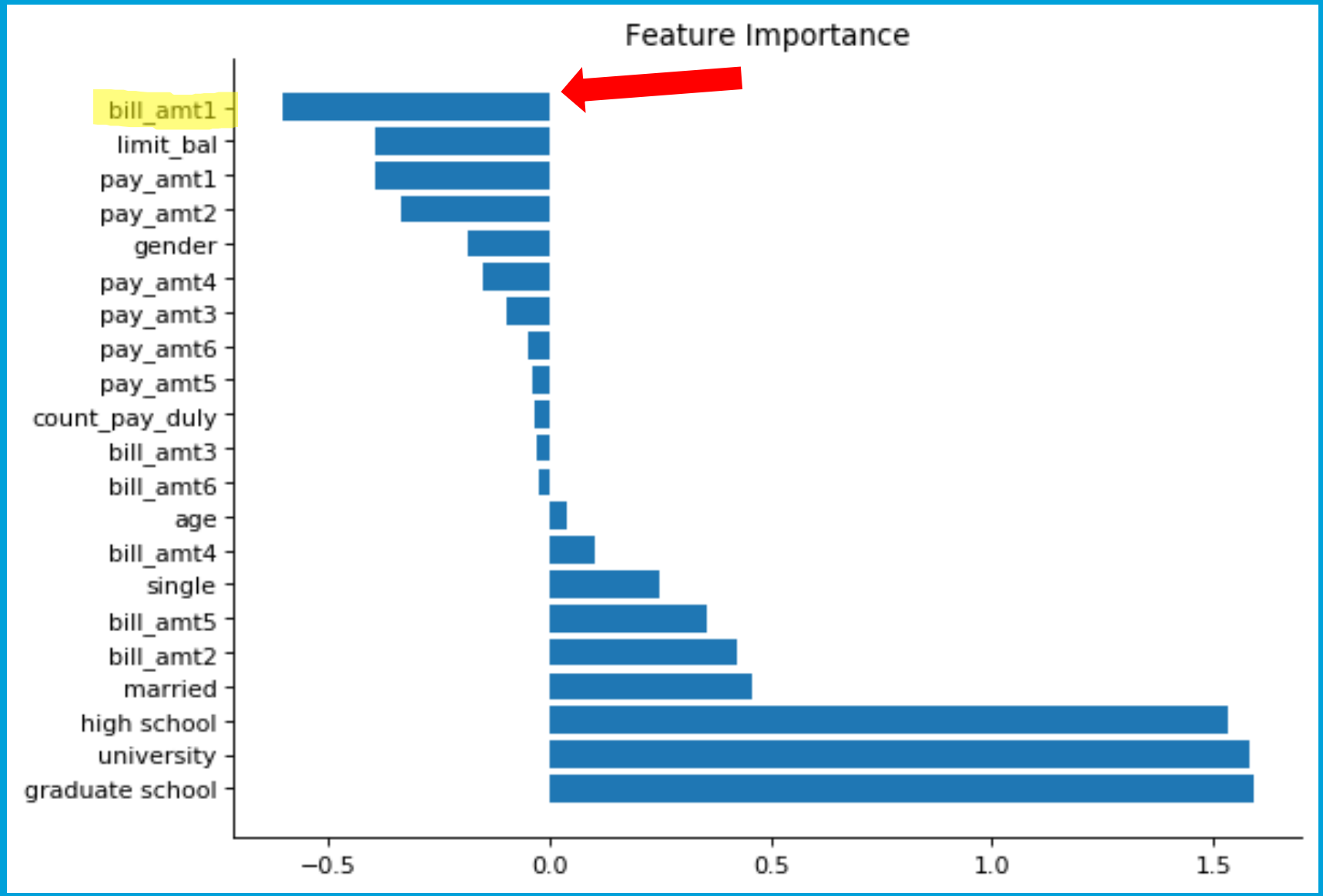
Results



Education ↑

Default ↑

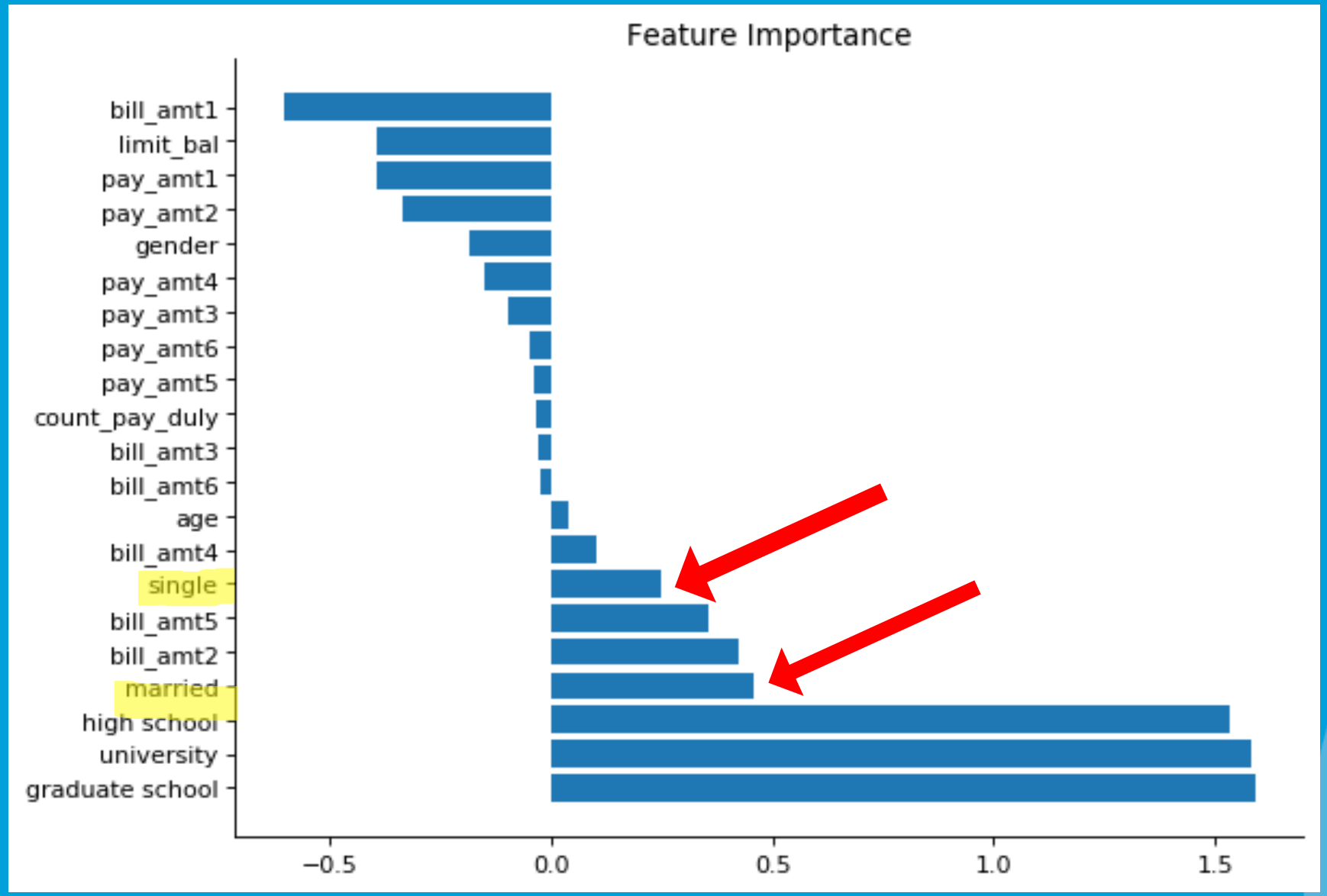
Results



First Month Bill Amount ↑

Default ↓

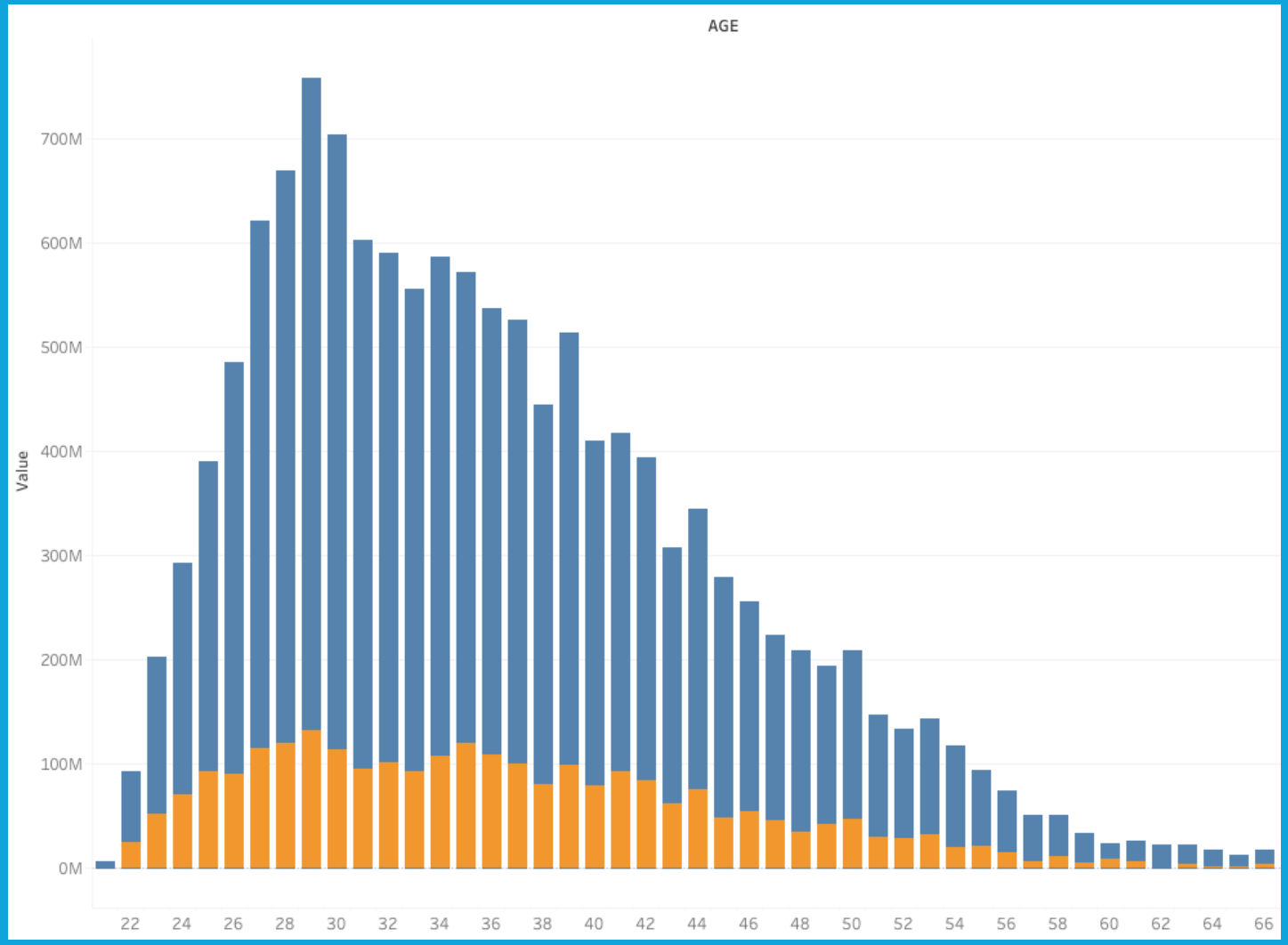
Results



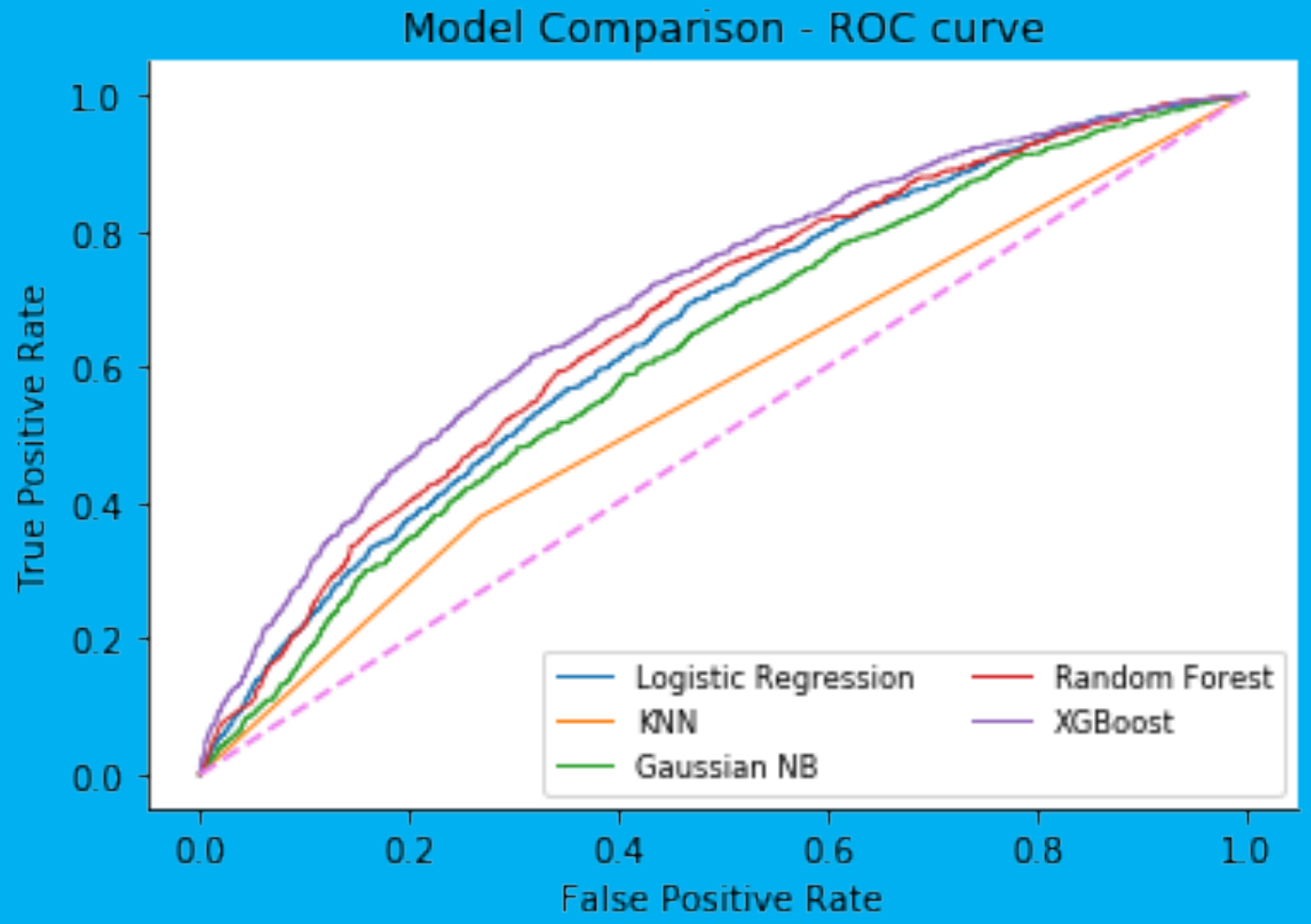
- Married clients are more likely to default than single clients

Future Work

- Feature Engineering
  - e.g. Categorize Age into bins
- Build a Flask App



Appendix


Model Comparison - ROC curve
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Logistic Regression
KNN
Gaussian NB
Random Forest
XGBoost

Appendix

- https://machinelearningmastery.com/classification-as-conditional-probability-and-the-naive-bayes-algorithm/
- https://scikit-learn.org/stable/model_selection.html
- https://towardsdatascience.com/beyond-accuracy-precision-and-recall-3da06bea9f6c
- https://www.ritchieng.com/machine-learning-efficiently-search-tuning-param/