STAT 536 - Case Study 4

Targeted Marketing

Sam Lee, Gavin Hatch

Model Evaluation

We first evaluate our models using the in-sample and out-of-sample metrics of the F_1 Score and Area under the ROC Curve (AUC).

Model	In.Sample.F1	In.Sample.AUC
LASSO	0.9018562	0.7036815
Probit	0.9097558	0.7045008
PCR	0.8520182	0.6589006

	Model	Out.of.Sample.F1	Out.of.Sample.AUC
LASSO	LASSO	0.8994684	0.6593630
Probit	Probit	0.9095978	0.7037272
PCR	PCR	0.8514603	0.6585632

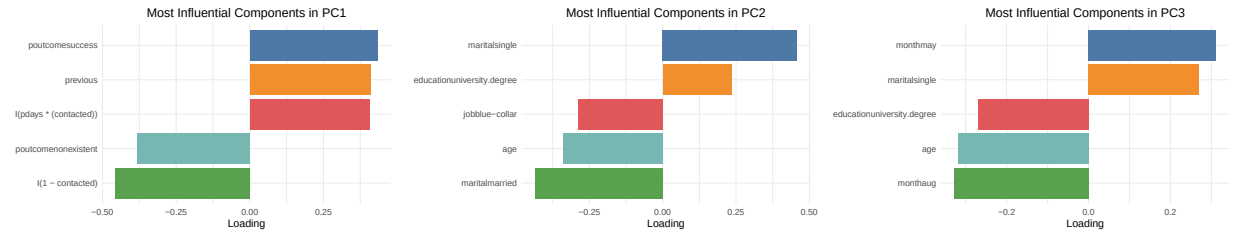

Figure 1: Most influential factors determined by PCR: As can be seen above, we classify the first partial component as *Marketing Interaction History*, the second partial component as *Customer Demographics*, and the third partial component as *Seasonal Timing*.

We classify the impact on the probability of opening an account through partial component analysis as seen in Figure 1. The first partial component captures the impact of past marketing interactions on customer receptiveness. This suggests that customers with favorable prior campaign outcomes are more likely to engage with the possibility of opening an account. We also point out the bisection here between the *timing* of when a customer was last contacted and whether they were contacted at all. Contacting customers is especially important, although the chances of a customer opening an account increases the more time has passed after they were contacted.

Figure 2 depicts the effect of contacting the campaigns have had on the probability of whether a customer has opened up a new account. We note that, in general, the more excessive contacting that takes place, the lower resulting expected probability that a customer will open up a new account. Despite the large spread, on average, social media contacts result in higher success rates than direct contacts.

Table 3: Estimates and 95% Confidence Intervals for Principal Component Regression

	Estimate (50%)	**95% Confidence Interval (2.5%, 97.5%)**
PC1		
I(1 - contacted) *	-0.1580	(-0.1826, -0.1422)
poutcomesuccess *	0.1501	(0.1346, 0.1726)
previous *	0.1420	(0.1256, 0.1647)
I(pdays * (contacted)) *	0.1406	(0.1263, 0.1617)
poutcomenonexistent *	-0.1316	(-0.1538, -0.1175)
PC2		
maritalsingle	0.0127	(0.0013, 0.0249)
maritalmarried *	-0.0119	(-0.0228, -0.0016)
age	0.0131	(0.0023, 0.0234)
jobblue-collar *	-0.0291	(-0.0365, -0.0211)
educationuniversity.degree *	0.0199	(0.0132, 0.0265)
PC3		
monthaug	0.0060	(-0.0003, 0.0142)
age	0.0131	(0.0023, 0.0234)
monthmay *	-0.0241	(-0.0314, -0.0172)
maritalsingle	0.0127	(0.0013, 0.0249)
educationuniversity.degree *	0.0199	(0.0132, 0.0265)
Intercept		
Intercept	-0.0001	(-0.0030, 0.0021)

Note: We include the top five most significant factors for each partial component. Significant factors (where the 95% CI does not include zero) are marked with an asterisk (*). Confidence interval was calculated using $B = 100$ # of bootstrap samples. Estimates are given the interpretation of its effect on the log-odds of whether a customer opens an account. Covariates are *scaled*.

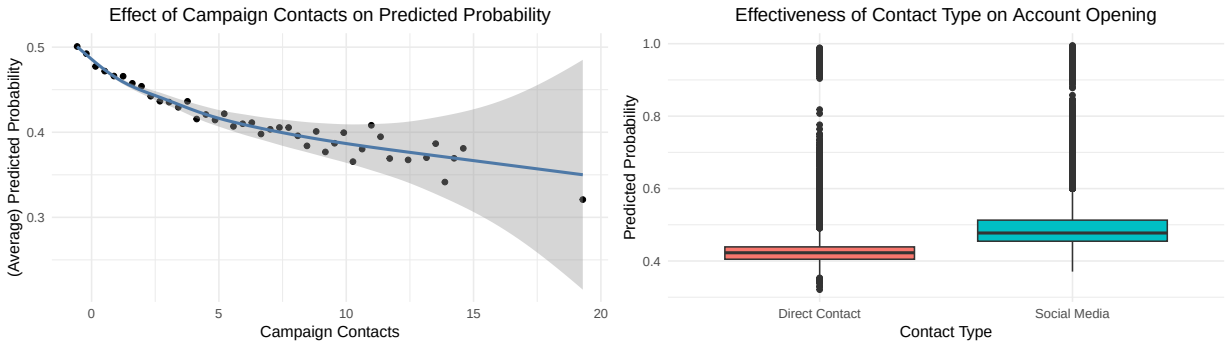

Figure 2: Model predictions with respect to contact strategy. The figure on the left was fit with a GAM smoother.