**File Name: bank.xls**

**Data Source:** Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., & Lichtendahl Jr, K. C. (2017). *Data mining for business analytics: concepts, techniques, and applications in R*. John Wiley & Sons.

Universal Bank would like to know which customers are likely to accept a personal loan. What characteristics would forecast this? If the bank were to consider expending advertising efforts to contact customers who would be likely to consider a personal loan, which customers should the bank contact first? By answering this question correctly, the bank will be able to optimize its advertising effort by directing its attention to the highest-yield customers

**Attribute Information:**

|  |  |
| --- | --- |
| ID | Customer ID |
| Age | Customer’s age in completed years |
| Experience | No. of years of professional experience |
| Income | Annual income of the customer ($000) |
| ZIP Code | Home address, ZIP code |
| Family | Family size of the customer |
| CCAvg | Average spending on credit cards per month ($000 |
| Education | Education level (1) Undergrad; (2) Graduate; (3) Advanced/Professional |
| Mortgage | Value of house mortgage if any ($000 |
| Personal Loan | Did this customer accept the personal loan offered in thelast campaign? |
| Securities Account | Does the customer have a securities account with thebank? |
| CD Account | Does the customer have a certificate of deposit (CD)account with the bank? |
| Online | Does the customer use Internet banking facilities? |
| CreditCard | Does the customer use a credit card issued by UniversalBank? |
| PersonalLoan\_Training | Training Data for Y variable (80:20) split |

1. Create dummy variables for the categorical predictors.
2. Explore the data and identify suitable predictors.
3. Build a logistic regression model
4. Comment on the significance and the explanatory power of the model
5. Write the estimated equations in three formats:
   1. The logit as a function of the predictors
   2. The odds as a function of the predictors
   3. The probability as a function of the predictors
6. Assess model fit.
7. To increase the percentage of correctly classified responders, should the cut-off probability be increased or decreased? By how much?
8. What is the next step?

**Recommendations**