USC Marshall

DSO528

Case 2 _ Logistic regression application
Trojan Horse: Unexpected Style at Your Door
Case Questions

- *Please submit a report addressing all the case questions. If you're preparing the report in Word, kindly include corresponding graphs, if applicable, within the description or in an appendix. If you're using a Colab notebook, ensure that the write-up is clearly labeled and organized within the notebook.*

- *The write-up portion should be no more than two pages.*

- *Please submit altogether any supporting analysis.*

- *Only one submission per team is sufficient. Please ensure that all team members' names are listed on the report cover.*

Case Questions:

1. Set the seed to **528** and randomly select **30%** of the data as the test set and **70%** as the training set. Build a decision tree model using the training data. You can use Colab to split the data and either Colab to build the model.

2. Select variables at your choice and train a logistic regression model to predict Success:

a. Briefly explain how and why you select those variables. Are these variables significant at 5%?

b. Describe your model using basic metrics such as pseudo-R^2, model equation and coefficients.

c. Explain the impact/business value of each variable on buy propensity. Do the coefficient values make sense?

d. In general, does the logistic model align with business logic?

3. Apply the model to make predictions on the test set. What cutoff probability will maximize the expected profit or payoff on the test set? Is the expected payoff on the test set with this cutoff better than what you found using the tree model in Case 1?

4. Following the policy from the previous case, the company limits the number of customers targeted in each campaign to no more than **10%** of a customer base of **500,000**. In other words, no more than **50,000** customers can be labeled as buyers. Given this policy and your logistic model, what is the maximum expected payoff? (Hint: Reconsider your analysis from the previous question.)

5. Compare the logistic regression model and the decision tree model you presented in case 1, overall, which one would you recommend? Explain.