

Trojan Horse: Unexpected Style at Your Door¹

Case questions require the following auxiliary data files, available on BB.

- *trojanHorseData.jmp*

Please write brief responses, and be prepared to discuss and present your work with the class. You are free to work in teams and use the Internet, but must write up your own solutions and create your own excel files.

Case Questions:**Decision Tree**

1. Using the training data, build a decision tree model to predict probability that a member will purchase a box. If you set a cut-off threshold of .15, how many members would you target in the training set? Describe (in words) the types of members that will be targeted.
2. What is the R^2 of your model? In simple words, what does this mean?
3. Examine each of the split variables. Do they make intuitive/business sense? Explain each one.
4. Using the testing data set and a cut-off threshold of .15, build the confusion matrix for this model. If you were to target at most an additional 48,000 members using this model, what do you estimate your profit to be?
5. (More challenging) The threshold .15 was somewhat arbitrary. Build a curve computing the expected profit on the 48,000 members for different values of the threshold. What threshold value would you propose?
6. For your chosen threshold value, what does the confusion matrix look like? How does this compare to your answer in part 4?

Logistic Regression

7. Build a logistic regression model to predict the probability a member will purchase a box. You should use JMPs "Step" feature to prune the variables.

¹ This case was developed for USC Marshall's BUAD 425 by Prof. Arif Ansari and Prof. Vishal Gupta.

8. What is the estimated logistic regression equation? How would you explain this equation intuitively to your boss?
9. What is the R^2 of your model? How does this compare to your answer in 2? Can you assert which model is better? Why or why not?
10. Which coefficients are significant at the 5% level? Do their values make intuitive/business sense? Explain each one.
11. Using the testing data set and a cut-off threshold of .15, build the confusion matrix for this model. If you were to target an additional 48,000 members using this model, what do you estimate your profit to be? How many members would you target?
12. (More challenging) The threshold .15 was somewhat arbitrary. Build a curve computing the expected profit on the 48,000 members for different values of the threshold. What threshold value would you propose? Compare this to the curve you built earlier for the decision-tree.

Towards a Recommendation

(Be prepared to present your responses to these questions to the class.)

13. Based on the results above, which methodology would you propose to Trojan Horse and why?
14. What other issues should Trojan Horse consider when adopting this new targeted campaign?
15. Is there anything you would recommend to improve the targeting campaign?