Freemium (Part 2)

Akshay Singh – akshaysi
Muskan Sharma - muskans
Pawanjeet Singh – pawanjes
Yashvi Thakkar – ypt

Our objective is to quickly improve the number of paid subscribers of High Note. We try applying the Decision Tree model to predict the likelihood of converting free users to premium users. After evaluating the initial Logistic Regression model, we found that it had correctly predicted that 61 users would convert from free to premium (True Positives). However, it also predicted that 99 users would have converted to premium, when in reality they would not have! (False Positive) It also misclassified 669 users as those who would not have converted, but in actuality, they were true adopters (False Negatives). We heeded Lisa's suggestion of exploring a Decision Tree model. Our best Decision Tree model gave the following results - 130 True Positives (correctly predicting users who actually converted), 291 False Positives, and 600 False Negatives. The Decision Tree was better at correctly predicting those who would actually convert to premium users, but it also had a poor performance in terms of a very high False Positive. To further compare the 2 models, we computed the predictions for the 10% of most likely adopters. We then calculated the ratio of how many we predicted to be adopters, to how many were actually adopters purely due to chance. (This is referred to as the Lift). The Lift for Logistic Regression (3.162) was better than the Lift we got for the Decision Tree model (3.088) .[2] We also looked at the ROC curve which essentially explains the diagnostic ability of the classifying model and found that Decision Trees (0.735) performed slightly better than Logistic Regression (0.734).[3].

We therefore dig deeper into a profitability analysis to finalize between the 2 models. We create 2 Confusion Cost matrices. One being the Cost matrix when the promotional offer was given to all users, the other being the Cost matrix when no action is taken. These 2 matrices represent the most expensive scenario and the baseline scenario, respectively. We then calculate the expected payoff (Profits) under each scenario to evaluate the Logistic Regression and Decision Tree models. We find that the baseline profit made when we take no action is $41,163. (Irrespective of the models). That is, when the promotional offer is not extended to any user, HighNote makes $41,163 in profits. If we extend the promotional offer to all users that the Logistic Regression model predicts to be True, High Note stands to make $40,576 in profits. And the same for the Decision Tree model is $39,883. Note that while the Profits tend to decrease for this offer strategy, the expected profit using Logistic Regression is higher than the Decision Tree. Therefore, we recommend using the Logistic Regression model.

After taking historical data into account, the model was able to explain the variation in the data better. Historical data seems to have an impact on whether a user adopts the premium subscription. However, the results do not seem to be extremely significant as it only improved the model accuracy by 0.4% and precision of predicting adopters by 0.02%. Not all changes are impactful – for instance, decrease in subscriber friend in the past 3 months does not necessarily imply that the user will also revoke their subscription. We can keep tracking historical changes and see if these results become more significant in the future. Visualizing the decision tree also made us aware that it was more important to understand that a customer is using the platform more frequently than as opposed to analyzing their total usage. Analyzing change in the number of songs listened gives a better understanding of whether a customer will be likely to adopt the premium service. This is confirmed by our decision tree model – the change in number of songs listened to provides the cleanest split, classifying over 65000 users. Hence, we should focus on whether the customer has increased activity when it comes to listening to songs as they are more likely to make the switch to premium. Moreover, the model also informs us that the total usage gives a poor estimate of adoption. A user may be frequently using the app over the first half of the current period and slightly inactive in the second – indicating a waning interest in the platform. Causes for these would most likely be curiosity during the first half and annoyance over adverts during the second. Because these customers are unwilling to pay for the premium subscription, they tend not to switch. In such cases, if we use total usage as a feature for prediction, we would inaccurately assume that the customer is a potential premium adopter because the overall usage will be high.

We now evaluate which users to target based on the profitability of launching the promotional offers. We proceed by making use of the 2 Confusion Cost matrices mentioned earlier. An important point to note here is that we evaluate the profitability of our actions by analyzing a test dataset of ~10,653 users. The profit High Note will make without taking any action (without offering free 3-month premium access to any user) is $41,153. This is the baseline against which we compare our models' performance. We make an assumption that if the promotional offer is extended to a user, their odds ratio of upgrading to premium will increase by 5 times. Using the updated odds-ratio, we back-calculate the probability that each user will upgrade to premium. We then tried a bunch of different approaches to find a balance between saving opportunity costs and mitigating loss in ad revenue, like - not targeting people highly likely to subscribe and excluding those with a very low probability to convert, and thereby getting people moderately likely to subscribe (40th to 70th percentile). However, through trial and error, we find that the maximum final profit (using the Logistic Regression model) would be ~$40,307. This does not even exceed the profits High Note would make without taking any action. Hence, we find that the strategy of offering access to premium services will not work as well as initially thought. We cannot be certain if such a strategy will help immediately increase the number of premium High Note users. The business might instead run on a loss. Therefore, the management should consider other alternatives to drive the number of High Note premium users.

APPENDIX

Exhibit - 1 (Confusion matrix of Logistic Regression and Decision Trees)

LR Model Test Data		TRUE	
		trueadopter	
Predictions	cadopter	0	1
	0	9823	669
	1	99	61

Decision Tree Test Data		TRUE	
		trueadopter	
Predictions	cadopter	0	1
	0	9631	600
	1	291	130

Exhibit - 2 (Lift chart - Combined for Logistic Regression and Decision Trees)

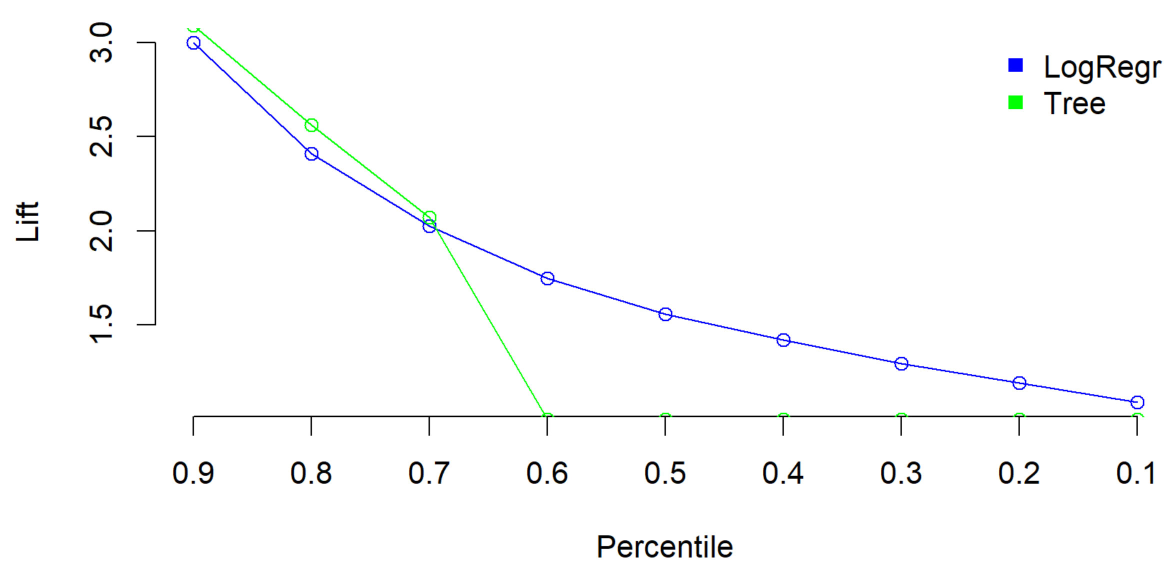

Exhibit - 3 (ROC Curves for Logistic Regression and Decision Trees)