Freemium(Part2)_Team G (Jingyi Ran, Suguru Isoda, Trista Lin, William Peng)

Required 1. Comparison between logistic regression model and a new decision tree

Per discussion earlier, the team has run the Decision Tree model. The comparison between this model and the Logistics Regression model ran previously is shown in Appendix 1. As you can see from the charts in Appendix2, results of the Decision Tree resemble that of the Logistic Regression.

However, we see nuances between the two. To choose target customers to give promotional offers, we focus our attention on users that seem not to want to pay for premium because with 3 month free access to premium, their probability of converting will increase. In this case, we want our model to make predictions as accurate as possible and the accuracy of decision tree is higher than that of logistic regression.

In addition to accuracy, we pay attention to AUC value because we need to consider the balance of type I and type II error, which means we need to find a good cut off between falsely predicting adopters to non-adopters and falsely predicting non-adopters to adopters. AUC value is a suitable indication. Considering the balance among accuracy, type I and type II error, the decision tree performs better.

Required 2. Suggestion for using the change from the historical period

Further, we have also examined the differences between the current period and the changes from the historical period. Echoing your suggestion, we also recommend building our model based on both pre and current data (Appendix4). From a technical perspective, consistent with previous model comparison in which we focus on AUC score, the decision tree with both pre and current data (AUC = 0.78) performs slightly better than the decision tree with current data only (AUC = 0.77) (Appendix5). More importantly, when interpreting the decision tree in the business context, including the pre data in the decision tree model could shed more insights on the factors in predicting whether the free user would convert to premium user in the current period. Similar to the model with current data (Appendix 3), the model with both pre and current data showed that the number of tracks loved is important in predicting the conversion. However, we found that, with pre data, the number of new tracks loved in the pre period is the most important factor in predicting the conversion. In other words, not only how many tracks the user currently loved, but also how many new tracks the user loved previously could influence, and even have a larger impact on, whether the user would convert or not. Moreover, in this new model, we found that the number of new songs listened to in the pre period also has an impact on the conversion, whereas, as opposed to the previous model, the cumulative number of songs listened to until the current period is no longer important in predicting the conversion. This result is more reasonable since a user could have been always listening to many songs and does not find a need to become a premium user, so the cumulative number of songs may not matter in predicting whether the user would convert or not. In contrast, if there is a sudden increase in the number of new songs the user listened to previously, it means that the user starts to use the platform more, and may be more inclined to become a premium user. Overall, the decision tree with pre and current data and the decision tree with current data only shared mostly the same important factors from the current data such as number of tracks loved, tenure and playlist, but the new model yielded a slightly different conclusion by showing the influence of changes in tracks loved and songs listened. Given that the decision tree with pre and current data has a higher AUC score and generates more insights, we decided to use the decision tree with pre and current data as our final optimal model.

Required 3. Targeted user and preferable result of promotional offer

In proceeding a new promotional offer, there are three business issues that need to be concerned to maximize the result of our investment on promotion.

1. Which users should we target?
2. What would the results be?
3. What kind of marketing activities are needed?

<1. Which users should we target?>

According to confusion matrix (Appendix6), the group that is reasonable in maximizing the effectiveness of promotional activities is the True Negative segment (29,100 users), because these groups will remain as Free Subscribers if there is no action from us such as marketing promotions. For the other segments such as False Negative segment and the True Positive segment, the return of marketing promotion measures for those segments is expected to be low because they have grown to become Premium Users even without marketing measures. As for False Positive segment, investing in marketing promotional activities on this segment is a waste of money, since the result will be no Premium Users from our promotion offer. Therefore, we should target the True Negative segment for promotional activities.

<2. What would the results be?>

In order for it to be reasonable in terms of decision-making to engage in promotional activities, the revenue in case of proceeding promotion must exceed the revenue from the case without promotional activities. From simulation table (Appendix7), in case of proceeding promotion, if there are more than 3.5% Premium User conversions from the targeted True Negative Segment, it is more reasonable in terms of revenue to promote than not to promote. Therefore, we should judge whether expecting a conversion rate of 3.5% is feasible as the result of promotional activities. By benchmarking other subscription services, Spotify shows 46% conversion rate, and by following Spotify's example, we can exceed our conversion rate target as well, such as reducing webpage friction and optimizing playlists to have specific appeal for target segments.

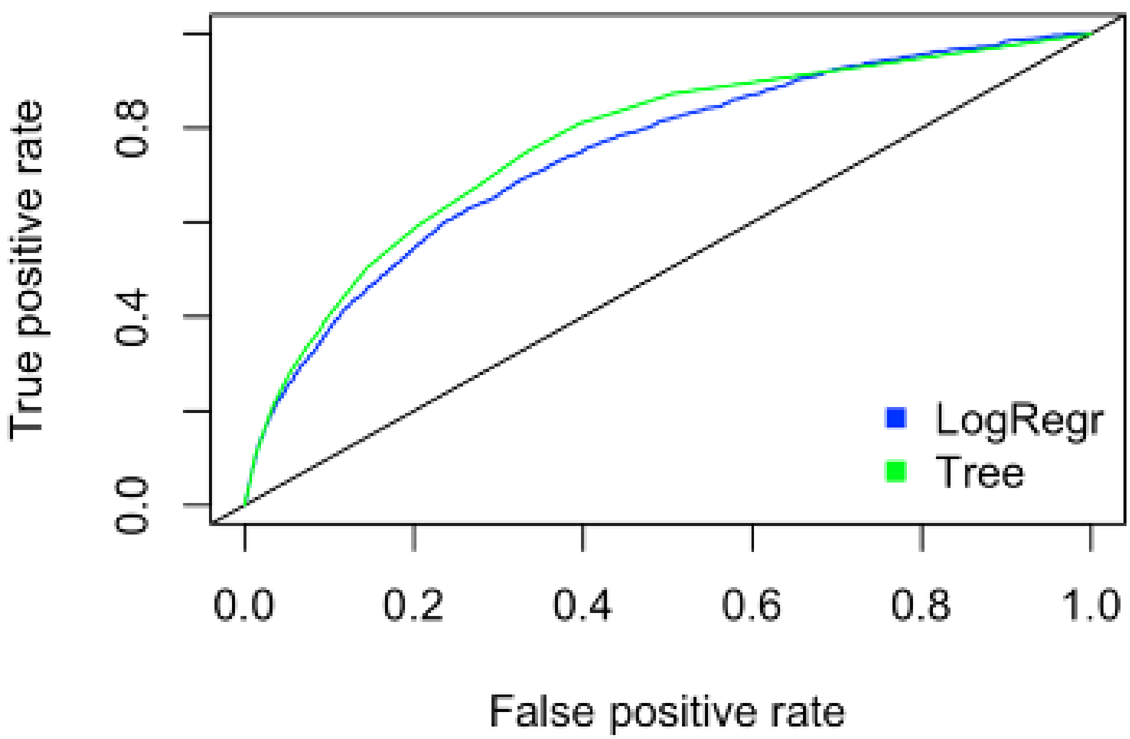
<3. What kind of marketing activities are needed?>

From our analysis above, the number of new tracks loved has a significant impact for converting Free Users into Premium Users. Therefore, the marketing approach which promote number of loved tracks has a positive impact. For example, for users in the True Negative segment, the target segment, the number of Loved Tracks can be used as a KPI to monitor the status, and if the number of Loved Tracks increases in targeted users, a three-month free subscription can be offered to them on a priority basis. Also we can offer that stop running advertises for a certain period of time, if the Number of Loved Tracks increases. We would like to achieve the target conversion rate by setting the variable like Loved Tracks increases as a KPI that captures the signs that a user may change into a Premium User.

Appendix1. Comparison between logistic regression model and a new decision tree

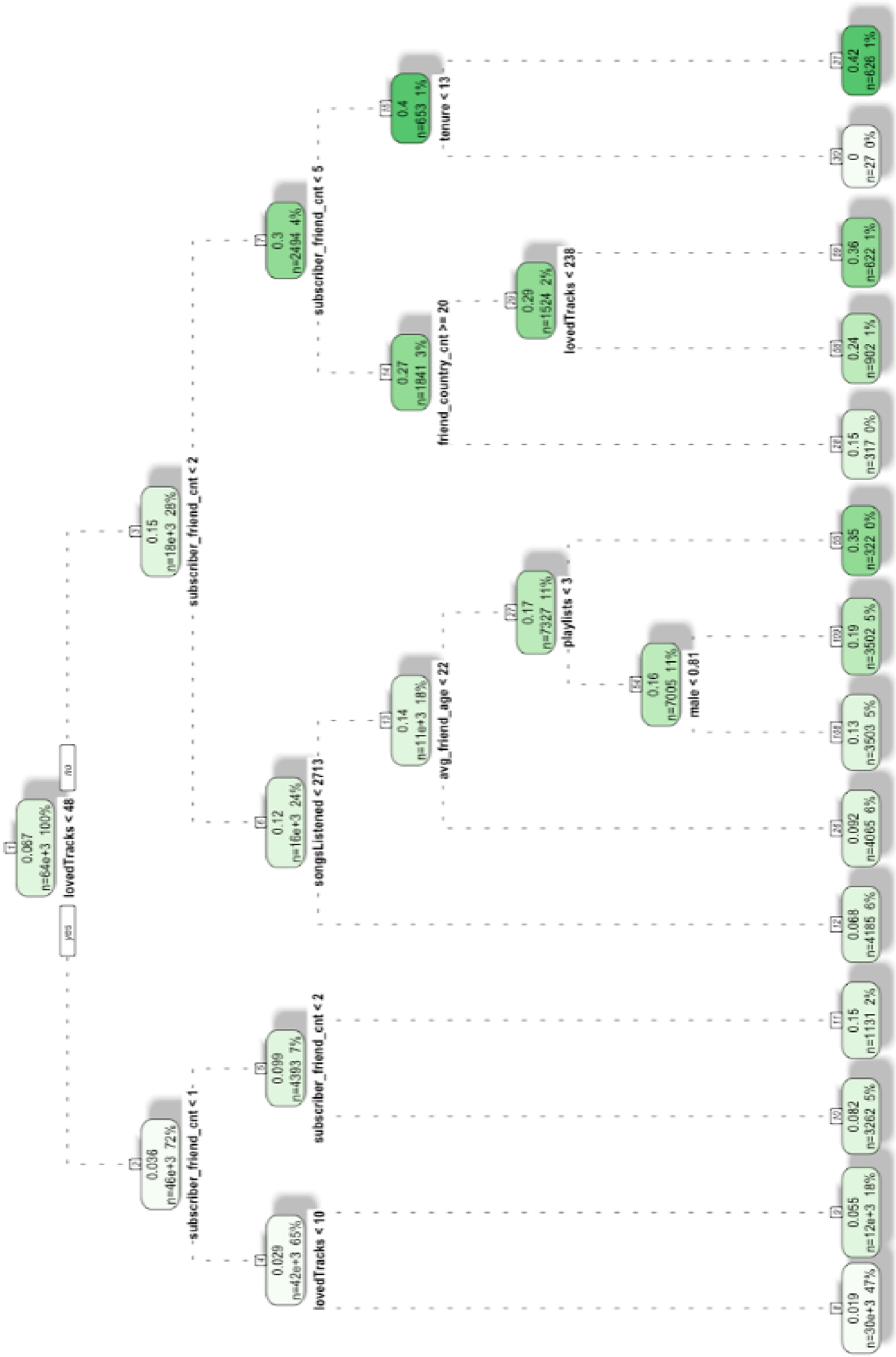
	Logistics Regression	**Decision Tree**
Accuracy	**0.93**	**0.96**
Precision	**0.35**	**0.34**
Recall	**0.09**	**0.14**
AUC	**0.74**	**0.77**

Appendix2. Accuracy Rate Comparison between two models



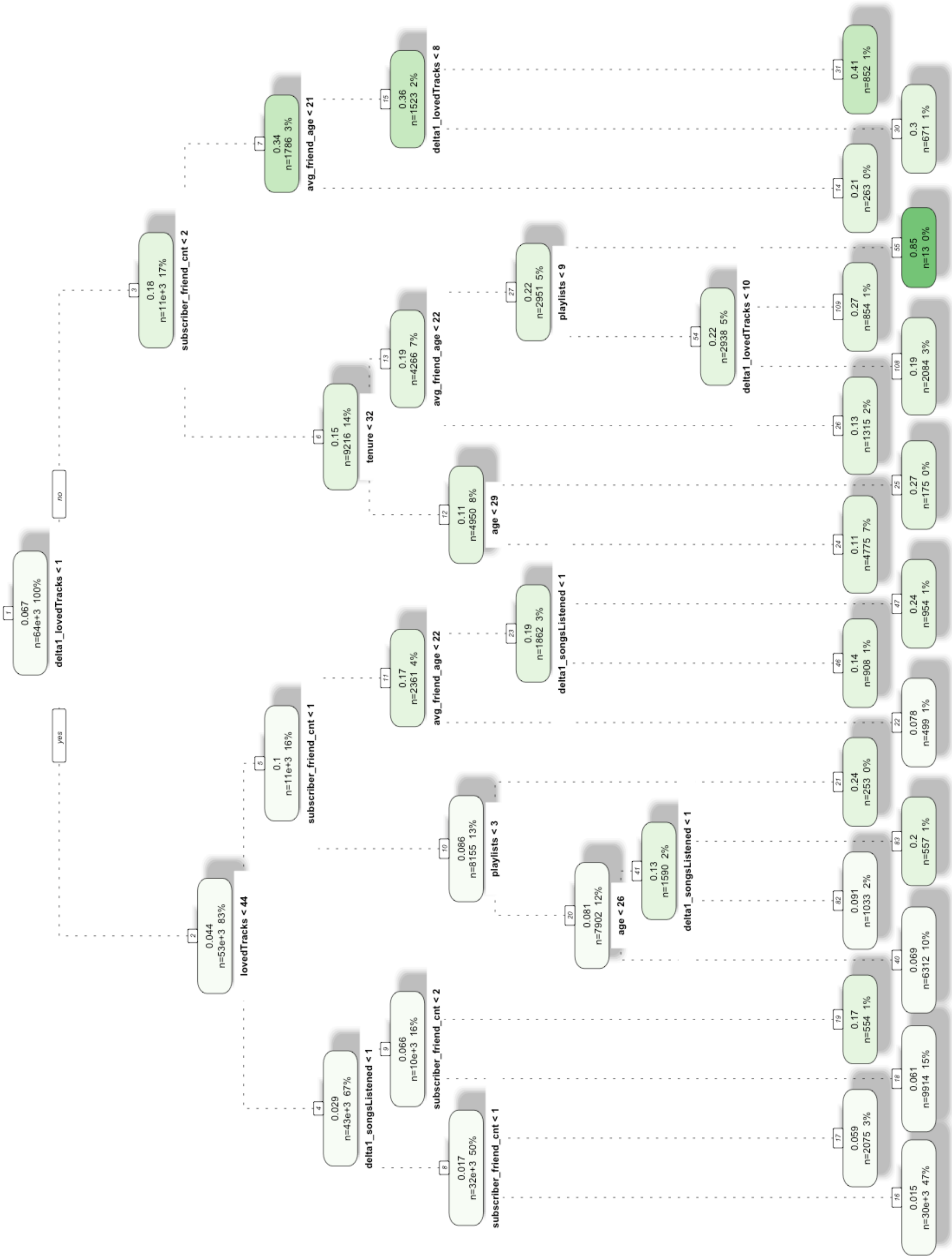
Freemium(Part2)_Team G (Jingyi Ran, Suguru Isoda, Trista Lin, William Peng)

Appendix3.First Decision Tree (Current)



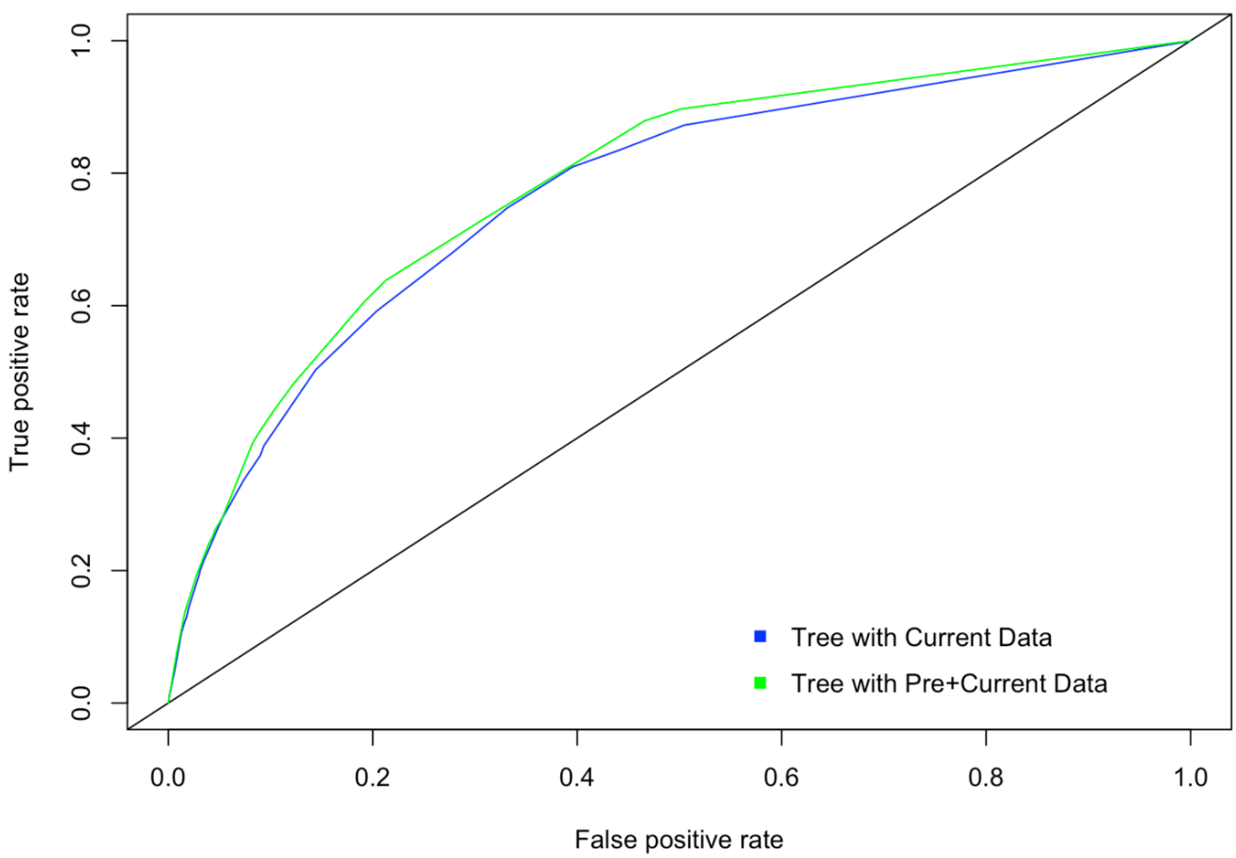
Freemium(Part2)_Team G (Jingyi Ran, Suguru Isoda, Trista Lin, William Peng)

Appendix4.Second Decision Tree (pre+current)



Appendix5. Comparison between Decision Trees using different term

	Decision Tree with Current Data	**Decision Tree with Pre+Current Data**
Accuracy	**0.96**	**0.92**
Precision	**0.34**	**0.33**
Recall	**0.14**	**0.19**
AUC	**0.77**	**0.78**



Appendix6. Confusion Matrix

```
           trueadopter
cadopter       0     1
       0   29100  1737
       1     838   419
```

(0,0): True Negative Group1
(1,0): False Positive Group2
(0,1): False Negative Group3
(1,1): True Negative Group4

Appendix7. Revenue Calculation

no promotion offer

	Number of Users	Number of months	fee/month	revenue
Group 1	29100	12	0.125	43650
Group 2	838	12	0.125	1257
Group 3	1737	12	3	62532
Group 4	419	12	3	15084
			Total Revenue	122523

Promotion offer

First 3 months

	Number of Users	fee/month	revenue
Group 1	29100	0	0
Group 2	838	0.125	314.25
Group 3	1737	0	0
Group 4	419	3	3771

Next 9 months

	Number of Users	fee/month	Conversion Rate	Revenue
Group 1	29100	3 for converted users; 0.0125 for other users	0.035254744	59283
Group 2	838	0.125		942.75
Group 3	1737	3		46899
Group 4	419	3		11313
			Total Revenue:	122523

The reuslt shows that if the conversion rate could reach 3.5%, the company can make profits