ECO 395 Project: How to acquire more "kiss" on the Lovoo?

Yu-Ting Huang, Jipeng Cheng, Weidi Hou

5/8/2022

Abstract

Introduction

The importance and influence of dating apps cannot be ignored in today's society. For example, "The Tinder Swindler" is one of the most famous movies in 2022, specially, global views was 45.8 million hours in its first week of release and hit the top 10 in 92 countries in Netflix. Furthermore, the application "Tinder - Dating New People" is super popular in the US and over 55 billion matches are made. Of course, except for Tinder, there are so many dating apps such as OkCupid, bumble, and Coffee Meets Bagel to name but a few, they have different features to attract different groups.

In this study, the data from Lovoo will be analyzed. The purpose of this research is to build the model and point out what profile and specific factors can acquire more "likes" (called "kiss" on the Lovoo), as an indicator to capture the potential charisma of people. On the other hand, as we know that more "likes" means the person may get more matches and more potential encounters. The one thing that should be noticed is that our data only collects samples from "female". Therefore, the results will be fitted for females rather than male.

To have a more meaningful explanation, this study created a new variable called "conversion rate", which can avoid the problem that the number of likes has highly correlation with the number of clicks (visit times) compared with using original `counts_kisses` as a dependent variable. Also, this indicator can effectively explain how to obtain more likes from other users by other independent variables.

$$\texttt{counts\_kisses} = \frac{counts\_kisses}{counts\_profileVisits} \ (1)$$

Throughout the series of following analysis, the consequence of this project can help many young people, who would like to make friends and/or have a pair, including our friends, classmates, and family, to have more wise strategies when using Lovoo as well as other dating app.

Data and Methods

Data Processing

The original data comes from "Dating App User Profiles' stats - Lovoo v3" gathered during spring April and May 2015. IOS version of the Lovoo app was in version 3 at that time. The original data includes 2940 rows and 39 variables, finally, this research left 22 as valuable variables and create 1 new variable.

The details of each variable presents in the following table.

Variable	Description
age	user age
counts_kisses	Number of unique user accounts that "liked" (called "kiss" on the platform) this user account, following formula (1), as a index for converting times of visits to times of likes
counts_details	The degree of account completion
counts_pictures	Number of pictures on the user's profile
counts_profileVisits	Number of clicks on this user (to see his/her full profile) from other user accounts
flirtInterests_chat	1 if the user indicated being in search for people to chat with
flirtInterests_friend	1 if the user indicated being open to making friends
flirtInterests_date	1 if the user indicated being open to dating people
isVip	1 if the user is VIP (this status came with benefits)
isVerified	Whether the user's account was verified through one of the methods (Facebook, phone number, . . .)
lang_count	Number of languages the user knows
lang_fr	1 if the user can speak French
lang_en	1 if the user can speak English
lang_de	1 if the user can speak German
lang_it	1 if the user can speak Italian
lang_es	1 if the user can speak Spanish
lang_pt	1 if the user can speak Portuguese
freshman	1 if the user register no more than one month
hasBirthday	1 if the user has birthday
highlighted	1 if the user's profile is currently highlighted (at fetch time)

Model

In order to analyze which feature will help app users attract more people, we decide to **Lesso regression** and **Random Forest** methods to study it. As we all know, Lasso regression as a linear regression will give us a better fitted coefficient value. Random forest method, on the other hand, will give us a better fit but it can't give us a good fitted coefficient value. In order to solve the above problems including in each method, we decide to combine these two methods in order to have a deep study on which characteristics will help app user attract more people.

Results

Lesso

Lesso approach is to make some regularization so that the regularized fit minimizes the deviance plus a **penalty** on the complexity of the estimate:$minimize_{\beta \in R} dev(\beta)/n + \lambda \times pen(\lambda)$ (2).

Here λ is the penalty weight, while **pen** is some cost function that penalizes departures of the fitted β from 0. In order to use **gamlr** function in r programming, we need to create our own numeric feature matrix.

Cross-validation Lesso

Then we use cross-validation Lesso regression method so that the result will be more robust.

Variable	beta_hat
intercept	-3.457606879
age	0.013822836
counts_pictures	0.004166058
lang_fr	0.040604945
lang_de	-0.127865953

Notice that beta coefficient for the following variables including `flirtInterests_chat`, `flirtInterests_friends`, `flirtInterests_date`, `isVIP`, `isVerified`, `lang_en`, `lang_it`, `lang_es`, `lang_pt`, `countDetails`, `freshman`, `hasBirthday`, and `highlighted` are **zero**.

The result shows that the optimal $log\lambda$ which minimize the test set mean square error is -8.99 and under this level of penalty, the fitted coefficients that are not zero are `age`, `counts_pictures`, `lang_fr`, `lang_de`. From these results, we conclude that age, number of pictures on the users' profile and proficiency in French have a positive effect on being loved by more people. However, proficiency in German has a negative effect on their level of charm. These results tell us that being mature will help people attract more people and more pictures on their app profile will also help them become popular since more pictures will tell other people more information about this themselve and help other people know the user better. In addition, people who have a good looking or who are more out-going and more confident prefer to post their picture on their app profile. These type of people also attract more people. Besides, French people are more romantic and they have a higher probability to attract more people on the app. On the other hand, German people are more serious and introverted, that's why the fitted coefficient of knowing German is negative. In addition, from our regression result, we found that our feature matrix is very sparse (i.e, mostly zero), this is especially true since we have lots of factors as features. lasso regression is a great way to improve the efficiency of our model since it screens and ignores zero elements in actually storing X.

Random Forest

A random forest starts from bagging. We still take B bootstrapped samples of the original data and fit a tree to each one, and we still average the predictions of the B different trees. However, it adds more randomness. With each bootstrapped sample, we don't reach over all the features in x when we do our greedy build of a big tree. Instead, we randomly choose a subset of a features sub sample to use in building that tree. The advantages of using fewer features in each tree are that it simplifies each tree, reducing its variance and it diversifies the B trees, decorrelating their predictions.

From the load.forest plot, we found that more trees shows smaller out-off sample MSE. After 500 tress, the partial decresing of MSE becomes very small.
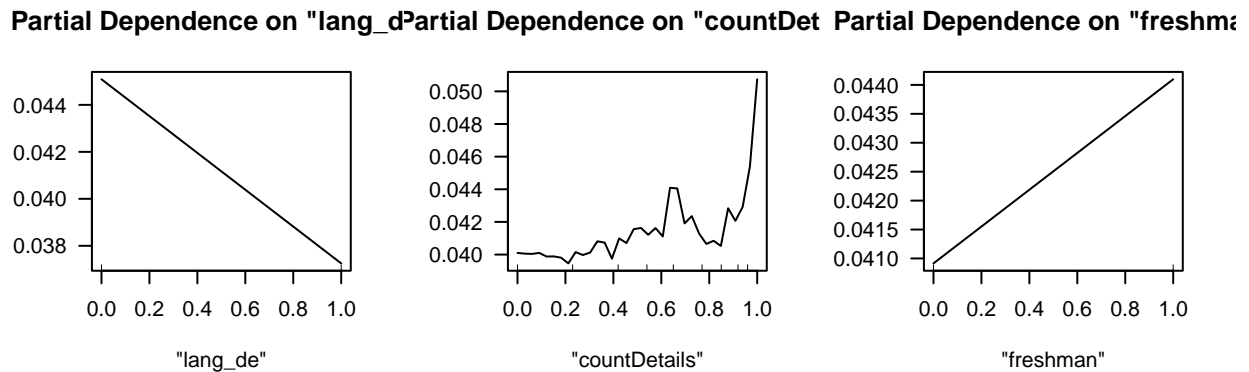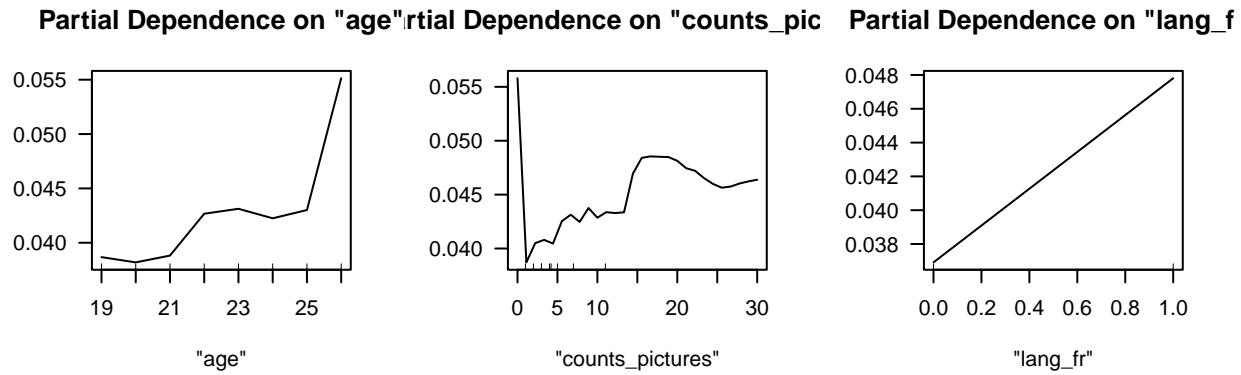
Then we study how random forests can give us a variable importance measure.

lvforest

IncNodePurity
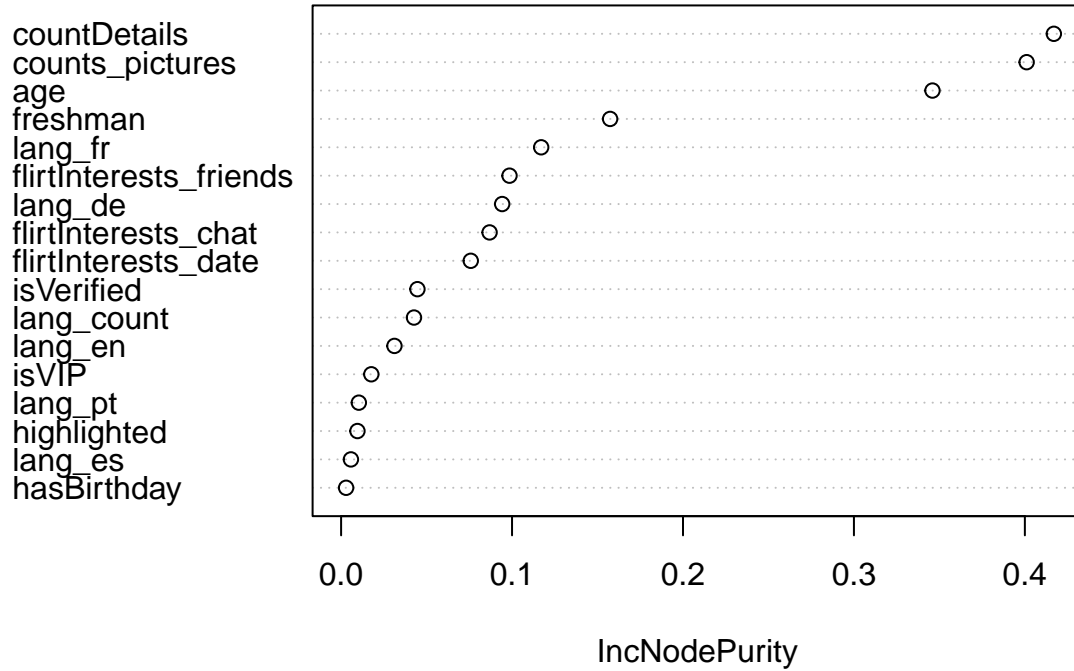
The x axis represents that percentage increase in mean square error and the y axis represents different variables. The variable importance plots shows how much omitting each of these variables inflates the MSE of the prediction, higher is worse. From our result plot, we found that `counts_Details`, `counts_pictures`, `age`, `freshman`, `lang_fr` and `lang_de` have higher value, which means these variables are more important in this model, they have higher influence on whether an app user will attract more people. This result is rational since completed information on their profile and more pictures on users' profile will provide more information of the user. In addition, users' basic information such as their age and whether they are new user will also provide some information their background and their communication proficiency, which are also key factors decide whether they can attract more people. Finally, people's language will tell us where they come from. People from different country may have different characteristics and their social development environment will also impact their probability of attracting more people.

Then we study the partial importance of each variable which we observe are more important in our variable importance plots.

Partial Dependence on "age" **rtial Dependence on "counts_pic** **Partial Dependence on "lang_f**

"age"

"counts_pictures"

"lang_fr"

Partial Dependence on "lang_d **Partial Dependence on "countDet** **Partial Dependence on "freshma**

"lang_de"

"countDetails"

"freshman"

First, from the PD on `age`, we found that when people above 20 years old, they are more likely to attract people, especially when people are above 25 years old, the slope of `age` is much higher. It's reasonable since age 20-26 is a period of dating and getting marriage so people during this age are more likely to log onto this app to find their dating mate. Second, from PD on `counts_pictures` and on `counts_Details` the result is rational since more pictures and more completed profile will tell people more stories about the user which have a positive effect on attracting more people. Third, from language perspective, people speaking French are more likely to attract people but people speaking German are less likely to attract people, since people's characteristic are influenced by their nationality and their growing up environment. Finally, from the PD of the `freshman`, it seems that new user are more attractive, since people always like new things.

Conclusion

As female user, in order to attract more people, they need to include more details in their profile and also post more pictures. These behaviors will provide more information about themselves. In addition, mature female seems to attract more people and new users seem to be more attractive.

Appendix

Reference

1. Tinder, https://tinder.com
2. 'The Tinder Swindler' Becomes First Doc To Lead Netflix's Weekly Film Chart, https://deadline.com/2022/02/the-tinder-swindler-first-doc-lead-netflixs-weekly-film-chart-1234928573/
3. Lovoo, https://about.lovoo.com/en/#app-features