Name: Andrea Fox

Class: DSC630-T301 Predictive Analytics

Date: October 1, 2020

Assignment: 5.4 Assignment – Predictive Analytics Case Study

**Overview**

Match.com is an online, subscription-based dating website where single individuals can go to try and meet a potential mate. This case study is about Match.com and their use of “intelligent matching” to predict potential matches that you would be interested in communicating with (Siegel, 2016).

**Business Understanding**

When Match.com started originally it was 1995 and the idea of online dating was not really known and somewhat taboo (Kessler, 2011). Now online dating is part of normal society with many couples meeting online, including my husband and me. The main problem was how to match individuals to potential mates. In 2003 a press release stated that Match.com planned to find matches using math to predict compatibility based on the results gathered from their personality test which include things like creativity, social style, and mood (Match.com Personality Matching Launches a First-of-its-Kind Product Innovation that Pairs up Singles using Ph.D. - Designed Compatibility Test, 2003).

**Data Understanding**

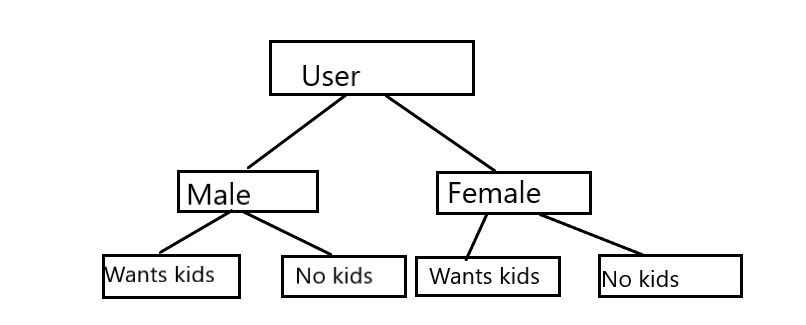
From my research and my own experience in online dating the first thing done by the user is fill out a brief survey that includes basic facts, preferences, and what you are looking for in a mate. From there you will start receiving suggested matches based on the initial data you provided. I also know if you add a picture that will drive more “views” from other users on the site. The personality test mentioned in the press release will actually add more nuanced data on top of the basic information. This means the more data I give the more specific my matches may get. However, since you are able to filter and search on your own as well you need to layer in “dissonance” where a user may say one thing and then do something different (Gelles, 2011).

**Data Preparation**

For data preparation I would need as many personality tests as possible to start training the model. As each test becomes somewhat individualized, I also think it would be important to start creating categories that would contain similarities. For instance, singles that want children would be one, while physical traits could make up several ranging from height to hair color. I also think this would be a continuous process as new users sign up daily, while old ones who have found their match may leave.

**Modeling**

Since the amount of data will be quite large, I think I will take the same approach the YMCA case study did and split the data 50/50 with half going to training and half going to testing. I hope by doing this I will get a better idea of how dissonance will play into the results. I think for modeling the data I would use decision trees with the user account being my primary key. I would then work through the basic qualifications as my leaves.



I would want to pay special attention to things that users would consider “deal breakers” or the things they will not compromise on with a partner. However, dissonance can play a role here. For example, a woman might say a man not wanting kids is a deal breaker, but after a few weeks may start reaching out to men who state they do not want kids. This model would have to be constantly updated taking these changes in mind. I would train my data on the train set and then validate it against the test set if I found my decision trees were working matching individuals based on their basic wants/needs as well as what their surveys introduce.

**Evaluation**

One of the first things I will need to evaluate with my data is overfitting. Since I plan to model with a decision tree approach, I will need to continuously look at the leaves and make sure they are not getting too specific. I will need to find a middle-road approach where enough qualifications match but have some wiggle room when it comes to other qualifications. Here I would focus on deal breaker decisions first and then layer in some other pieces based on importance. What I would use as my success feature is if the model accurately predicts when a match was made ending in some form of communication between the two parties.

**Deployment**

For deployment I would try a phased approach since I am a paranoid person by nature. I would start off with a few users and verify that what we saw in training and testing held up against the ongoing receipt of data. If the sample size performs well enough which means enough match suggestions ended in some communication, then I would deploy to the whole population.

# References

Gelles, D. (2011, July 30). *Inside Match.com*. Retrieved from Slate: https://slate.com/human-interest/2011/07/inside-match-com-it-s-all-about-the-algorithm.html

Kessler, S. (2011, July 13). *The Love Equation: How Match.com Calculates Your Ideal Mate*. Retrieved from Mashable: https://mashable.com/2011/07/13/match-com-equation/

Match.com Personality Matching Launches a First-of-its-Kind Product Innovation that Pairs up Singles using Ph.D. - Designed Compatibility Test. (2003, December 17). IAC.

Siegel, E. (2016). *Predictive Analytics: The Power to Predict who will Click, Buy, Lie, or Die.* Hoboken: Wiley.