

Top Predictors of Marriage Success

Roland Bennett

12/30/2019

R Introduction

In this project survey data is used to identify indicators of a successful marriage. The data used a survey of 170 Turkish couples who were asked 54 questions about their relationship. Of the 170 couples, 84 were divorced and 86 were married with no plans of becoming divorced. This data can be found at <https://archive.ics.uci.edu/ml/datasets/Divorce+Predictors+data+set#>. The questions were about behavior, opinion, and personality, such as: most of our goals are common to my spouse.

Developing an algorithm to predict whether or not a couple would get divorced based on their answers to the survey turned out to be rather simple and 100% accuracy was achieved. The primary goal was to determine which of the questions are most useful in determining whether or not a couple will get divorced. The four questions that were identified as being most useful were then used as predictors in a new algorithm.

Methods

Each respondent answered each question with a number from zero to four, with 0 being strongly agree and 4 strongly disagree. The questions asked are as follows:

1. If one of us apologizes when our discussion deteriorates, the discussion ends.
2. I know we can ignore our differences, even if things get hard sometimes.
3. When we need it, we can take our discussions with my spouse from the beginning and correct it.
4. When I discuss with my spouse, to contact him will eventually work.
5. The time I spent with my wife is special for us.
6. We don't have time at home as partners.
7. We are like two strangers who share the same environment at home rather than family.
8. I enjoy our holidays with my wife.
9. I enjoy traveling with my wife.
10. Most of our goals are common to my spouse.
11. I think that one day in the future, when I look back, I see that my spouse and I have been in harmony with each other.
12. My spouse and I have similar values in terms of personal freedom.
13. My spouse and I have similar sense of entertainment.
14. Most of our goals for people (children, friends, etc.) are the same.
15. Our dreams with my spouse are similar and harmonious.
16. We're compatible with my spouse about what love should be.
17. We share the same views about being happy in our life with my spouse
18. My spouse and I have similar ideas about how marriage should be
19. My spouse and I have similar ideas about how roles should be in marriage
20. My spouse and I have similar values in trust.
21. I know exactly what my wife likes.
22. I know how my spouse wants to be taken care of when she/he sick.
23. I know my spouse's favorite food.
24. I can tell you what kind of stress my spouse is facing in her/his life.
25. I have knowledge of my spouse's inner world.
26. I know my spouse's basic anxieties.
27. I know what my spouse's current sources of stress are.

28. I know my spouse's hopes and wishes.
29. I know my spouse very well.
30. I know my spouse's friends and their social relationships.
31. I feel aggressive when I argue with my spouse.
32. When discussing with my spouse, I usually use expressions such as 'you always' or 'you never'.
33. I can use negative statements about my spouse's personality during our discussions.
34. I can use offensive expressions during our discussions.
35. I can insult my spouse during our discussions.
36. I can be humiliating when we discussions.
37. My discussion with my spouse is not calm.
38. I hate my spouse's way of open a subject.
39. Our discussions often occur suddenly.
40. We're just starting a discussion before I know what's going on.
41. When I talk to my spouse about something, my calm suddenly breaks.
42. When I argue with my spouse, I only go out and I don't say a word.
43. I mostly stay silent to calm the environment a little bit.
44. Sometimes I think it's good for me to leave home for a while.
45. I'd rather stay silent than discuss with my spouse.
46. Even if I'm right in the discussion, I stay silent to hurt my spouse.
47. When I discuss with my spouse, I stay silent because I am afraid of not being able to control my anger.
48. I feel right in our discussions.
49. I have nothing to do with what I've been accused of.
50. I'm not actually the one who's guilty about what I'm accused of.
51. I'm not the one who's wrong about problems at home.
52. I wouldn't hesitate to tell my spouse about her/his inadequacy.
53. When I discuss, I remind my spouse of her/his inadequacy.
54. I'm not afraid to tell my spouse about her/his incompetence.

Initially, the data is divided into a training set with 80% of the data and a test set with 20% of the data. An ensemble was used to predict whether or not the couples in the test set were married with every single question used as a predictor. The ensemble consisted of a generalized linear model, linear discriminant analysis, naive bayes, support vector machines, k-nearest neighbour, gam logical regression, random forests, and adaboost. Each algorithm yields a prediction of either 1 or 0 and the final prediction is the mode of the ensemble. Using the answers to every question as a predictor, the ensemble was able to correctly predict whether or not every couple in the test set was married or divorced.

In order to further test the precision of this ensemble, a Monte Carlo simulation was employed. 136 (80%) of the respondents were selected at random to use as the test set, and this was run 100 times. In order to save time, the adaboost algorithm was not employed.

In order to determine which questions were the most accurate in predicting marriage success, each question was used individually to predict marriage success. For each number zero through four, I predicted divorced (1) if the response was greater than or equal to that number and married (0) if less than the number. Then the maximum accuracy out of these five values was assigned as the accuracy for that predictor. The predictors with the four highest accuracies were then used in an ensemble model on their own.

Results

The initial ensemble model had a perfect 100% accuracy. When run as a Monte Carlo Simulation, the mean accuracy was 97.69% and 427 out of 1000 of the simulations had perfect accuracy. The standard deviation was 0.02 The most useful questions in predicting marriage success were:

18. My spouse and I have similar ideas about how marriage should be
11. I think that one day in the future, when I look back, I see that my spouse and I have been in harmony

with each other

17. We share the same views about being happy in our life with my spouse

19. My spouse and I have similar ideas about how roles should be in marriage

They have accuracies of 98.24%, 97.65%, 97.65%, and 97.65%, respectively. These accuracies were all found when divorced (1) was predicted for responses of value 3 or higher.

When these four predictors were used as the only predictors in an ensemble machine learning model, the model still had perfect accuracy. In order to improve the veracity of this, another Monte Carlo simulation was run. A model using these four predictors was compared against a model using four random predictors. The mean accuracy when Questions 11, 17, 18, and 19 were used is 97.71% and the mean accuracy when four random predictors were used is 93.88%. The standard deviations were 0.02 and 0.04, respectively. When only the top four predictors were used, the accuracy was slightly higher than when all predictors were used. However, the difference was smaller than the standard deviation.

Conclusion

In the final Monte Carlo Simulation, the chosen predictors had a considerably higher mean and lower standard deviation than the random predictors. This supports my conclusion that they do in fact have greater predictive power than other questions. Curiously, the mean accuracy of the Monte Carlo simulation for the chosen predictors was not significantly higher than just guessing divorced when the response had a value of 3 or greater.

This survey is a strong method of predicting marriage success. However, there are multiple potential sources of error. For one, the length of the survey could cause the taker to lose interest and not think as carefully about the answers at the end. Honesty is also an issue when giving surveys to humans. Another is that it was Turkish people who were surveyed, so cultural differences between Turks and Americans could make it difficult to apply to American couples, which is how I would apply this knowledge since I am an American. The sample size is also relatively small: only 170 respondents. The monte carlo simulations all have relatively high standard deviations.

In order to improve this experiment, a larger sample size should be used. Also, additional, more objective questions covering topics such as relative age, income, socioeconomic and cultural background, diet, and consumption of alcohol and tobacco products could lead to more insight.