**Divorce Prediction: Man vs. Machine**

**Objective:**

If you have read the book Blink by Malcolm Gladwell, then you might be familiar with Dr. John Gottman who is renowned in the field of psychology for being able to predict, with a high percentage of accuracy, couples that will divorce. By simply meeting with a couple and observing their interaction Dr. Gottman can identify negative communication patterns that reveal feelings of criticism, contempt, defensiveness, and stonewalling. So, is there a way that we can use data science to identify who will divorce with more accuracy than Dr. John Gottman? Dr. Gottman boasts an accuracy rate of over 94%. The goal of my case study is to learn if machine learning can be used to predict who will divorce, with higher accuracy than the human expert.

**Dataset:**

The dataset being used was retrieved from UCI Machine Learning Repository.

With Dr. Gottman’s research in mind a group of other researchers created a dataset using a list of 54 statements that were meant to measure the respondent’s feelings toward their spouse or ex-spouse and the relationship. The statements were evaluated during an interview of 170 respondents containing both married and divorced participants and measured on a Likert style scale. 0=Never, 1=Seldom, 2=Averagely, 3=Frequently, 4=Always. One thing to keep in mind is that the researchers and respondents were from Turkey, so there are some places where the translation does not make grammatical sense in English, but you will be able to understand the gist of the statements.

Attributes:

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, ı 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.

**EDA Graphical Analysis:**

The first visualization I choose to do was to plot the mean scores for all attributes by Class.

Chart, line chart

Description automatically generated

We can see that the general trend is that married persons scored statements in the higher ranges, with the exception of statements 6 and 7. Upon reviewing those statements, they do appear as negative statements, were most of the rest of the statement are more positive. But it’s important to note that all statements are not created equal. 0 is not always equal to bad and 4 is not always equal to good. We also see that statements 43, 46, and 48 scored the highest among the divorced persons.

The next visualization I choose was a correlation heatmap.

A picture containing text

Description automatically generated

As you can see, due to the sheer number of features this correlation heatmap is not very valuable. When looking at the correlation matrix, I can see that there are some features that have higher correlation scores, like Atr 9 and Atr 15 at ~0.95. This information may come in handy when conducting feature selection, but for exploratory data analysis I was more interested if there were any features that correlated with Class.

That’s why I choose a visualization to display features correlated with class next.

Background pattern

Description automatically generated

This visualization at least gives me a place to start. There are a number of variables that score pretty high, but for the next visualization I’m going to take a look at just the top five features.

Chart, bar chart

Description automatically generated

At least with the top five features that are correlated with Class, we can see that there is a clear divide and very little overlap between divorced and married counts of answers on statements. This may be why they were identified as correlated.

This information has given me a place to start, some ideas to ponder, but it’s really going to come down to feature selection to create the best model to predict who will get divorced.

**References:**

Lisitsa, E. (2013, February 18). The Research: Predicting Divorce from an Oral History Interview. The Gottman Institute. Retrieved from <https://www.gottman.com/blog/the-research-predicting-divorce-from-an-oral-history-interview/>.

UCI Machine Learning Repository. (2019, July 24). Divorce Predictors data set Data Set. Center for Machine Learning and Intelligent Systems. Retrieved from <https://archive.ics.uci.edu/ml/datasets/Divorce+Predictors+data+set>.

Yöntem, M., Adem, K., İlhan, T., Kılıçarslan, S. (2019). DIVORCE PREDICTION USING CORRELATION BASED FEATURE SELECTION AND ARTIFICIAL NEURAL NETWORKS. Nevşehir Hacı Bektaş Veli University SBE Dergisi, 9 (1), 259-273. Retrieved from <https://dergipark.org.tr/tr/download/article-file/748448>.