Predicting Marriage Outcomes Using Couples Therapy Questions

Name: Zachary Stiles Email: zmstiles@crimson.ua.edu

*Abstract*—This electronic document is a “live” template and already defines the components of your paper [title, text, heads, etc.] in its style sheet. *\*CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract*.

Keywords—component, formatting, style, styling, insert

# Introduction

This section describes the background and significance of the research questions.

# Research Question

My research question is “Can the outcome of a marriage be predicted by a spouse’s opinions of the state of their marriage?” In other words, “Can we use a spouse’s current feelings and views about his or her marriage to predict whether or not that marriage will end in a divorce?” This question is a simple binary classification problem, with our two classes being “will divorce” and “will not divorce.”

This question is obviously not to be taken seriously. The outcomes of any models created for this project are not intended to diagnose the state of a marriage or suggest a married couple file for divorce.

# Related Work

This same topic was discussed by Dr. Yöntem et al. (henceforth, “Researchers”) on the same dataset used to answer my research question. Researchers used three different models to evaluate varying numbers of features in their research. Researchers received their features from the Gottman Insitute, a Turkish-owned couples therapy business. Gottman couples therapy contains a 54 question survey known as the Divorce Predictors Scale (DPS). Each question served as an initial feature in their research. They obtained the data via in-person interviews and Google Drive survey submissions from married and divorced couples. This is the same dataset that I used for my own analysis (discussed more in IV.).

After feature acquisition, Researchers used a method of correlation-based feature selection to reduce the total features down from 54 to six, selecting only the features with the highest correlation to a couple’s divorce: Q2: “I know we can ignore our differences even if things get though sometimes.”; Q6: “We don't have a common time we spent together at home.”; Q11: “We don't have a common time we spent together at home.”; Q18: “We have similar ideas with my spouse about how a marriage should be.”; Q26: “I know the basic concerns of my spouse.”; Q40: “We're starting to fight before I know what's going on.”

Researchers ran both sets of 54 and six features on three different machine learning models: artificial neural networks (ANN), random forest, and radial basiss function (RBF) classification. For the 54 feature set, their models predicted with similar degrees of accuracy: ANN: 97.64%; RBF: 98.23%; random forest: 97.64%. Interestingly, the six feature set scored better, and one of them scored worse: ANN: 98.82%; RBF: 97.64%; random forest: 97.64% [1].

While this research was on the same dataset, I hoped to provide a more refined approach to selecting features and use additional models to further explore the data and take of note of any differences in performance of the models used in this project.

# Methods

## Data Acquisition

As discussed in III., the data used for this project was previously collected through in-person interviews and Google Drive surveys for the research discussed above. The dataset was obtained on Kaggle from [2], sourced from a different party than Researchers above. The file for the data was obtained in a

The data came in the form of a .csv file with semicolons as delimiters, containing 55 columns and 170 rows. The first 54 of these columns represented answers to the 54 DPS questions, respectively, ranging from integers zero (0) to four (4). These numbers represent a linear scale of how much a couple or spouse agrees or disagrees with the question or statement: zero (0) meaning “strongly disagree” or “never”, one (1) meaning “disagree” or “seldom”, two (2) meaning “neutral” or “averagely”, three (3) meaning “agree” or “frequently”, and four (4) meaning “strongly agree” or “always”. The last column contained a binary value of zero (0) or one (1) to represent whether a couple was divorced or not divorced, respectively. Each row of the data contained the entire set of answers and marital status of a single couple [2].

## Data Cleaning and Labeling

First it was important to ensure the data contained no null values. I decided to remove any rows that contained a null or missing value from the data. Fortunately, every row was fully populated.

One of the best qualities of the dataset is the zero (0) to four (4) scale that every value has. Normally on a numerical dataset, it would need to be standardized or normalized so that the varying ranges of numbers do not interfere with the feature engineering process. However, since all values are on the same scale, there is no need to standardize or normalize the data.

Another benefit of the dataset is the “Divorce” column. Because I am trying to predict whether a not a couple will get a divorce, this column serves as the target variable for the machine learning models. However, I found it confusing that zero (0) was used to indicate “divorced” and one (1) was used to indicate “not divorced.” Following the tradition of zero (0) representing false and one (1) representing true in binary, I decided to switch the values in the dataset to make it less confusing, which is the extent of the data labeling done in this project. Lastly, I decided to create a count plot to display how balanced the target variable is for the dataset [3]. Fortunately, there was a very even balance of “not divorced” to “divorced” couples in the dataset, which will greatly improve the reliability of the models to be developed.

Chart, bar chart

Description automatically generated

Fig.1. A count plot of the target variable.

## Feature Engineering

I found the research in III. to be overly simplistic for the feature engineering process, so I wanted to expand on the subject. However, I believe looking at correlation values was a great place to start. I decided to create a heatmap of the absolute magnitude of the correlation values for every feature in the dataset [4]. Furthermore, I wanted to use the Predictive Power Score (PPS) as a metric for feature selection. While correlation is fine for summarizing linear relationships between variables, it completely neglects nonlinear relationships. Furthermore, a correlation plot or heatmap is symmetric across the diagonal, but the PPS of one variable to another is completely different from the PPS of the second variable to the first [5]. Thusly, I created a heatmap of the PPS as well. The main rows to look at for the first step of the feature engineering process are the “Divorce” rows in each heatmap. Lighter colored intersections with these rows represent higher correlation or PPS with the variable.

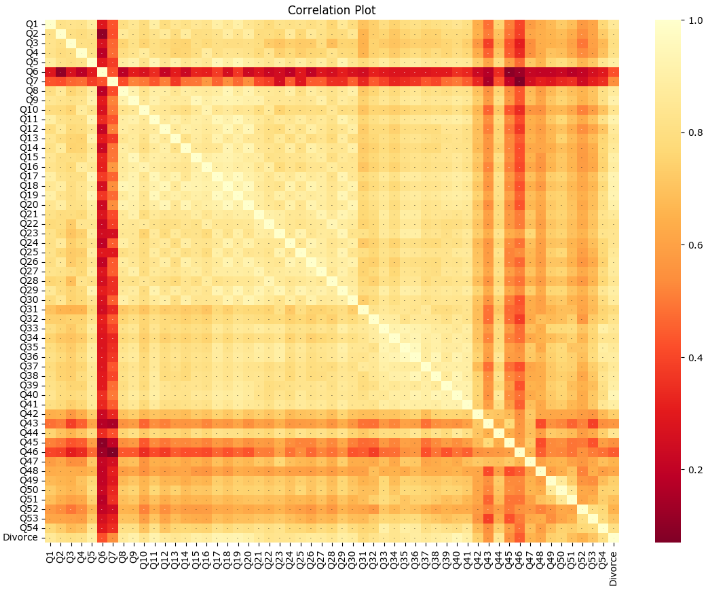


Fig. 2. The correlation heatmap between all 54 features and target variable “Divorce”.

**A picture containing text

Description automatically generated**

Fig. 3. The Predictive Power Score heatmap between all 54 features and target variable “Divorce”.

Looking at the heatmaps, it is interesting to note that there is a very high correlation between most features. Also, the row containing the highest values on the PPS heatmap is the “Divorce” row, which is promising for model performance.

I started by selecting features using PPS in reference to the target variable. Because the dataset is so small, I needed to drastically reduce the number of features to prevent overfitting of models [6]. I set a minimum requirement of 0.7 PPS, which immediately eliminated half of the features. With this subset, I took after the original researchers of the project and took only the features with the highest correlation values, setting a minimum of 0.8 for the correlation with the “Divorce” variable.

After this first stage of feature selection, there was a subset of 9 values remaining. While the original translation from Turkish to English was poor, I referenced the Gottman Institute’s website to find the appropriate translation: Q5: “The time I spend with my spouse is special for the both of us.”; Q9: “I look forward to vacations with my spouse and enjoy the travel we do together.”; Q17: “My spouse and I have similar ideas about how we find happiness in life.”; Q18: “My spouse and I have similar ideas about how marriage should be.”; Q19: “My wife and I have similar ideas about how roles should be in marriage.”; Q20: “My wife and I have similar values in trust.”; Q29: “I know my spouse very well.”; Q36: “I am humble in discussions with my spouse.”; Q40: “I know why my partner is upset before we have an argument.” [7]. It is important to note that this changed the meaning of questions 36 and 40. Question 40 was used in the original research of this dataset, but notably had a translation that no longer exists on the Gottman couples therapy’s questionnaire [1].

Even after greatly reducing the number of features, I wanted to reevaluate the correlation plot once again. The heatmap displayed very high correlation values between most of the features, so for good measure I created new heatmaps for both correlation and PPS and take a closer look.

Chart, treemap chart

Description automatically generated

Fig. 4. The correlation heatmap between the reduced set of nine features and target variable “Divorce”.

**Chart, treemap chart

Description automatically generated**

Fig. 5. The correlation heatmap between the reduced set of nine features and target variable “Divorce”.

With the reduced feature set, the correlation and PPS values could be displayed on the heatmaps. However, the correlation between many of the features is very high. The first goal of this feature engineering process was to select features that maximize correlation and predictive power, but the second is to reduce redundancy. Having variables with too high of correlation can cause models to overcompensate for these features and skew prediction results [8] [9].

Thusly, I decided to remove “duplicates” of features having correlation values over 0.95 with each other. While this seems like a very high value, I think a degree of leniency is required when evaluating these particular features, as they are based on psychological answers to questions about personal relationships; I do not want to discount the relevancy of a spouse’s opinion of their marriage. However, this small filter eliminated questions 19 and 20 from the feature set, as they overlapped with questions 17 and 18, respectively.

This reduced the total number of features for predicting the target variable to seven. Interestingly, the only features in my reduced set that overlap with the original researchers’ feature set were questions 18 and 40.

## Model Development

I decided to develop three models for this research using the methods of linear regression, decision tree classification, and randomized forest classification.

The first method I used was logistic regression, a classic classification method despite being called a “regression” technique. While linear regression is also fine for classification problems, predicting whether a couple will get divorced is not fully black and white; there may be a chance closer to 50 percent that the couple will indeed divorce. Logistic regression will benefit the deployment of this model as the probabilities of each category, “not divorced” and “divorced” can be displayed alongside whatever classification the model predicts [10]. I decided to maintain the default values used in scikit-learn’s logistic regression model. Most notable of these parameters is the default 100 maximum number of iterations permitted to converge [11].

I decided to develop a decision tree classification model as the second model, largely because I wanted to compare its performance to a randomized forest classification’s performance. For the decision tree classifier, I used Gini impurity for the classification criteria, which calculates the probability of for incorrectly classifying features [12]. I did not set a maximum depth of the decision trees, as I found setting values lower than six worsened the prediction performance of the model [13].

Qr code

Description automatically generated

Fig. 6. A decision tree used on the training dataset assigned at random state 13.

The last model I created was a randomized forest classification. Because random forest models take several decision tree classifiers on subsets of the data and then averages results to increase accuracy, I created the third model to compare its performance to that of the second model. I also was interested in comparing this model’s performance to that of the randomized forest model used for this dataset’s original research project. I decided to maintain use of the Gini impurity criteria and not set a maximum tree depth; the same settings used for the decision tree classification model [14].

## Model Training

While model training, I opted to split the dataset to allow for 75 percent training set size. This resulted in a training size of 127, leaving the remaining 25 percent for a test size of 43. Initially, I divide the dataset for a 70 percent training size, but upon increasing it to 75 percent saw an increase in prediction accuracy. Further increasing it to 80 percent resulted in negligible change. All models used the same training and test datasets, which were split at a random state of 27, although changing this value did not affect performance. As aforementioned, most hyperparameters were left in default values and the decision tree classification and randomized forest classification models used Gini impurity.

## Model Evaluation and Validation

To evaluate the three models, I used two main techniques: accuracy scoring and a confusion matrix. While essentially these display the same values in a binary classification problem, the confusion matrix helps achieve the data visualization aspect of data science.

## Model Deployment

Please describe how you deploy the models to web.

# Results

Please show your results as below.

## Mode Performance in Evaluation

#### Please show the prediction performance of the models during evaluation and validation.

## Feature Importance

Please show how the features contribute to the model performance.

## Runtime Monitoring

Find other 2 classmates and let them use your web-based models. Each of them should create a testing data sample to test your web-based model performance. Please describe your web-based model’s testing performance here.

# Discussion

## Answers to Research Questions

Did your results answer the research questions. If yes, how much is being answered? If no, why?

## Limitations of your research.

Please explain any limitations of your research

## Lessons Learned

Please describe what kind of lessons you learned from the final project.

##### Acknowledgment

Please acknowledge your classmates and any others helped you during the research process..

##### References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

1. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. K. Elissa, “Title of paper if known,” unpublished.
5. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
6. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
7. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.**