

ISyE 601 Project Proposal: Divorce Prediction and Explanation

Akhilesh Soni (9079896453), Zach Zhou (9071567169)

Divorce is devastating—it is draining both emotionally and mentally, and can also become a massive financial burden. But as much as couples would like to avoid a divorce, it seems all too prevalent in society. According to the American Psychological Association, 40 to 50 percent of married couples in the United States divorce, with the divorce rate for subsequent marriages being even higher. If divorce likelihood of married couples can be estimated early, many divorces can be prevented.

Gottman (1992) found that he was able to predict which couples would eventually divorce with 94% accuracy. Since then, Gottman’s research has been to understand the factors that lead to divorce. More recently, Yöntem and İlhan (2018) developed the Divorce Predictors Scale (DPS) based on Gottman’s work in couples therapy. The DPS is a behavior rating scale (BRS), i.e., a questionnaire with responses coming from a 5-point scale (0=Never, 4=Always) to questions such as “I enjoy traveling with my wife” and “Our discussions often occur suddenly.” By analyzing DPS responses, researchers could hypothetically predict whether someone will ultimately divorce one’s partner. Conversely, it may be discovered that certain questions are more predictive of divorce than others.

Our goals are twofold. We aim to create a machine learning model which can predict whether a couple will divorce or not based on responses to the DPS. We would also like to identify the most telling questions, in an attempt to explain what factors are key to understanding the likelihood of divorce.

We will use the **Divorce Predictors dataset**. The dataset consists of 170 couples’ responses to the DPS; 84 are divorced and 86 are married. The features correspond exactly to the 54 questions appearing on the survey; each feature is a response in the form of a discrete rating on scale of 0–4. Note that features can certainly be correlated. For instance, a response to the question “I feel right in our discussions” is unlikely to be independent from a response to the question “I wouldn’t hesitate to tell my spouse about her/his inadequacy.” The target is a binary indicator of whether the couple is divorced.

Our first task is to develop an effective classifier for predicting whether a couple is divorced based on responses to the DPS. Classification models we will apply include logistic regression, decision trees, random forest, and support vector machine.

Our second task is to identify the most important features in the dataset i.e. questions which play the most crucial rule in determining the likelihood of a divorce. Here we approach the problem as an explanation problem instead of a prediction problem. We plan to use correlation statistics and selection methods to identify most significant features and train our model on this subset of features. We plan to study performance of the model with reduced features against the model inclusive of all the features.