The Matchmaker’s Dilemma:
Predicting Unsatisfying Marriages

by

Casey La Honta

Table of Contents

1. Introduction

Marriage. Marriage is what brings us together. Indeed, the United Nations Department of Economic and Social Affairs reports that over 90% of adults will be married at some point in their lives. Married couples and their family units often form the building blocks of societies. Marriage is also correlated with more positive outcomes for each of the marriage partners, their children, and their communities.

To this end, many matchmaking services, dating apps, and other products exist to create the types of relationships that lead people to lasting marriages. It's no easy task to initiate a lasting relationship. Divorce rates are also high, especially in the Western world, which raises the stakes for matchmakers. This project uses machine learning to build a tool for explaining a piece of that puzzle.

The intended use of this model would be for matchmakers to screen clients that will be difficult to find a long-term satisfying match for. This can help these individuals and companies maintain strong reputations based on the clients they do take on, rather than having their reputations stained by divorce.

2. Dataset Information

The data set for this project comes from a Marital Satisfaction survey (2017) given to married couples in 33 different countries and contains 7178 responses. The participants were 3827 women and 3351 men, ranging in age from 17 to 88 years old. The average age of participants was 40.7 years old. The participants' marriages ranged from less than one year to 70 years, with an average marriage duration of 14.8 years.

The questionnaire includes four sections to be completed by the participants. First, a *Personal information* battery collected basic data on demographic characteristics of the participants. The next scale was taken from the Marriage and Relationships Questionnaire (MRQ), nine items where participants rated themselves on a 1 (No) to 5 (Yes) scale. Then, the participants completed the three-question Kansas Marital Satisfaction Scale (KMSS), where participants rated their satisfaction on a 1 (very dissatisfied) to 7 (very satisfied) scale. Finally, participants completed a subset of questions from the GLOBE survey that specifically addressed familial topics. Once again, participants rated themselves on a 1 (strongly agree) to 7 (strongly disagree) scale.

3. Data Wrangling

This dataset required some processing in order to prepare it for analysis. Notably, only 86 null values existed in the entire dataset. The 86 null responses were all from one country (Uganda) on one survey item (`religion`). I replaced these null responses with “Did not answer” to denote the missing values.

The three rating scales also needed to be cleaned up, starting with the GLOBE survey. This survey was the only one where a response of “1,” “2,” or “3” denoted a positive or affirmative rating. In order to demonstrate consistency with the other scales, I made the decision to reverse-score this survey. Our scale for analysis then became 1 (strongly disagree) to 7 (strongly agree).

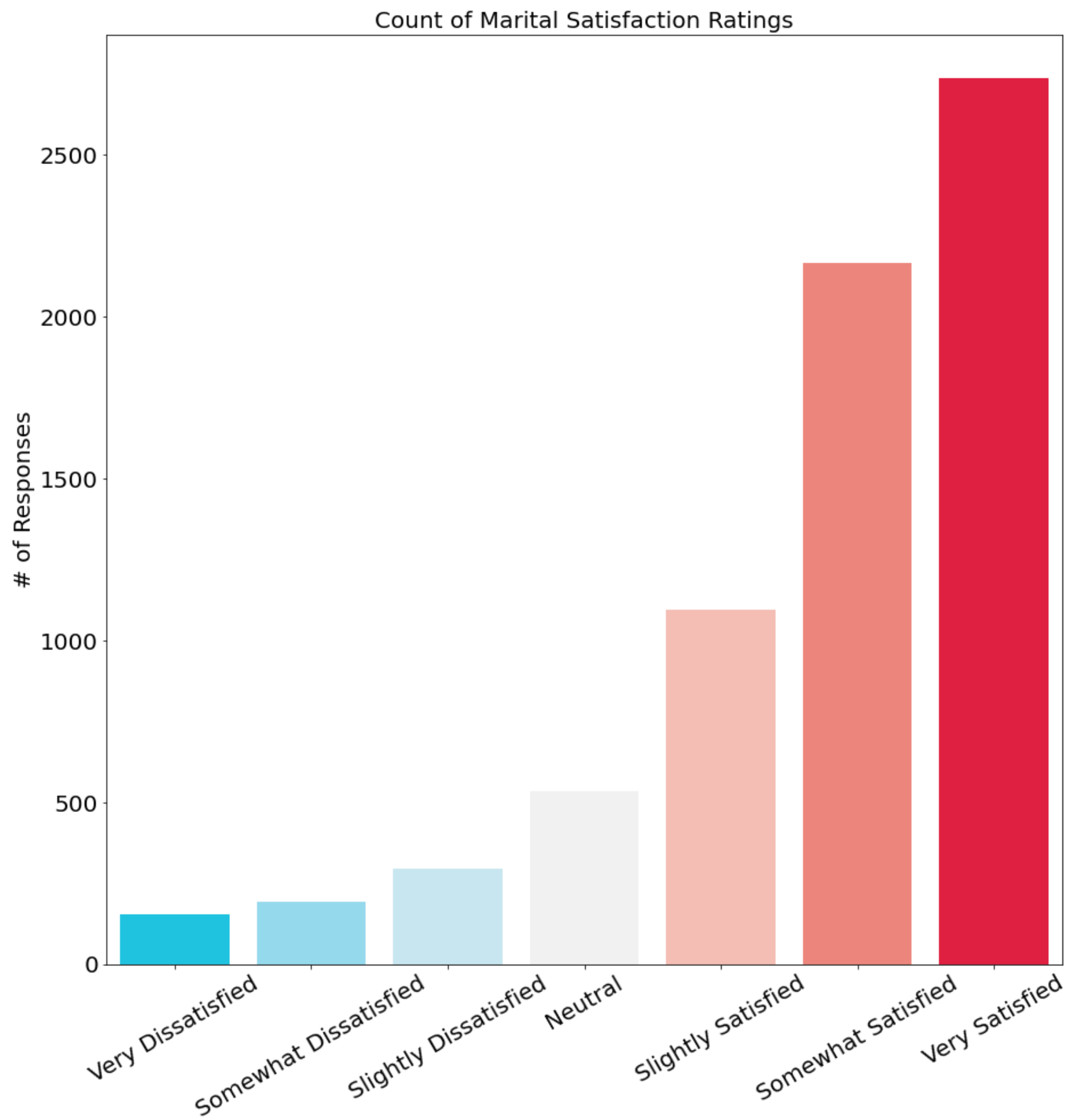
The three scales were then rescaled to make the neutral response equal to zero. This scoring assists interpretability of data, where a positive value indicates a positive or affirmative rating, and a negative value indicates the opposite. The three scales then became:

MRQ: -2 (No) to 2 (Yes) (0 = Neither Yes nor No)
KMSS: -3 (very dissatisfied) to 3 (very satisfied) (0 = neutral)
GLOBE: -3 (strongly disagree) to 3 (strongly agree)

Various numeric data (e.g. `religion` or `education level`) were also converted to categorical data for the purposes of the Exploratory Data Analysis. These would be one-hot encoded later on for Model Selection. The missing religion values were the only ones that needed imputing. These were dealt with by replacing the null data with the string “Did not answer” (*n* = 86). The final step was to set the data types to integers rather than objects, to allow the project to utilize scikit-learn’s many statistical techniques. The final data set kept all 7178 responses.

4. Exploratory Data Analysis

For this project, our target variable for prediction will be the self-reported `marital_satisfaction` from the KMSS. Here, participants rated themselves from 1 (“Very Dissatisfied”) to 7 (“Very Satisfied”):



Well, this is encouraging! In other words, it’s good to see that the surveyed population enjoys some satisfying marriages. For this analysis, however, the extreme skew of the responses is likely to lead to some difficulty in deriving an effective model.

This regression heatmap identifies a number of interesting things. First, the KMSS items (which include the target variable) vary strongly with each other. Therefore the alternate items (`spouse_satisfaction` and `relationship_satisfaction`) will be dropped for the model building. The MRQ features relate the highest with our target, so I would expect to use many of these features during modeling. The GLOBE survey, however, shows the weakest coefficients among the rating questionnaires.

5. Pre-processing

When preparing the data for model selection, I first one-hot encoded the two nominal-scale data sets, `religion` and `country`. This process provides a numerical measure of each of the nominal-scale data points vs the absence of that data point for a given observation. For instance, whether or not someone is Catholic is represented by a 0 (not Catholic) or 1 (Catholic). To properly calibrate the new features, I dropped `rel_Catholic` and `cnty_Croatia`, which would now be represented by a row of zeros among the other features from the nominal scale. To continue with the previous example, a response from a Catholic person would now be represented by a 0 value in each new `religion` feature.

Next, I performed the variance inflation factor test from statsmodels to examine the data for examples of high collinearity or outlier influence. This process identified three features for removal (`rel_Muslim`, `cnty_India`, and `marriage_duration_years`) that correlated so strongly with other features that they became redundant.

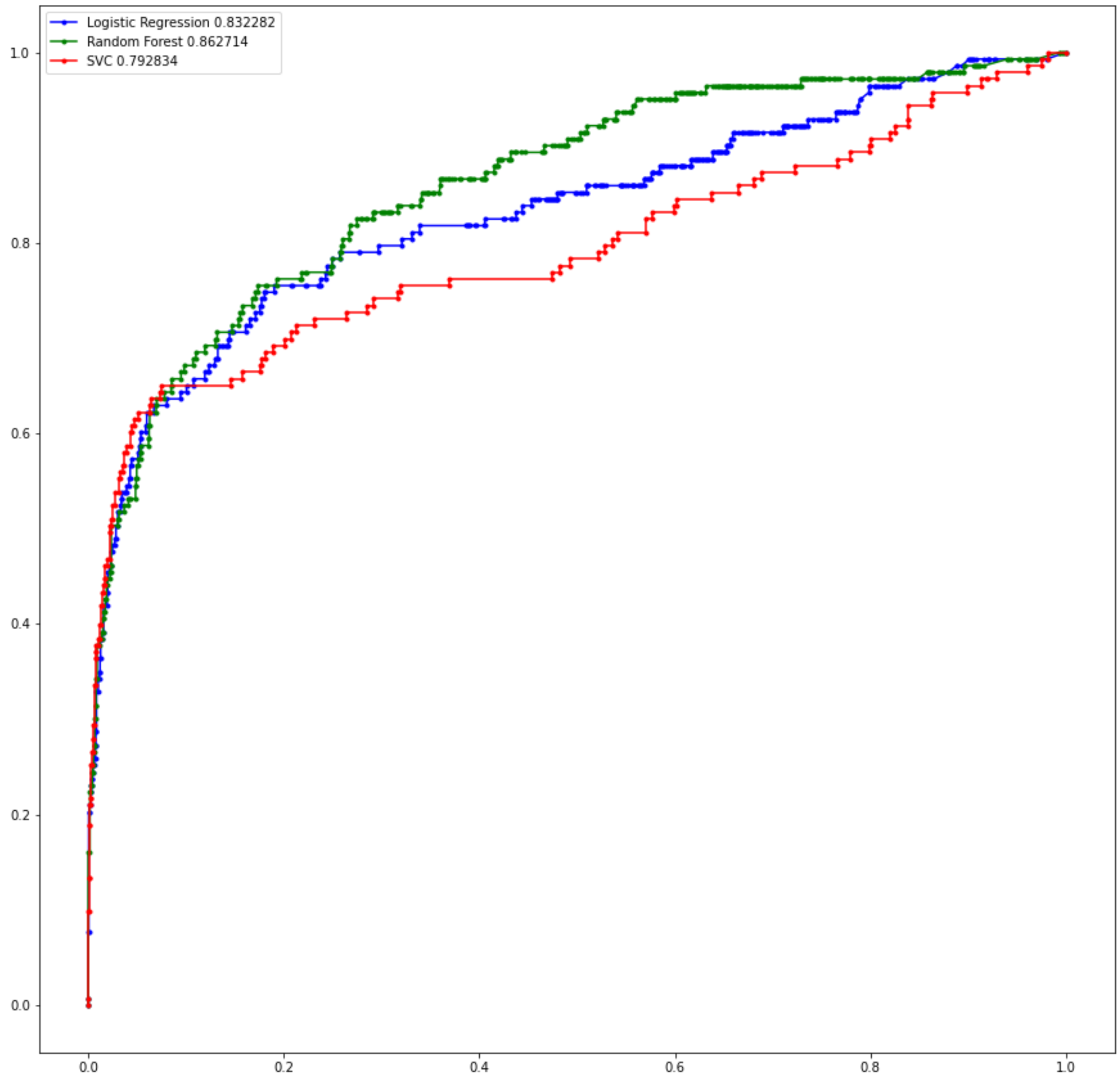
By the time I started the feature engineering and pre-processing, I was met with a new challenge. Responses to the target variable, `marital_satisfaction`, are skewed towards the “satisfied” end of the spectrum enough that it isn’t fun to predict. Fortunately, it appears that people are capable of finding satisfaction in their marriage very well on their own! For this project, however, it is not particularly interesting in terms of predictive analytics and algorithm-building, so I made the decision to reverse the prediction and focus on the *dis*satisfied participants instead. For our new target variable, `marital_satisfaction` was coded into the following:

0 = *neutral* and *satisfied* responses (*n* = 6533)
1 = *unsatisfied* responses (*n* = 645)

Additionally, coding the target into a binary variable provides us with more model-building opportunities. With the appropriate features selected and the target properly coded, I began the model selection. For each of the classification algorithms, I performed sklearn’s GridSearchCV to determine the best hyperparameters. I fit and classified using Logistic Regression, Random Forest, and a Support Vector Classifier, all from sklearn’s libraries. I used AUC score to compare the models, as it is threshold independent and helps us understand how well the models perform. The results and ROC curves are shown below:

Algorithm	AUC score	Parameters
Logistic Regression	.832	C = 0.01, l1_ratio = 0.5, max_iter = 5000, penalty = 'elasticnet', solver = 'saga'

Random Forest	.862	max_depth = 2, n_estimators = 75
SVC	.792	C = .001, probability = True



Random Forest clearly won out during model selection, with `spouse_pride`, `spouse_attraction`, `enjoy_spouse_company`, `spouse_love`, and `happiness` being the most important features. The fact that these features rose to the top of the importance ranking, however, brings us to an issue that can be best explained through the lens of our business case.

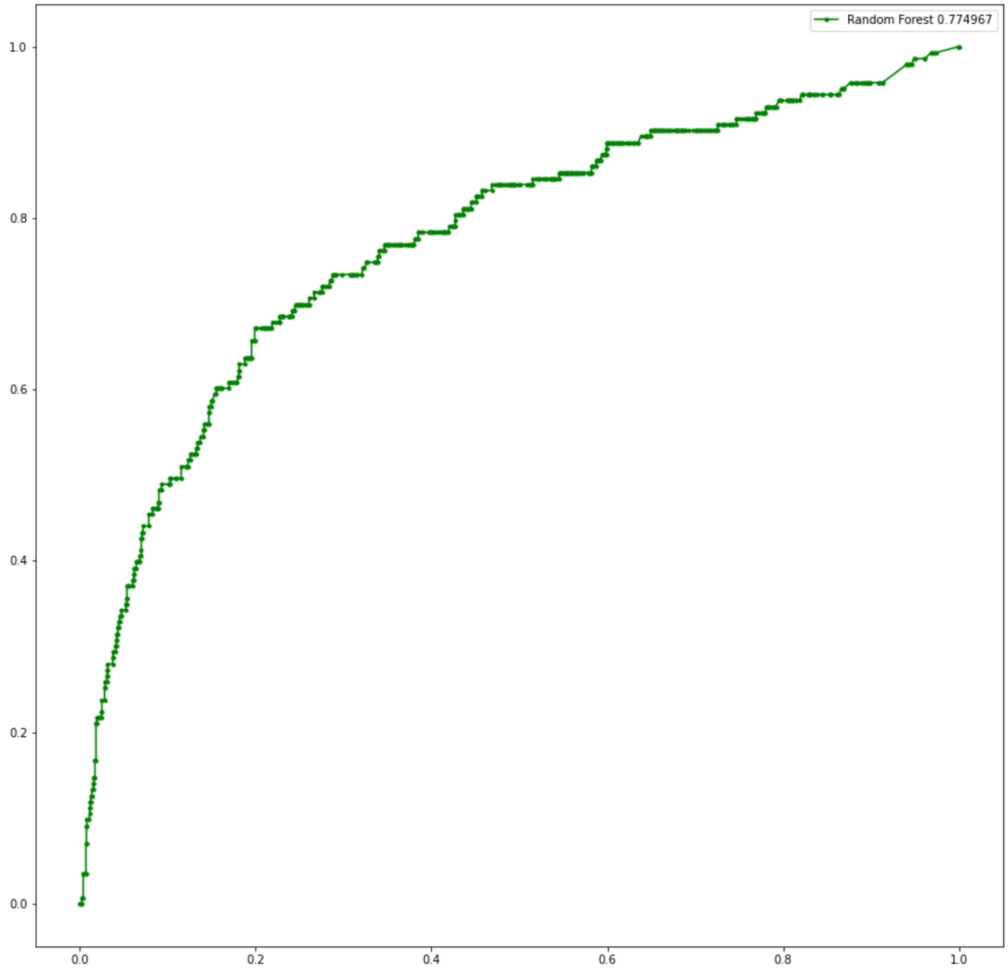
6. Business Case and Feature Selection

The primary issue with using these features to predict relationship dissatisfaction is that the most important features are all measures of the relationships' quality. Features such as these have little predictive value when it comes to, say, running a matchmaking service (e.g. It's Just Lunch) or dating application (Tinder, Coffee Meets Bagel, etc.). For this project, my objective is to predict the potential dissatisfaction a single person may have with their possible matches, in order to avoid taking that person on as a client. Therefore, questions such as "Are you proud of your husband/wife?" lose their validity despite their importance to the Random Forest model.

Shifting the focus to the *beginning* of relationships, a total of 8 features will be dropped from the model that evaluate relationships. These features represent the entirety of the MRQ except for one item, `happiness`. This item, which is among the most important from the Random Forest model, remains included due to the wording of the item on the questionnaire. The item simply asks "Are you happy?" without any reference to relationship quality. While there is little doubt that one's happiness and the quality of one's romantic relationship influence each other, this model assumes that happiness is independent enough to merit inclusion.

7. Model Selection

The Random Forest model was applied once more, this time using the new feature set:



With the new model, the top features become `happiness`, `material_situation`, `indv_pride_in_children`, `natl_pride_in_parents`, and `indv_pride_in_parents`. The 6th feature on the list is `age`, which will also be included because these products already collect this data, and it will be an easy inclusion into the model with some benefit. Running the Random Forest once more with this final feature set returned an AUC score of .784, slightly improving on the model and confirming the feature set as an effective one for this project.

In order to improve the classifier, I first went to the confusion matrix to locate instances of misclassification:

Threshold = .5 (default)	Predicted 0	Predicted 1
Actual 0	1652	0
Actual 1	143	0

Without even calculating the classification scores, there is a clear issue here: not one dissatisfied relationship was predicted at all! This is clearly due to the default .5 probability threshold in sklearn, so my next step was to explore thresholding options.

Returning to the business case, the purpose of this model is to predict whether or not a potential client is predisposed to be dissatisfied with a relationship. For a service intending to avoid making bad relationship matches, this algorithm is meant to help avoid taking on these clients in the first place. With a higher cost associated with an incorrect prediction of a satisfying relationship, this model should therefore look to favor the *recall* of unsatisfying relationships.

After testing a number of thresholds, I identified the following:

Threshold = .075	Predicted 0	Predicted 1
Actual 0	1221	431
Actual 1	47	96

This threshold provides a high number of predictions of unsatisfying relationships with a low precision of .18, but a recall score of .67 indicating greater ability to find *actual* unsatisfying relationships. The model is counterbalanced by predicting a number of true negatives as well (that is, satisfying relationships). Potential good clients are now less likely to be identified than with the old threshold of .5, but more importantly the model is better at predicting bad clients.

8. Conclusion

This model is designed to be a tool in the belt of matchmakers, and it accomplishes that task. Though the model is not perfect in any sense, it could help provide an extra screening for new clients or users. Furthermore, this product could be combined with other tools to form a broader service designed around detecting ultimate compatibility. Romantic relationships are complex, but this model provides a piece of the puzzle.

The major finding of this project is that a person's culture has an effect on the potential dissatisfaction of their relationship. The features that made it into the final model (with the exception of `happiness`) describe cultural attitudes of the participants. These cultural attitudes had greater predictive power than education level, religiosity, financial security, and other features that would seem to be more important at face.

Next steps would focus on delving into the positive side of marital satisfaction, rather than dissatisfaction. Unraveling the inner workings of relationship compatibility is the goal of any matchmaker or dating app, and using machine learning techniques to solve this riddle would be a great but fulfilling undertaking.