A Model of Whether Someone Decides to Get Married

Model:

$$Desire_{marriage} = \beta_0 + \beta_1 * \left(1 - e^{-k*age}\right) + \beta_2 * gender + \beta_3 * IR + \beta_3 * ES$$

$$Decision_{marriage} = \begin{cases} 1 & if\ Desire_{marriage} \geq 0 \\ 0 & if\ Desire_{marriage} < 0 \end{cases}$$

In the model, $\{\beta_i\}$ and k are constant coefficients that are currently unspecified. The inputs *age*, *gender* (0 = female, 1 = male), *IR* (0 = not in a relationship; 1 = in a relationship), and *ES* (0 = unemployed, 1 = employed) are the four exogenous variables. $Desire_{marriage}$ (the extend the person wants to get married), and $Decision_{marriage}$ (0 = not get married; 1 = get married) are endogenous variables.

My reasoning behind choosing these variables is that, I believe the decision to get married is largely influenced by three broad collections of factors, namely basic demographic features, factors of romantic relationship, and economic factors. From various demographic features, I believe age and gender can be two of the most deciding factors when it comes to marriage. Second, whether a person is currently in a romantic relationship is perhaps the most important that influences the decision of marriage. We can safely assume that people that are currently in a relationship are significantly more likely to get married as compared to those not in a relationship. Finally, employment status can influence the decision of marriage because employment is important for a solid economic foundation on which a marriage can reply upon.

Too keep the model relatively simple and straightforward, I have excluded some variables that might potentially play a role in influencing the decision of getting married. For example, while race and nationality may certainly influence a person's decision to get married, these variables are much harder to code and analyze than age and gender, and therefore are excluded from the model for simplicity. Moreover, for those people that are currently a romantic relationship, their level of satisfaction toward their current relationship may also affect whether they want to get married. However, I also excluded it from the current model because introducing the level of satisfaction toward the current relationship would necessitate an interaction term between the level of satisfaction and the indicator variable of being in a relationship, thus making the model much more complex.

To perform a preliminary test of the model, we can first collect a small sample of data either directly from Internet (if readily available) or through a survey, and then use machine learning to perform a classification of the following model to learn about whether the coefficients are statistically significant:

$$Marital\ Status = \hat{\beta}_0 + \hat{\beta}_1 * \left(1 - e^{-k*age}\right) + \hat{\beta}_2 * gender + \hat{\beta}_3 * IR + \hat{\beta}_4 * ES$$

While the model is not the same as our previously proposed model, the coefficients should be somewhat close. Therefore, the classification result should be a good indicator of whether the selected variables are significant.