**Final Project, Part 2**: Project Design Write-up and Approval Template

Miranda Remmer

**Project Problem and Hypothesis**

Using information gathered from a speed dating experiment, covering from 2002-2004, the data aims to explore the relationship between attributes (self and other), demographics, habits, and beliefs and how they influence relational goals, perceptions and choices.

*Key questions for exploration:*

* Does race, religion, age,
* Least desirable attributes in a male partner? Least desirable attributes in a female partner?
* Does the number of people the dater met in a given night influence whether or not they found matches? (round (# of people), match (0/1))
  + Data has also used ‘condtn’ for binary classification: 1 = limited choice (10 and under) / 2 = extensive choice (11 and up)
* Does order matter?

- What's the project about? What problem are you solving?

- Where does this seem to reside as a machine learning problem? Are you predicting some continuous number, or predicting a binary value?

- What kind of impact do you think it could have?

- What do you think will have the most impact in predicting the value you are interested in solving for?

**Datasets**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Description** | **Type of Variable** |
| order | Order met | Continuous |
| match | Binary (1 = yes | 0 = no) | Categorical |
| dec\_o | Decision of partner the night of event (1 = yes | 0 = no) | Categorical |
| age | Age of self | Continuous |
| age\_o | Age of partner | Continuous |
| race | Race of self (see below) | Categorical |
| race\_o | Race of partner (see below) | Categorical |
| samerace | Participant and the partner were the same race. 1= yes, 0=no | Categorical |
| imprace | Importance of same ethnic background of partner (scale 1-10) | Categorical |
|  |  |  |
|  |  |  |
| imprelig | Importance that a person you date be of the same religious background (on a scale of 1-10) | Categorical |
| field\_cd | Your field of study (coded) | Categorical |
| goal | Goal for participating in the event (1-6 – coded below) |  |
| date | How frequently do you go on dates (1-7: coded below) |  |
| go out | How often do you go out (not necessarily on dates) (1-7 coded below) |  |
|  | How interested are you in the following activities, on a scale of 1-10 |  |
|  |  |  |
|  |  |  |
|  |  |  |
| match\_es |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

race:

Black/African American=1

European/Caucasian-American=2

Latino/Hispanic American=3

Asian/Pacific Islander/Asian-American=4

Native American=5

Other=6

field\_cd: field coded

1= Law

2= Math

3= Social Science, Psychologist

4= Medical Science, Pharmaceuticals, and Bio Tech

5= Engineering

6= English/Creative Writing/ Journalism

7= History/Religion/Philosophy

8= Business/Econ/Finance

9= Education, Academia

10= Biological Sciences/Chemistry/Physics

11= Social Work

12= Undergrad/undecided

13=Political Science/International Affairs

14=Film

15=Fine Arts/Arts Administration

16=Languages

17=Architecture

18=Other

goal:

What is your primary goal in participating in this event?

Seemed like a fun night out=1

To meet new people=2

To get a date=3

Looking for a serious relationship=4

To say I did it=5

Other=6

date:

In general, how frequently do you go on dates?

Several times a week=1

Twice a week=2

Once a week=3

Twice a month=4

Once a month=5

Several times a year=6

Almost never=7

go out:

How often do you go out (not necessarily on dates)?

Several times a week=1

Twice a week=2

Once a week=3

Twice a month=4

Once a month=5

Several times a year=6

Almost never=7

- Description of data set available, at the field level

Domain knowledge

- What experience do you already have around this area?

- Does it relate or help inform the project in any way?

- What other research efforts exist?

- Use a quick Google search to see what approaches others have made, or talk with your colleagues if it is work related about previous attempts at similar problems.

- This could even just be something like "the marketing team put together a forecast in excel that doesn't do well."

- Include a benchmark, how other models have performed, even if you are unsure what the metric means.

Project Concerns

- What questions do you have about your project? What are you not sure you quite yet understand? (The more honest you are about this, the easier your instructors can help)

- What are the assumptions and caveats to the problem?

- What data do you not have access to but wish you had?

- What is already implied about the observations in your data set? For example, if your primary data set is twitter data, it may not be representative of the whole sample. (say, predicting who would win an election)

- What are the risks to the project?

- What's the cost of your model being wrong? (What's the benefit of your model being right?)

- Is any of the data incorrect? Could it be incorrect?

Outcomes

- What do you expect the output to look like?

- What does your target audience expect the output to look like?

- What gain do you expect from your most important feature on its own?

- How complicated does your model have to be?

- How successful does your project have to be in order to be considered a "success"?

- What will you do if the project is a bust (this happens! but it shouldn't here)?