DS 707
Business Understanding Document
Dating Analytics

Anirudh Ravi(IMT2013005)
Nigel Fernandez(IMT2013027)
Revanuru Karthik (IMT2013033)
Rishabh Manoj(IMT2013035)

October 12, 2016

1 Organizational Structure

Our team consists of four members. All members have the required skill set to contribute in major aspects of the project including data exploration and mining.

2 Problem Area Description

Our project "dating analytics" broadly involves obtaining actionable insights in the dating process. We aim to find association rules about dating preferences in either gender. For example, does a certain attribute increase the likelihood of being accepted? Are these attributes common to both men and women? Does shared culture, interests or race have an impact? These are some of the questions which we'll be exploring. From a business perspective, in a dating application our results can help increase accuracy of match probabilities thereby increasing customer satisfaction, expanding the customer base and reducing churn.

Our first goal is to find association rules, and correlations between desirable attributes and acceptance rates. After this, depending on time, we plan to use data mining to construct a prediction algorithm which classifies possible partners for a participant into three categories(likely match, neutral match and unlikely match). The association rules and correlations will be helpful to design such an algorithm. We'll be using the data set generated from a speed dating experiment involving the students of Columbia university as explained in a later section [1].

3   Current Existing Solution

There exists data from speed dating experiments, match prediction algorithms (Tinder [2]), correlation analysis between attributes and acceptance rates and results on the dating process. "Racial Preferences in Dating" [3] and "Gender Differences In Mate Selection" [4] are two papers which illustrate few of these results using the same data set from Columbia university. We'll be exploring existing algorithms, results and data mining techniques which will helpful to achieve our goals.

4   Business Objectives Definition

Our project is more academically inclined. As explained in the problem description section, our business objective could be to use the actionable insights obtained to improve existing dating applications (Ex: Tinder). The sorting algorithm ranking possible partners on the basis of match probabilities can be improved. These will achieve the following business objectives. There will be increased customer satisfaction due to relevant recommendations increasing application usage times. The increased popularity can lead to an increase in the customer base. There will be reduced churn in customers due to improved recommendations.

5   Business Success Criteria

From an academic perspective our primary goal is to obtain actionable insight(s) in the dating process. Our next goal, depending on time, is to design a match prediction algorithm. These results can be mapped to business objectives of a dating application. As mentioned above it can increase customer application usage times, increase the customer base and reduce customer churn. Improving the above business parameters by 10% is a realistic estimate.

6   Situation Assessment

We will provide a description of the data set in this section. The data set was compiled through a speed dating experiment involving students from Columbia university between 2002 to 2004. Participants engage in a four minute date and decide whether there are romantically interested in their partner. Surveys were conducted on the participants before the date, immediately after the date and three weeks after the date. Each participant was assigned a unique ID serving as the primary key in the database. Before the date, participants provided data points on various attributes including demographics, dating habits, self perception on various parameters, beliefs and opinion on the dating process and lifestyle parameters. Immediately after the date, participants decided whether there were romantically interested in their partner and rated their partner across

six attributes namely attractiveness, sincerity, intelligence, fun, ambition, and shared interests. After three weeks participants provided data points on attributes capturing their perception on the importance of various factors in the dating process after being through actual dates. This data obtained from three surveys was compiled into a database.

7 Resource Inventory

We'll be using the above data set for our analysis. Data is available is a csv file with numbers (mainly ordinal along with interval) as the major data type increasing the convenience of carrying out a statistical analysis. We have a sufficient team size (four members) with the required skill set to carry out the analysis. We have the required computing power (personal laptops) to carry out data mining since the data set in small (order of 10^3).

8 Requirements, Assumptions and Constraints

There is no legal restriction to use the data set for academic use. The data set is in the public domain. There are no economic factors involved since the project is academically in nature. We are making the data quality assumption that the data set taken in 2004 is a valid substitute for data from 2016.

9 Risks and Contingencies

The major risks involved are data quality and time. Our major assumption of the data set taken in 2004 being a valid substitute for data from 2016 may be incorrect. We may not have sufficient time to complete all the goals of our project namely the two parts where part one is identifying association rules and correlations and part two is designing a match prediction algorithm.

10 Terminology

1. **Date:** A social or romantic appointment or engagement

2. **Match:** Both participants accepting their romantic interest in their partner

3. **Churn:** Number of customers discontinuing use of service provided by the application
(Additional words will be populated with time.)

11 Cost/Benefit Analysis

There is no major economic factor involved in the project. The data set is freely available in the public domain. The tools and techniques used will be open source or obtained through a student license (Ex: Tableau). In return the actionable insights obtained can be used in a dating application to provide increased customer application usage time, increased customer base and reduced churn.

12 Data Mining Goals

Our main data mining goal is to perform a descriptive analysis on the data set. This will help us obtain association rules which can be transformed to actionable insights. Time permitting, these rules can further be applied to predictive analysis is which we aim to design a match prediction algorithm. A simplified version of the match predictor can use classification techniques to classify available partners for a participant into three categories (likely match, neutral match and unlikely match). This will help participants sort and filter out partners to choose to go out on a date with.

13 Data Mining Success Criteria

To evaluate the success criteria of our data mining results, we will divide the data set (order of 10^3) into training data and test data. We will use 80% of the data as training data and the remaining 20% as test data. Our association rules and classification techniques will be trained on this training data and used on the test data. A success rate above 60% on the training data will be aimed for.

14 Project Plan

Phase	Time	Date
Business Understanding	1 week	12/10/16
Data Understanding	1 week	19/10/16
Data Preparation	1 week	26/10/16
Modelling	2 week	09/11/16
Evaluation	1 week	16/11/16
Deployment	1 week	23/11/16

References

[1] Speed Dating Experiment
http://www.stat.columbia.edu/ gelman/arm/examples/speed.dating/
12 October 2016

[2] Tinder.
https://www.gotinder.com/ 12 October 2016

[3] Fisman Raymond, Iyengar Sheena, Kamenica Emir and Simonson Itamar. `Racial Preferences in Dating`. Review of Economic Studies 75,2008

[4] Fisman Raymond, Iyengar Sheena, Kamenica Emir and Simonson Itamar. `Gender Differences In Mate Selection: Evidence From A Speed Dating Experiment`. The Quarterly Journal of Economics,2006