Introduction

Our client MyHealthTeams is an independent startup based in San Francisco that provides social networks for people facing chronic conditions. Via the online platform, users have the option to share information regarding their diagnoses, symptoms, comorbidities, treatments, and details about each treatment. Using this self-reported data, our group performs data visualization and statistical analysis to find correlations and provide insight into the perceived efficacy of different treatment options.

Objectives

1. Visualize medication data for understanding treatment pathways;
2. Enable insights regarding the prescription prevalence of different treatments, prescription variability over time, and correlations to reported efficacy;
3. Make the process of data exploration and visualization easily replicable for the future data.

<table>
<thead>
<tr>
<th>Data Cleaning</th>
<th>Data Visualization</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>R-Studio:</strong></td>
<td><strong>Sunbursts:</strong></td>
</tr>
<tr>
<td>● Remove unnecessary symbols and spacing</td>
<td>● Used html, javascript, and css to plot radial tree maps for treatment pathways</td>
</tr>
<tr>
<td>● Uniformize the treatment list by replacing all off brand drug names/active ingredients with their more well known name</td>
<td></td>
</tr>
<tr>
<td>● Initially spell checked by hand</td>
<td><strong>Likert Scale:</strong></td>
</tr>
<tr>
<td>● Replaced this with an automated drug spell checker to expedite process, using an NIH online database</td>
<td>● Created a stacked bar graph to visualize users’ perceived efficacies of treatments</td>
</tr>
<tr>
<td><strong>Java:</strong></td>
<td><strong>Predictive Model for Response to Medications</strong></td>
</tr>
<tr>
<td>● Manipulate the data into a usable format</td>
<td>● Use the R package rpart to create random decision forests</td>
</tr>
<tr>
<td>● Program was written to expand user information from a single cell entry into columns separated by category</td>
<td></td>
</tr>
</tbody>
</table>
Analysis & Results

Visualization for Categorical Data

Figure 1. Percentage of treatments listed among total treatment entries

Figure 2. Frequency of users from different countries/states

Figure 3. Proportion of users reporting treatments

Random Forest

Figure 4. Random Forest: Predictive Model for Response to Medication

- Accuracy based on an 80% training set and 20% testing set: 51.8%.
- Only significant factors in predicting efficacy were currently or not currently using the treatment and treatment used.
Analysis & Results

Sunbursts: Treatment Pathways

- 12251 viable users out of 47052 total users;
- 1.48 treatments listed per user on average;
- 686 unique treatment entries;
- Most users reported only one treatment, and thus longer trails do not have enough users to be displayed.

- Treatment pathways: the chronological sequence of treatments that a patient takes;
- A radial treemap (sunburst) displays treatment pathways for a subset of users;
- Users grouped by type of depression, gender, symptoms, and comorbidities.

Reference:

Likert Scale: Efficacy of Treatments

- Efficacy Scale: 0 for ineffective; 1 for effective; 2 for very effective

Figure 5: Sunburst for users with Seasonal Disorder Depression (134 users)

Figure 6: Sunburst for all viable MyDepression Team users

Figure 7: Average Effectiveness 1.09

Figure 8: Average Effectiveness 1.25

Figure 9: Average Effectiveness 0.73