Survival Random Forest to Predict Time to Fill

Summer M. Husband, PhD
Senior Director, Data Science
Randstad Sourceright
time to fill

- Large volume of jobs to be filled for a client
- Time to Fill (TTF) – days from the date the requisition was open to the date it was filled
- Target TTF SLAs set based on complexity of the job
- Predict TTF based on job features? Early in the process, flag reqs with a high probability of remaining open at target TTF.
- Assign additional sourcing to high risk reqs. Adjust recruiter load based on complexity of reqs.
Survival Random Forest Algorithm:
- Draw bootstrap samples from the data
- Survival tree for each sample
- Cumulative hazard function for each tree. Averaged for ensemble CHF
- Calculate prediction error with out of bag data.
## Feature List

- Candidate Submitted
- Candidate Interviewed
- Month Opened
- State
- City Category
- State Desirability
- Business Unit
- Job Level
- City Job Recruiter Frequency
- Job Title Group
- Job Function
- National Active Candidates per Job Posting
- HR Manager Previous Experience
- Recruiter Experience in Job Category

### Target TTF Predictor Features

<table>
<thead>
<tr>
<th>Feature Name</th>
<th>Description</th>
<th>Relative Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Candidate Interviewed</td>
<td>Candidate interviewed in 1st 2 weeks, 3rd week or neither</td>
<td>1.0</td>
</tr>
<tr>
<td>Candidate Submitted</td>
<td>Candidate submitted in 1st week, 2nd week or neither</td>
<td>0.45</td>
</tr>
<tr>
<td>Target TTF</td>
<td>Target TTF</td>
<td>0.37</td>
</tr>
<tr>
<td>Job Level</td>
<td>0-4, indicates job complexity</td>
<td>0.31</td>
</tr>
<tr>
<td>City Job Recruiter Frequency</td>
<td># times the recruiter has been assigned this job title in this city</td>
<td>0.24</td>
</tr>
<tr>
<td>Job Category</td>
<td>Broad job category</td>
<td>0.17</td>
</tr>
<tr>
<td>National Active Candidates per Job Posting</td>
<td>Active candidates/job posting in the job category and level, pulled from Career Builder</td>
<td>0.14</td>
</tr>
<tr>
<td>Month Opened</td>
<td>Month of year</td>
<td>0.11</td>
</tr>
<tr>
<td>State</td>
<td>State or province</td>
<td>0.10</td>
</tr>
<tr>
<td>City Category</td>
<td>Most common cities in the data or Other</td>
<td>0.10</td>
</tr>
<tr>
<td>Job Title Group</td>
<td>Most common job titles in the data or Other</td>
<td>0.09</td>
</tr>
<tr>
<td>Business Unit</td>
<td>Client specific business unit</td>
<td>0.09</td>
</tr>
<tr>
<td>State Desirability</td>
<td>State desirability pulled from census data</td>
<td>0.07</td>
</tr>
<tr>
<td>Country</td>
<td>Country</td>
<td>0.01</td>
</tr>
</tbody>
</table>
**ttf predictor**

### Data Sources
- Historical TTF data from client
- Recruiter experience data from client
- CareerBuilder stats
- State desirability

### Algorithm
**Survival Random Forest**

- `R randomforestSRC`
- `nsplit = 10`
- `ntree = 1000`
- `na.action = na.impute`
- `importance = permute`

### Output to Recruiters

- **Red** [0.75, 1.0]
- **Yellow** [0.5, 0.75)
- **Green** [0, 0.5)
recruiter dashboard
predictor performance

% Met/Missed Target

Low | Med | High
---|-----|-----
Met Target | Missed Target 1-15 days | Missed Target 16-30 days | Missed Target 31-60 days | Missed Target 61+ days

- Low: 100% Met Target, 0% Missed Target
- Med: 80% Met Target, 20% Missed Target
- High: 60% Met Target, 40% Missed Target
sourcing resources

<table>
<thead>
<tr>
<th>Truth</th>
<th>Met Target</th>
<th>Missed Target 1-15</th>
<th>Missed Target 16-30</th>
<th>Missed Target 31-60</th>
<th>Missed Target 61+</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6271</td>
<td>1933</td>
<td>749</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>756</td>
<td>1051</td>
<td>1045</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>420</td>
<td>710</td>
<td>890</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>459</td>
<td>774</td>
<td>1185</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>432</td>
<td>915</td>
<td>1383</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

24.8%  64.1%  85.7%
future steps

- Roll out predictor for other clients
- More market data
- Develop more sophisticated location features
- Compare with other models