Eros Erdős

Michelle Heeney, Abigail Boulin, Kirs Imsong, Roy Jackman, and Jack Kenney

March 25, 2018

Abstract

Trying to help employees find communities, and helping employers communicate effectively.

1 Objective

For our Data Fest project, we decided to apply machine learning methods to produce usable data for both the employees and the employers. Given the distribution of job offers across the United States, we built a model that can predict the quality of a company job posting based on a number of different parameters, so the employers can tailor their posts for specific locations and employee types. For the employees, we have made a heat map based on industry, normalized using city population so you can find a job in a community of similar, or maybe not so similar, industries.

2 Methodology

1. We made several linear regression models, the best of which explained 65% of the variation in the data. Thinking that we could do better than this, we applied a machine learning model known as Multilayer Perception Classification.

2. Applying the machine learning model of neural network classifiers using the fields employeeJobCount, employeeCount, descriptionLengthChars, licenseRequiredJob, noEducationRequirementsJob, highschoolEducationRequirementsJob, higherEducationRequirementsJob, supervisingJob, jobAgeDays, admin1, and city. We predicted the overall company rating using these fields with 75% confidence.

3. We wanted to represent interesting metrics geographically since jobs are so closely tied to the job markets of the cities in which they are posted. This is why we chose to use Tableau which has powerful pre-built mapping tools. We were able to plot industry size by city but we were running into a problem that larger cities were overly represented, which makes sense. To better show different city industries, we decided to normalize our data by city population. We went ahead and pulled in a US Population data from data.un.org and performed a left-outer-join with our original dataset from Indeed.com. We were now able to show the best places to post jobs, based off of industry, and normalized to better represent smaller cities. Heatmap Image

3 Citations

1. Indeed Jobs Data Set

2. UN Population Data