Translational Data Analytics @ Ohio State
Hackathon Challenge

Nov. 14-15, 2015

Thank you for taking up the TDA@OhioState hackathon challenge! Data science is among the fastest growing job markets in the country, and among the hardest jobs to fill. Where's the disconnect between employers and graduates? What can be done to help students land that dream job or internship? Help us find out!

Our challenge is designed to test your skills by asking teams to develop algorithms to study raw data. Resume and job data by nature tend to be difficult to standardize. So we challenge you to study, link, and visualize raw data sets on internship/job opportunities in data analysis and student resumes.

The Data

Teams will be given two raw data sets pulled from the Buckeye Careers Network database, Ohio State's university-wide online resource to connect students with employers.

- The first data set includes 262 full-time positions and internships from Sept. 2015. Data is broken out by job title, description, position type, majors/concentrations, and qualifications.
- The second data set contains resume data from 91 current and recent students, including major/minor, graduation year, employment experience, skills (computer, language, or otherwise), courses, and projects.

The Challenge

Teams of **no more than 4 individuals** are expected to:

(1) develop matching algorithms to evaluate the raw data,

(2) identify what similarities and gaps exist between what employers look for and what students list on resumes,

(3) visualize their findings,

(4) make 2 recommendations for addressing any issues found, and

(5) upload their code, results, etc. to the Hackathon GitHub.

THE OHIO STATE UNIVERSITY

Teams may use any as open-source/access products or tools at their disposal to develop the algorithms. However, the results should garner significant knowledge of what similarities and gaps exist between the data sets. Specifically, teams are asked to discover similarities and differences in:

a) word(s) used by employers and students,
b) key skills required vs. key skills listed on resumes,
c) level of experience required and how that matches with resume data,
d) the most common major/minor listed for positions,
e) the most useful courses for positions, and
f) other creative / useful data points gleaned from the data (i.e. not listed above).

Teams are **highly encouraged** to be creative in their study of the data sets, searching to find unique relationships (similarities and differences) between the data.

After creating algorithms to address the above areas, teams should visualize their findings in a pictorial or graphical format. This can include charts, graphs, animation, hierarchies representations, networking, or interactive objects. Teams may use any software or tools at their disposal to visualize the data. Visualizations will **not** need to be printed, but should be readily available on a laptop / screen for judging.

Teams should come up with a minimum of 2 recommendations for addressing the issues found. One should focus from the student perspective (i.e. students should use the word "analysis" more in their resumes) and one from the employer perspective (i.e. employers should include more majors).

Finally, teams are required to upload their code, results, and products to the Hackathon GitHub.

Evaluation

Teams will be judged on their ability to:

1) meet the minimum requirements of the challenge,
2) draw reasonable conclusions using the data provided,
3) the depth of their analysis in addressing the similarities and differences noted above,
4) the value of their analysis in terms of what similarities and gaps exist between the data sets,
5) how well the findings are visualized, and
6) the quality of their recommendations.

Each team will be expected to speak with judges regarding their work, including answering questions regarding the process for developing algorithms used, results, how data was visualized, and your recommendations.

THE OHIO STATE UNIVERSITY