Domain

I've always wanted to do an NLP project identifying how much of a match certain people are based on their profiles to roles they're interested in. Eventually, I would like to focus on a project that could identify how much of a match someone is to any job role based on their school, work experience, skills and other information. This would help people decide whether they're a match for certain jobs or not (if they make industry switches).

Data

Gathering data on resumes is difficult, but I will be obtaining 30,000 resumes from Indeed.com – mostly Data Analyst, Data Scientist and Data Engineer resumes. The resumes are formatted according to some common fields such as summaries, job titles, companies, education, skills, current location, work start date and end date.

Known Unknowns

I'm not decided on what kinds of models would work best. Should I use an RNN? How should I vectorize key words and skills? I'm excited to work this out because I want to see how much of a match I am to data scientist roles (praying I get a job here), specifically roles related to resume matching. #Projectception