

Think Topic Questions

Cully West

✉ cullywest@gmail.com ☎ 785-218-3303 🌐 github.com/supwest

1

I'm familiar with a variety of machine learning techniques. I have a strong background in linear and logistic regression, and I also enjoy the tree based methods for regression and classification. I've also very familiar with Bayesian Networks. Recently I've been working on a project using matrix factorization for a recommendation system based on latent preferences.

2

For a minimum viable product I imagine being able to allow a recruiter to specify the language(s) and proficiency they're looking for in an engineer, and then delivering a list of choices most similar to their specification. So, for the initial dataset I would get the users/engineers who are hireable, then get the number of lines of each language present in their repos. This could be used to give a language proficiency profile for each engineer. Then you could use a similarity/distance measure to rank engineers by closeness to the ideal candidate. That's the low hanging fruit. It allows a recruiter to at least narrow down how many people they need to look at, and only be looking at people who meet some basic qualifications.

Going forward there are several ways to get a more sophisticated and hopefully better performing model. An interesting approach might be to look at the actual code itself, and use some text processing methods to get an idea of a type of engineer's specialty in terms of both language and application. Additionally I'd be interested in building a matrix, then decompose it and use the new features to find engineers similar to the recruiters ideal candidate.

I am most experienced with Python, so for the minimum product I would think of using the traditional Python tools like pandas, numpy and scikit-learn. The more advanced model would involve some text analysis tools and linear algebra libraries. Analyzing all Github users would be a fairly large task, so it might be beneficial (or even necessary) to move into some big data tools and a language with functional capabilities.