What kind of data would be necessary to build the model?

In order to train a model, the data must basically be structured data. A label is needed to determine whether a person is a high-performing software engineer or not. Label can be binary(0 : low performing, 1 : high performing) or numerical(0 ~ 100, the higher the number, the higher performing). Except for the label, the remaining columns must be numerical, categorical or text-type variables.

List that can be helpful to hire software engineers.

[programming languages, the number of projects]

1. Programming languages used by high-performing software engineers at the company  
   Because people who know the programming languages can achieve high performing in a short period of time.
2. the number of projects they do  
   I think the larger the number of projects, the easier it will be to analyze and modify code that has already been written, and write code for people to see. Also, I think the more collaborative projects there are, the easier communication and collaboration will be.