

Critique of Data Demystifiers (Benefi) Progress Video

What we liked:

- Their explanation of the business problem, and converting it into a machine learning problem was made simple and easily understood
- A decent explanation of the two datasets (LinkedIn and Glassdoor) they had
- A good explanation of what features to be engineered. I like how they are developing their own features for this model.
- Liked how they explained what their next steps were.

What we didn't like:

- Not sure if we understood how the feature "number of months worked per period". If the team did a one-hot encoding for the number of months worked per period, this may not work very well; it may very well overfit to the data. Overall, I am not sure if it is so useful in engineering "number of months worked per period" feature. Perhaps a more useful take is to take the average number of months worked per period.
- Would have liked to see a more in-depth explanation of some of the other features they would be using from LinkedIn and Glassdoor, for there appears to be some good features that could be used and further developed.
- We would have liked to have been shown feature importance and how they decided to get rid of the rest of the data.
- Would have liked to see more about how the education level, job experience and certifications was encoded.

Suggestions:

- Maybe don't use NN as it would not be interpretable, which may be needed to be used for loans decisions, auditing, tweaking, or regulations requirements