

IDEAS ON ISTHISAREALJOB.COM

Data Collection:

- 1) Allow users to input through various ways like text of email, link to a site offering job opportunities, name of the company, picture form like screenshot of email, etc
- 2) For every label made there should be an upvote or downvote and comments made should be analyzed by conversational analysis to understand what people think about the place and the label given to it.
- 3) After each labeling of a job, it stores it in its database for future references to that company, in case anyone asks in relation to it.

Analysis:

- 1) For text input, the structure of the mail is analyzed, to check for wordings mistake and grammar structures, and check for the name of the company in the email. If unnecessary mistakes are found, it is justified high-risk. (Conversational analysis, Natural language processing)
- 2) When the name of the company is found or if the input is the name of the company, it checks it against the registered companies by CAC. If found there, it is justified low-risk, if not found it is justified high-risk.
- 3) If screenshot is seen by computer, it is also analyzed like 1) above. (Computer Vision, Natural language processing(Nlp))
- 4) You can also upload your documents and for each job uploaded that have been labeled Not-Scam, it can check for your eligibility too and if compatible, advises you to apply for it. (Nlp)
- 5) If the name of the company is not easily found online and they do not have a company profile, it will be labeled high risk. (Nlp)
- 6) The location is also checked, to check the number of check-ins, and name of place if given to analyse if it's supposed to be a residential address or a company exists there so as to justify high-risk or low-risk respectively. (Computer vision, bot)
- 7) If comments have been made towards it before, it check the analyses the comment to know the type of comment made, if it's a complaint, it is labeled high-risk.

Labeling/Prediction:

We'll be using a discriminative model:

- 1) The bot visit the site if a link is given and goes through it, if the company is not registered under CAC, and former complaints have been made towards it before, it labels it scam.
- 2) If for a particular company, it receives more than one high-risk caution, it is labeled Scam.
- 3) If for a particular company, more than one low-risk tag with no high-risk tag at all, it labels it Not-Scam.

4) If both high-risk and low-risk is found, it takes focus on the high-risk, and if the number of high-risk is beyond two, it is labeled Scam.

Correction:

Sometimes the model may predict incorrectly:

- 1) A place to vote is there, if the number of upvotes for a particular company labeled scam is overwhelming, it is relabeled Not-Scam and vice versa.
- 2) Analysis of conversation to know what people say concerning the company and if views contradict the label given, it is rechecked again and if found correct, label would be changed.