Capstone Project Proposal

YAN QIN

Business Goals

Project Overview and Goal What is the industry problem you are trying to solve? Why use ML/AI in solving this task? Be as specific as you can when describing how ML/AI can provide value. For example, if you're labeling images, how will this help the business?	Current job seeking websites are more supply oriented. There is pool of job advertisements. Job seekers look at ads then submit resumes. Employers have to go through large amount of applications, setup interviews then make decisions. It's expensive, time consuming and does not guarantee to get the best candidates. I am going to design a ML model to streamline and simplify the process. For example, every open position is labelled with years of experience, key skills, previous employers and leadership qualities based on existing employee with similar role. Job applicant just needs to submit their resumes to an online system which would extract certain labels then match back to the position and see what's the best chance of fit. This matching process is wholly handled by ML in real time. It reduces labor effort to manually read through documents and human bias for subjective judgements. Companies can save lots money to pay for job advertiser and reduce internal headcounts for manual resumes processing which is very time consuming.
Business Case Why is this an important problem to solve? Make a case for building this product in terms of its impact on recurring revenue, market share, customer happiness and/or other drivers of business success.	Current job recruiting process is not automated enough. Instead it's expensive, very time consuming and with fairly low matching rate. This machine learning system will automate supply/demand through a labeling system and provide better and faster decision making. It also will generate significant financial benefits by saving money from pay job advertisers and reducing labor cost of human resource.

Application of ML/AI What precise task will you use ML/AI to accomplish? What business outcome or objective will you achieve?	Key tasks: 1. Every open position is labelled and cross checked with existing employees with similar attributes. 2. Once resumes are submitted online, all key sections and key attributes are atomically identified and extracted. 3. ML system will do real-time matching through an underlying algorithm to put high quality candidates in a pool. Business goals: 1. Automate recruiting process, improve efficiency and reduce human bias. 2. Reduce recruiting time and cost. 3. Create a better and faster decision making process.

Success Metrics

Success Metrics What business metrics will you apply to determine the success of your product? Good metrics are clearly defined and easily measurable. Specify how you will establish a baseline value to provide a point of comparison.	1. How automate for the whole recruiting process without human involvement. 2. What's money value saving by implementing such system. 3. How much time is saved by going through this automated process.

Data

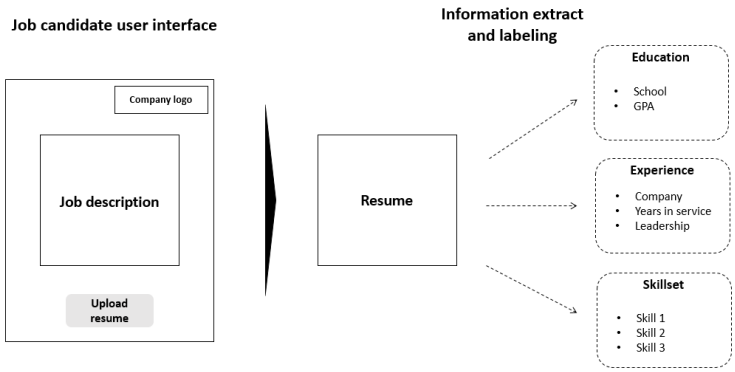
Data Acquisition Where will you source your data from? What is the cost to acquire these data? Are there any personally identifying information (PII) or data sensitivity issues you will need to overcome? Will data become available on an ongoing basis, or will you acquire a large batch of data that will need to be refreshed?	There are huge amounts of existing resumes in company career websites. Data model can be built based on previous positions and resumes then tested to see if job candidates in place can be predicted from the model. It's an ongoing process as long as there are open positions. Data privacy policy can be parallel with existing policy as job candidates can see all disclaimer when they submit their resumes through company career site.
Data Source Consider the size and source of your data; what biases are built into the data and how might the data be improved?	Job candidates always try to match their resumes with job descriptions. Undifferentiated resumes would make it difficult for ML model to decide best candidate.
Choice of Data Labels What labels did you decide to add to your data? And why did you decide on these labels versus any other option?	Label would be based on job functions and ranks. Algorithm would be in place to create a context based on labels. General label would be years of experience, key skills, previous employers, education and leadership qualities.

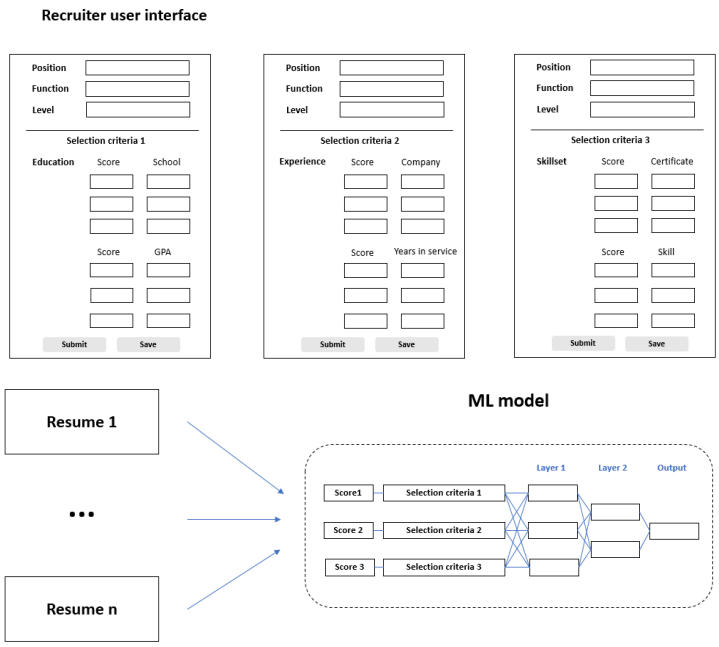
Model

Model Building How will you resource building the model that you need? Will you outsource model training	Data is already existed. To be more capital efficient, can choose Amazon AWS or MS Azure as platform for model building and training. The cost is subscription based so can pay as you go without much string attached.

and/or hosting to an external platform, or will you build the model using an in-house team, and why?	
Evaluating Results Which model performance metrics are appropriate to measure the success of your model? What level of performance is required?	1. Recruits feedback. They are part of feedback loop and will always do random check and see if model result is ideal. 2. If model is hosted through platform, a cost benefit analysis can be done to compare from traditional recruiting. 3. How much time to get right candidate onboard can be measured as a benchmark,

Minimum Viable Product (MVP)

Design What does your minimum viable product look like? Include sketches of your product.	



Use Cases

What persona are you designing for? Can you describe the major epic-level use cases your product addresses? How will users access this product?

The ML model is to automate resumes checking process and improve the speed and accuracy and reduce workload of human resource so they can offload the work and focus on other more value-added projects. The system will automatically process resumes and label key attributes for job candidates based on their profiles.
For job candidates, the only thing they need to do is to upload their resumes to career site.
For recruiters, they need to help define the labels, key attributes and score based on open positions so ML model can train itself based on these labels from existing resumes data then make context-based prediction enabled by underlying algorithm on the best candidates with a score number.

Roll-out

How will this be adopted? What does the go-to-market plan look like?

1. MVP is rolled out in 8 weeks with 4 iterations.
2. Users feedback is collected and put into next version. New version is rolled out every 4 weeks after MVP.
3. Scrum team (PM, data scientist and engineers) work together with recruiters to fine tune algorithm and label methodologies and cross check prediction result.

Post-MVP-Deployment

Designing for Longevity How might you improve your product in the long-term? How might real-world data be different from the training data? How will your product learn from new data? How might you employ A/B testing to improve your product?	ML model would be further classified by job functions and ranks. Label methodologies and algorithms would be more customized to match selection criteria and improve prediction success rate. ML model is prone to trust more about training data. Real world data needs to be fed into the model to test model output and justify the result. Model can be further trained based on 80% training data and 20% real world data to check result discrepancy.
Monitor Bias How do you plan to monitor or mitigate unwanted bias in your model?	Build a feedback loop to have scrum team (PM, data scientist and engineers) work together with recruiters and cross check prediction result so ML can be adjusted and tested with the most ideal combination.