Capstone Project Proposal

Hamad S. Alassafi

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?	The problem that I will try to solve is in Human Resources (HR) industry by detecting whether an individual who works in certain organization will likely to leave the organization or will stay, Machine Learning can help us in this project to classify the employees in two categories (Leave, and Stay) by taking multiple criteria from each employee for example (Promotions, Salary increasing, other Benefits such as discount, etc…), we will probably uses a supervised learning algorithms for classification such as (Random Forest).
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.	The problem that I have mentioned earlier is highly important, because in each organization the most valuable asset on it their employees, so if we build this machine learning model we will understand the issues that faces the unhappy employees and wants to leave, then we will have the ability to convince them to stay in the appropriate manner.

Application of ML/AI What precise task will you use ML/AI to accomplish? What business outcome or objective will you achieve?	We will probably uses a supervised learning algorithms for classification such as (Random Forest), to classify each employee to decide whether he/she will (Stay, or Leave), the outcome of using this particular model is to increase the employee happiness and satisfaction

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.	Multiple metrics can be used. • Employee Satisfaction Increases • Decreasing in Employees Leaving the Organization • Increasing in Employees Productivity I will take the average leaving employee on year for the past five years and compare it with the leaving after deploying the model, and make it as baseline

Data

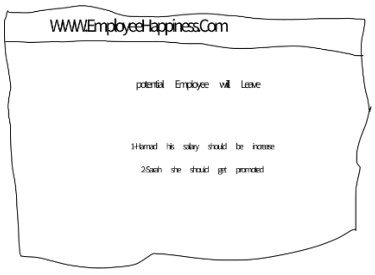
Data Acquisition Where will you source your data from? What is the cost to acquire these data? Are there any personally identifying information (PII) or data sensitivity issues you will need to overcome? Will data become available on an ongoing basis, or will you acquire a large batch of data that will need to be refreshed?	The Data Acquisition can be done from the Human Resources (HR) for each employee, there are not cost in gathering the data because it is in the organization's database, but there are problem of Personally Identifying Information (PII), because we will get the data of each employee from his salary to promotions, etc…, the data will be supposed as always refreshed from the organization database

Data Source Consider the size and source of your data; what biases are built into the data and how might the data be improved?	Algorithm Bias: The biases that can happens in algorithm, that we do not know why the past employees have leaved the organization so maybe that will causes some biases Measurement Bias: Overfitting of the model, so we think it perform well but it actually perform well in the training set only Exclusion Bias: We may faces outliers in the dataset, which is considered as bias, we should try to handle it Human Bias: The model developer could neglect certain type of employees which will causes a bias in the model.
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?	I have decide to use these labels • Salary • Promotions • Years of Experience Because these in my opinion is the most important labels to decide whether to leave a organization or not, because if you have large years of experience and you salary is low also you did not get any promotions you will most likely leave the organization

Model

Model Building How will you resource building the model that you need? Will you outsource model training and/or hosting to an external platform, or will you build the model using an in-house team, and why?	I will use an in-house team, because the sensitivity of the information will force us to make it in-house.
Evaluating Results Which model performance metrics are appropriate to measure the success of your model? What level of performance is required?	There are multiple evaluating criteria can be taken • Accuracy • There are no Overfitting It should perform very accurately to make it feasible to be applied, the accuracy should be (90%) at least, also to spot the overfitting we should look for the accuracy of the test set

Minimum Viable Product (MVP)

Design What does your minimum viable product look like? Include sketches of your product.	 Portal that contains the potential employees that will leave
Use Cases What persona are you designing for? Can you describe the major epic-level use cases your product addresses? How will	The Human Resources (HR) employees can uses web page or portal to access the model, after accessing the portal they will find a list of all employees that will potentially leave the organization, and the reason behind their leaving by the model

users access this product?	
Roll-out How will this be adopted? What does the go-to-market plan look like?	1. Who: The targeted people is the employees 2. What: Machine Learning model to decide whether the employee will stay or not 3. Why: to increase employee happiness, and make them want to stay 4. Where: in Human Resources (HR) Department 5. How: Through portal or web page 6. When: After testing and deploying the model, and make sure it works correctly

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?	Using large amount of data, from multiple sources and organizations will enhances the model performance, there are a margin of error for example (employee is happy and have good salary, but he received a bigger offer from another company then he/she will most likely leave the organization), it will learn from applying large amount of data on it using real-world data, I can use A/B testing to decide if I used years of experience or not will affect the model accuracy, so the null hypothesis will be using years of experience in thee model, and the alternative hypothesis will be not using the years of experience in the ML model.
Monitor Bias How do you plan to monitor or mitigate unwanted bias in your model?	I can make follow-up to employees who leaved from the company to find out the reasons behind their leaving and feed it to the model, which will mitigate unwanted biases from the model in my opinion.