Capstone Project Proposal

Ekta Bharti

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?	• Industry problem - Manufacturing industry • ML/AI would be used in solving the use case of predictive maintenance. • ML/AI will help predict when exactly in future the centrifugal pumps will break or defect so that as Industry experts, we can plan to put preventive measures for the same. This will prevent down time and increase uptime saving lots of cost wasted when the centrifugal pump is not working. The temperature, vibration, pressure, centrifugal pump rotation speed, centrifugal pump age and other time series data of product telemetry will be used to train the model. • The label in this case will be the time when the centrifugal pump has failed.
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.	• Its important to solve this problem because failure of the centrifugal pump will end up in shut down of the entire process and will lead to downtime unless the centrifugal pump is repaired. • The down time will lead to delayed delivery of the product, wastage of man hours of the labourers for downtime, the B2 B customer will get the product late, the quality delivered will also suffer and hence will decrease the rating of the manufacturing company.
Application of ML/AI What precise task will you use ML/AI to accomplish? What	• ML/AI will be used to predict when exactly in future the centrifugal pump is going to fail. • It will help be prepared beforehand to handle situations of failure and downtime by replacing

business outcome or objective will you achieve?	the centrifugal pump or by providing maintenance activities on the centrifugal pump.

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.	Business metric -> Productivity per hour= value of good produced in n number of hours/n hours n-> number of hrs in a shift To mark the baseline, we will calculate the current productivity without applying predictive maintenance and compare it after applying predictive maintenance algorithm outcomes to determine failure ahead and be prepared for centrifugal pump failure beforehand.

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?	All data received via sensors attached to centrifugal pumps would be gathered. Cost to acquired would be the cost of sensors to measure temperature, pressure, vibration etc. No PII data. Data required: Large batch of past data would be gathered and it would be refreshed once a day.
Data Source Consider the size and source of	Extra weightage needed on recent data Biases could be built based on the following:-

your data; what biases are built into the data and how might the data be improved?	1) Periodicity based on weather 2) Quality of supporting tools 3) Corrosion – dependent on point 2 4) Over heating– dependent on point 2 Data can be improved only for point 1 a) where we can add a periodic factor to weather conditions which affects the temperature of the centrifugal pump and its working condition. b) where we can ensure equal data points for all months. Quality of supporting tools, Corrosion and Overheating factors will automatically be captured in vibration, temperature related data points captured and hence would not need to be dealt with separately.
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?	• Labels will be the date and time stamp when centrifugal pump fails • These labels are chosen because we are predicting the date and time in future when the centrifugal pump will fail.

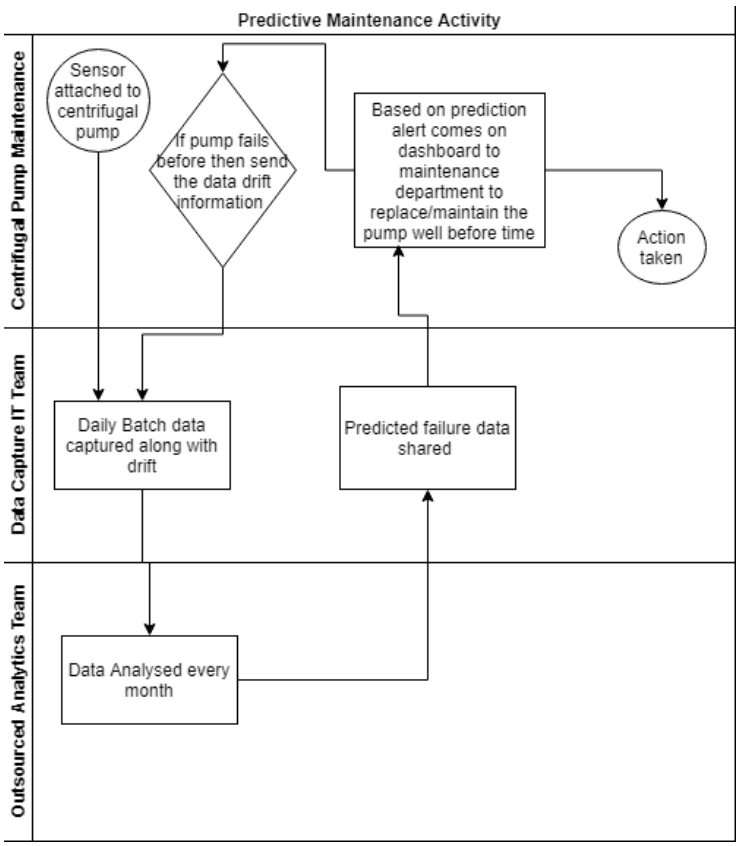
Model

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?	Usually manufacturing companies do not have analytics team of their own. They can help with obtaining the data by attaching sensors to measure temperature, pressure, humidity, and velocity (rpm) etc. and also capture the time series data from company records , for example when centrifugal pump was replaced and after how long the centrifugal pump failed. So, this data would be collected from manufacturing units itself. The model training can be outsourced and hosted outsourced to an external platform since we do not need to setup a permanent team to carry out this activity daily. This could be a periodic activity in a year when the modelling services of a third party can be outsourced.

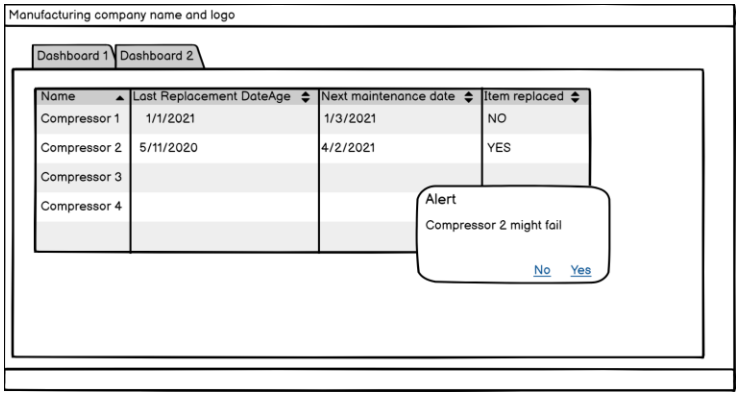
	Moreover, from data security point of view, the data can be hosted to external platforms since the leakage of data won't harm the company as it is not sensitive
Evaluating Results Which model performance metrics are appropriate to measure the success of your model? What level of performance is required?	Given that our business metric is productivity and that can only be improved if the accuracy of future failure of centrifugal pump is predicted correctly. Centrifugal pumps are used to supply low viscosity liquids. If the liquid is not pumped then the delivered output will be ruined because of excessive heating. This will affect the number of delivered output/products manufactured. Hence productivity is the business metric. The model could be regression given the date-time stamp is continuous and the model performance could be achieved by accuracy in predicting timestamp. When predicting the date-timestamp when centrifugal pump fails, we need to be accurate as it will impact the entire chain. Hence Mean Square error will be measure of error since it will penalize the deviation in predicted date-timestamp. It even captures slight deviation and penalizes the same. Level of performance would be MSE error ~1% deviation off the predicted date-timestamp

Minimum Viable Product (MVP)

Design What does your minimum viable product look like? Include sketches of your product.	**Predictive Maintenance Activity Flow Diagram**



Manufacturing company dashboard



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?

- Persona is the Maintenance engineers in Manufacturing company.
- They will be able to see a dashboard telling them which centrifugal pump is going to fail and needs attention and after what date-time duration.
- Users access it via a dashboard. And they provide additional feedback if the pump fails

	before the intended/predicted date-time. This acts as data drift. It is then fed back to training data.
Roll-out How will this be adopted? What does the go-to-market plan look like?	Meetings need to be set with Manufacturing companies where the use cases could be showcased/demoed. The go to market plan would be to highlight the predicted yearly monetary benefits derived from the number of failures repaired in time.

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?	In the long term I can use the deviation and loop the feedback back to the model. Real world data might be different in terms of weather conditions, environment factors since all this will change. The specification of the centrifugal pump might also change, and the quality of sensors used to capture data can change. The environmental factors can be accommodated by adding an additional factor of periodicity. The other changes will have to be manually added with time-> data set will change, features will change and hence modelling needs to be done on updated data set. A/B testing can be planned where post the first roll out predicted failure date-time stamp, it will be matched against the actual failure date-time. In case there is deviation from the ones predicted in the model, then the same should come as feedback before the actual model is rolled out in production.
Monitor Bias How do you plan to monitor or mitigate unwanted bias in your model?	• I will take the periodic factor and take it into account while training the data and also in case there are environmental changes I will incorporate those changes and retrain the model basis the changes in data and changes in

	parameters required with time. • Since in the model being proposed there is a risk of the model being trained towards predicting failures early and only capturing feedback in production when failure happens before the predicted date-time of failure. This risk is being mitigated in the A/B testing proposed where we are letting the pump fail and verifying the data against the predicted date-time of failure.