Capstone Project Proposal

Hussain Al-Balhareth

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 objective is to improve the effectiveness of maintenance of PV panels on top of large solar car park (~10MW) with minimal disruption to parking lots users. Using AI/ML is going to help analyzing hundreds of visual and infrared images taken from flying drones. This would switch the maintenance from planned or preventive to predictive which reduces the job (which involves man-hours and heavy equipment) and increases the customer satisfaction.
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 output electric energy production is the project cashflow the asset managers have to ensure proper availability and performance. Shading effect has a negative impact (hotspot) that could burn a cell or cells jeopardizing the performance or a string if not early detected. Also, sand factor (if beyond a normal threshold) could reduce the amount of solar irradiance which then reduces the generation. So, monitoring of both factors are better be automated and analyzed based on real case database to drive better business decisions, minimize failures and sand effect on energy production.
Application of ML/AI What precise task will you use ML/AI to accomplish? What business outcome or objective will you achieve?	AI/ML is proposed to be utilized to analyze hundreds of visual and thermal images taken from flying drones that are cheaper to use across large parking area (see figure below).

	 10 MW Solar Carpark

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.	The energy production in terms of performance and availability, maintenance cost, and cost avoidance are amongst some metrics used to measure the success. All of them are measurable through the metering station and maintenance reports. The baseline should be the business as usual prior to activating the AI/ML solution, say the average past three years cost and performance.

Data

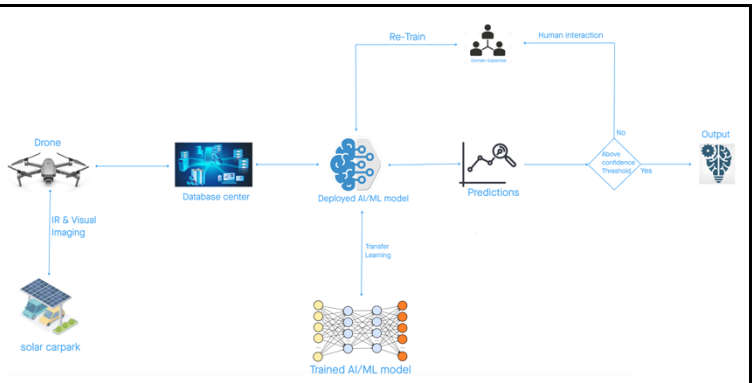
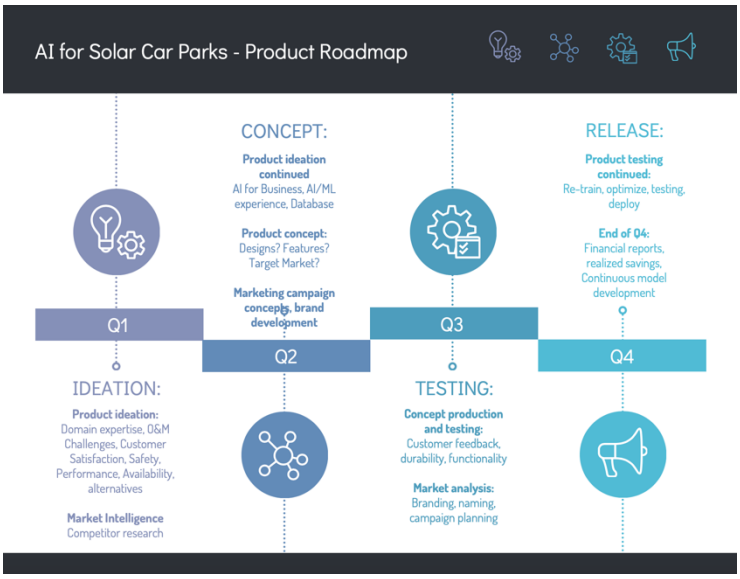
Data Acquisition Where will you source your data from? What is the cost to acquire these data? Are there any personally identifying information (PII) or data sensitivity issues you will need to overcome? Will data become available on an ongoing basis, or will you acquire a large batch of data that will need to be refreshed?	The solar carpark energy production is measured at a metering station that is linked to a smart monitoring center where both performance and availability reports could be automatically generated by the engineers. Hence the cost of acquiring such data is the man-hour spent on this task which is predictable and controlled with no PII or data sensitivity issue. This database is available and maintained within the company owned servers 24/7. However, the team can also buy additional infrared database with labeled data to help train our ML models while the future could be generated in-house.

Data Source Consider the size and source of your data; what biases are built into the data and how might the data be improved?	The source is most likely from academic and/or ongoing solar asset management businesses with a size not considerably large but rather decent, e.g. ~10,000 rows, which could be improved with in-house data from the drones with visual and infrared images.
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?	The labels are simple and straightforward such as: Is panel shaded? (0/1), sand factor (5-50%), cell temperature (°C), IR image patterns, hotspots detection (°C), physical condition (discoloration/cracks/corrosion), is it shaded by just leaves or worse bird droppings, etc.

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?	Considering lack of in-house AI/ML experience, the in-house domain expertise is still available. So, there are two options: AI/ML in-house training which could delay the project but useful in the medium term, or hiring external specialized services to build the model with expectations of being static so future training is not of high urgency.
Evaluating Results Which model performance metrics are appropriate to measure the success of your model? What level of performance is required?	Both precision and recall are key metrics for the success of any selected model to ensure proper performance of the solar carpark system. The conventional base case scenario vs. the proposed solution should help determine the acceptable levels. Using the proposed metrics, the team can measure the ongoing results compared to the base scenario. So, a significant increase in performance is required to justify the investment with decent IRR and ROI acceptable by the key stakeholders. We may test more than one model and select the optimum one upon agreed contract T&C.

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?	Designing for eexecutive management to make a business decision. The management is well-informed of current spending and challenges and expected to be interested. 
Roll-out How will this be adopted? What does the go-to-market plan look like?	With the massive number of panels installed and the cost and safety issues associated with the O&M, it's appealing for business to minimize the work and move to predictive maintenance rather than reactive which would improve the IRR and ROI as well.

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-house training of machine learning engineers to understand the algorithms and developing new models utilizing cloud services with computing capabilities to improve the model metrics. Also, training data should be relevant to the actual PV panels installed. A/B testing could be applied to compare new scenarios to baseline cases.
Monitor Bias How do you plan to monitor or mitigate unwanted bias in your model?	Model bias is expected to be caused by unbalanced dataset. Also, dirty data or special case data or irrelevant data could lead to biased or low performing models.