

Capstone Project Proposal



<Chidimma Dominic Nweke>

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?	Identification of Crop Diseases from Images of Crop Leaves. Visual estimation which is a primary and quickest method of identifying / detecting crop diseases based on characteristic plant disease symptoms and this method is always dependent on individual experience, subject to variation and just like other disease detecting methods, very time consuming. These downsides can be enhanced using machine learning systems.
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.	It is estimated that plant diseases constitute precisely 42% of losses in crop yield, hence the need for proper control measures on plant health. However, without proper diagnosis of the diseases and or its causal agent, disease control measures can be very ineffective and a waste of capital. Proper diagnosis on the other hand can be very expensive and time consuming, therefore, developing a means that can help cull these disadvantages will not only save money but also time.
Application of ML/AI What precise task will you use ML/AI to accomplish? What business outcome or objective	The task that I will use ML/AI to accomplish is plant disease detection / Identification. Albeit this software will not totally replace plant pathologists, the agricultural business or agriculturist that holds this technology will not only save money

will you achieve?	which goes into ordering expensive plant inspection and disease detection but also save time which it might take the pathologists effectively inspect and detect crop diseases. The business can also roll out the product to local farmers, gardeners and greenhouse farmers and make profit. Although expertise is important, anyone who can use the product can automatically identify crop diseases with an acceptable accuracy, culling the constant need of experts and saving money.
-------------------	---

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.	For the business metric to determine the success of the product, the following metrics are put in place: <ol style="list-style-type: none"> 1. By user feedback and traffic to the product (app and webapp/website) through SEO improvement, increasing processing speed and rating. 2. Product performance on MVP testing. 3. User retention rate. 4. Performance comparison with prior systems existing in farms. 5. By measuring product's baseline performance traffic with a user benchmark of five thousand within three months post-launch.
--	---

Data

Data Acquisition Where will you source your data	For starting, the data is sourced from camera and drone images of crops including the diseased ones from farms.
--	---

<p>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?</p>	<p>There are no personally identifying information (PII) or data sensitivity issues as the data extracted is just plant images with no additional information.</p> <p>It might also cost additional money and more time to acquire a more robust data as there are already existing datasets in this area.</p> <p>A large batch of data will be acquired and refreshed as more data are acquired and added.</p>
<p>Data Source</p> <p>Consider the size and source of your data; what biases are built into the data and how might the data be improved?</p>	<p>Based on the size and source of the data, the possible biases built into the data might arise from labelling which might either be introduced during annotation or by not having enough labels to capture all types of diseases affecting a particular crop. Also, diseases on genetically enhanced or crossbred plants might not be properly identified.</p> <p>The data can be improved by:</p> <ol style="list-style-type: none"> 1. Employing the services of expert annotators 2. Acquiring a more robust data on all possible diseases affecting a particular crop and also on genetically enhanced or crossbred crops.
<p>Choice of Data Labels</p> <p>What labels did you decide to add to your data? And why did you decide on these labels versus any other option?</p>	<p>The labels added to the data are dependent on the type of crop and its related diseases. For example:</p> <ol style="list-style-type: none"> 1. Apple – Healthy 2. Apple – (Some Disease) 3. Apple – (Some Disease) <p>These labels provide more clarity and narrows down the detection to a particular type.</p>

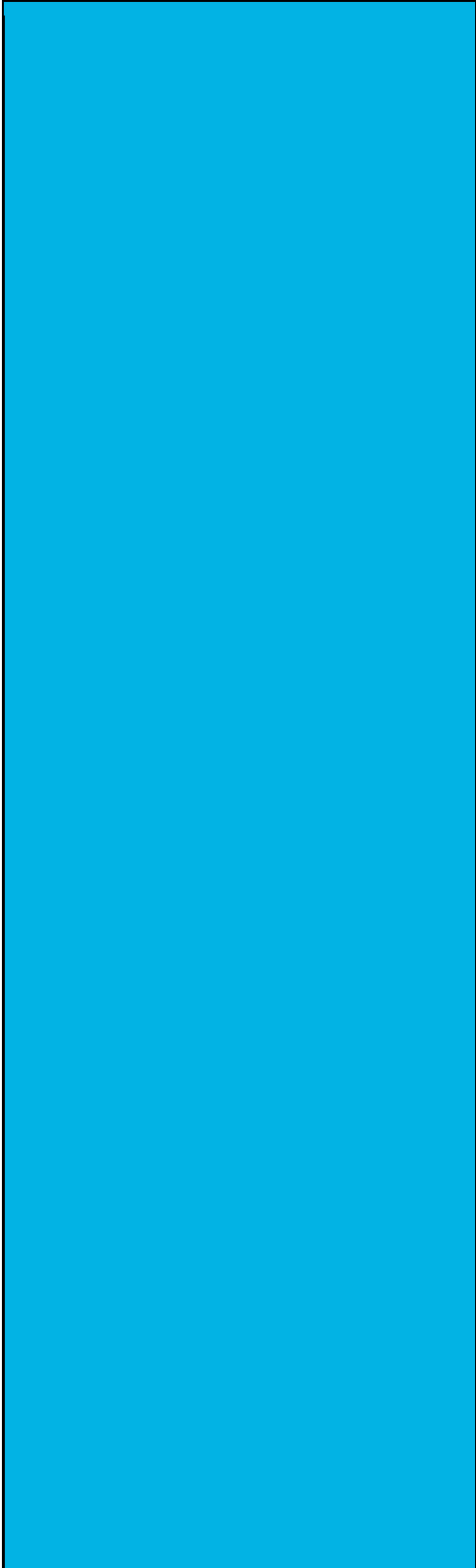
Model

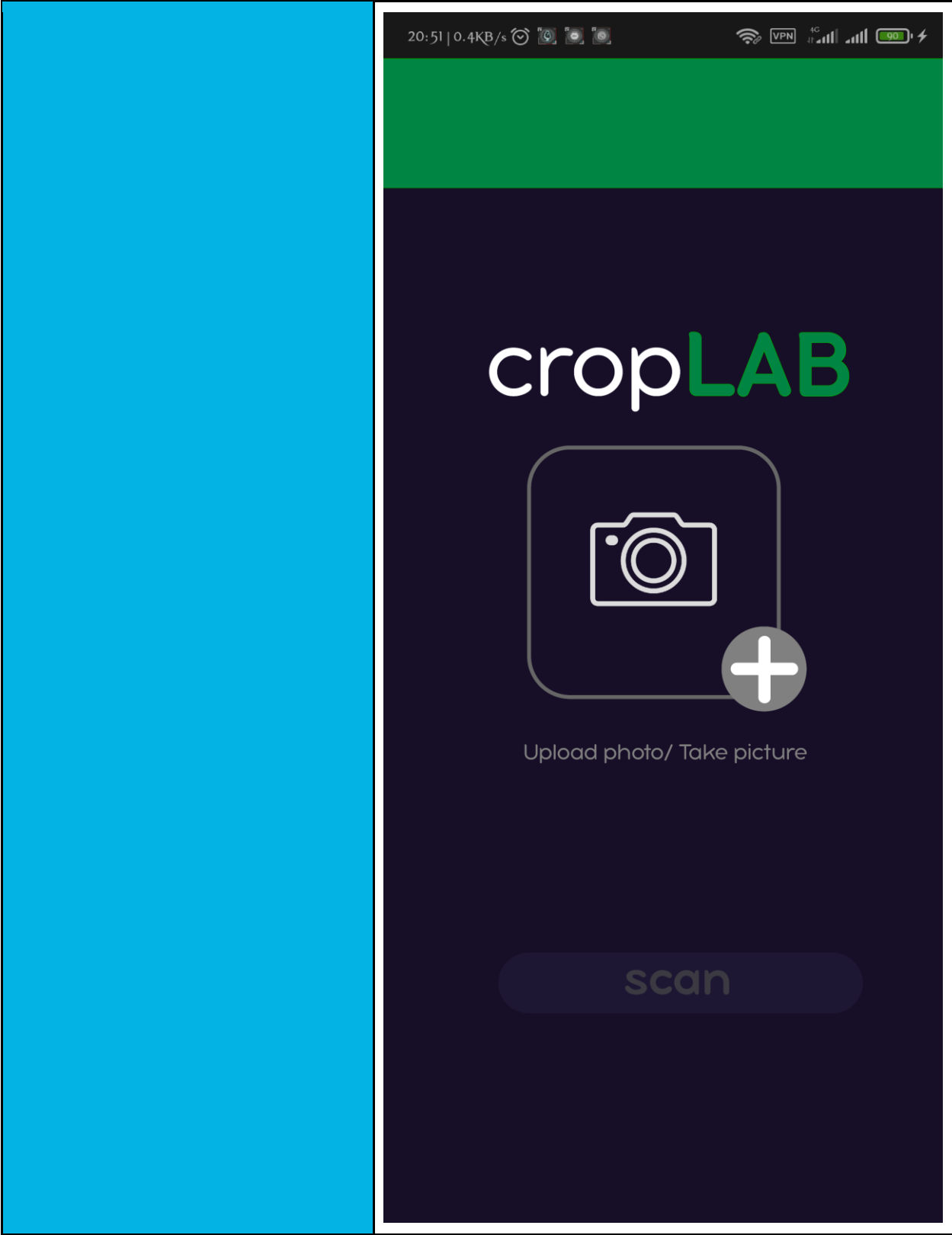
<p>Model Building</p> <p>How will you resource building the model that you need? Will</p>	<p>The model training will be outsourced and hosted to an external platform. This is mainly to optimize development time and cost as the model will be performing basic image classification.</p>
--	---

you outsource model training 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?	The performance metrics to be employed for the model are: Logarithmic Loss, Precision and Recall, with a performance level of 90%. As this model is proposed for disease detection, it needs as little false positives and false negatives as possible.

Minimum Viable Product (MVP)

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





	
Use Cases	The product is designed for both large- and small-scale farmers, including gardeners and greenhouse farmers.

<p>What persona are you designing for? Can you describe the major epic-level use cases your product addresses? How will users access this product?</p>	<p>Major Epic-Level Use Case: Supposed a gardener was inspecting his/her crops on a bright morning, s/he notices brown patches on one or several leaves of some crops. The product is used to take and scan the image of the leaf/leaves with the patch/patches and then the possible disease affecting the crop is detected.</p> <p>Users can access the product via mobile devices with camera as the product will be available as a mobile and web app.</p>
<p>Roll-out</p> <p>How will this be adopted? What does the go-to-market plan look like?</p>	<p>Since the product is an app, it can be downloaded from Google play store or IOS store, or be used directly from the web.</p> <p>The roll-out will have these steps:</p> <ol style="list-style-type: none"> 1. A team consisting of data scientists, machine learning engineers, plant pathologist and software engineers is assembled in a time frame of about two months. 2. With the help of an annotation company and the plant pathologists, the sourced data is labelled, this should take about a month. 3. Since the building and training of the model is going to outsourced and hosted on an external platform, it will take about two weeks to arrive at a model that meets the set benchmark before launching the MVP. 4. The product launch can happen in about six months. 5. As new data and or labels are generated and added, the model will be retrained.

Post-MVP-Deployment

<p>Designing for Longevity</p>	<p>Primarily, the product seeks to detect the disease affecting a plant. However, to improve the effectiveness of the product on the long-run, more data will be</p>
---------------------------------------	--

<p>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?</p>	<p>collected in addition to the existing ones and re-annotated to tailor the detection down to causal organism i.e., Bacterial, Fungal or Viral.</p> <p>There won't be any difference between the training and real-world data as both come from the same source – a camera.</p> <p>The product will use the extra features/examples the new data provides to improve and provide more accurate results.</p> <p>On addition of a new model, 80-20% would be used on the current and new model respectively, then the performance is evaluated and compared.</p>
<p>Monitor Bias</p> <p>How do you plan to monitor or mitigate unwanted bias in your model?</p>	<p>User feedback on the product will be used to both monitor and mitigate unwanted bias in the model.</p>