|  |  |
| --- | --- |
| Capstone Project Proposal |  |

*Noureddine Ettayyeby*

**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? | I want to optimize spraying herbicide and fertilizer in crops field so we can save on the costs of these two and have a better harvest in the end, also this will help the environment by reducing the usage of herbicide and the need to produce more of them which bad for the planet. I will use computer vision to train models that can recognize bad plant and spray them with herbicide and spray good plant with fertilizer. I will build a model using neural network using labeled images with labels like bad/good or planet name/other. This will help the business by decreasing the amount of money spend on fertilizer and herbicide. |
| **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. | Building this product will help the company reduce cost on fertilizer and herbicide which will increase revenue. This reduction of cost and increased revenue can be direct to develop better products for the costumers which will lead to a better costumer’s satisfactions which in turn can lead to a positive impact on the market share of the company. Also building this product will help the environment by reducing the use of herbicide which will help the planet and can lead to positive images for the company. |
| **Application of ML/AI**  What precise task will you use ML/AI to accomplish? What business outcome or objective will you achieve? | The task I will solve with ML/AI is recognizing bad plant and good plant so I can spray them with the right substance. The business outcomes are reducing the consumption of fertilizer and herbicide and the overall spending on productions. |

**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. | A business metrics we could use to measure success is how much we decreased spending on fertilizer and herbicide when we start using the products compared to previous years we could set a target of reducing our spending by 20% for example, that’s will be our main metric, But we also could take a look at revenue but here we have so many variables that cloud influence our revenue, Another way to see if this product have a positive impact on the company is to look at how the company is received on social media. |

**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 will be images of plants so we will have some data available for in our databases but I don’t think it will be sufficient since deep learning require usually a large amount of data to train good models so either will buy data from some other source if possible or will have to hire someone to take pictures from our fields and use a labeling platform like appen to label it. Other than buying the right to use the data if it’s not ours I don’t think will have a problem with PII. We will need at least a 1000 example to start experimenting with building the products then to get our model to a better performance we may need more data we will also need more data to maintain the model be retraining as its performance deteriorate over time. |
| **Data Source**  Consider the size and source of your data; what biases are built into the data and how might the data be improved? | We will need at least 1000 examples to start experimenting with building the model and we may need more depending on the performance of our model but usually deep learning models need a lot of data to perform well, since we are working in control environment the need to frequently retrain our model may not be necessary so may need more now and again  But not often. |
| **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 will use good plant versus bad plant labels since we are building a model to classifier whether a plant is good and need to be fertilized or is bad and need to be treat it.  This will make our model simpler because a two classes classifier is usually less of pain to train in deep learning. |

**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? | Since training a computer vision model usually requires a lot of compute power, we would use a cloud platform like Google cloud, azure or AWS for the training process because being able to provision resource when you need them and turn them off when you don’t will save cost for the company especially with the training process being an iterative. As for the team we will use our own team and maybe add some contractor depending on our need during the Development process. |
| **Evaluating Results**  Which model performance metrics are appropriate to measure the success of your model? What level of performance is required? | We will use a confusion matrix to evaluate our model because it’s a classification model. Using accuracy to get an overall estimation of how well our model is doing and digging deeper using recall and precision, our aim will be to get these three to be more then 90%. |

**Minimum Viable Product (MVP)**

|  |  |
| --- | --- |
| **Design**  What does your minimum viable product look like? Include sketches of your product. | The product will be a model which will be put on the spraying machines using a camera to take a picture of the plant send to the model for inference and depending on the prediction spray with fertilizer or herbicide. |
| **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 product addresses the problem of spraying vast fields with fertilizer or herbicide which inefficient and increase the cost of producing process for the agriculture industry, our user will be the engineer responsible making the machinery a way to make our model available for them is to put it in the cloud and give these engineers permissions to download it. |
| **Roll-out**  How will this be adopted? What does the go-to-market plan look like? | We will put the model on some machine and test how they do if it goes well will put the model on the rest of the machines if not will make the necessary adjustment and try again, also we need to test the roll-out against the current system we have on place. |

**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? | I would monitor the model to see how is doing in the real world. Collect more data and retrain if necessary and keep testing it against prior version using A/B testing to improve it and if the new model get better result than the old one, I will deploy it to production. Real world could be messy sometimes in this we expect that image may contain other objects so we should account for that and find ways to deal with it, so it won’t interfere with our predictions. |
| **Monitor Bias**  How do you plan to monitor or mitigate unwanted bias in your model? | We should look first in this instance for class imbalance mitigate that by under sampling over sampling techniques for example. And there is the problem of other objects presence in the images when the model is in production which cloud lead to problems. |