

# Capstone Project Proposal



*Vinícius da Silva Vale*

## 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 pokédex is a handheld electronic encyclopedia device created by Professor Oak as an invaluable tool to Trainers in the Pokémon world. It gives information about all Pokémon in the world that are contained in its database, although it differs in how it acquires and presents information over the different media.

Pokédex entries typically describe a Pokémon in only two or three sentences. They may give background information on the habitat or activities of a Pokémon in the wild or other information on the Pokémon's history or anatomy. Pokédex entries also include height, weight, cry, footprint, location, other forms, and a picture of the Pokémon. Completing the Pokédex is a common goal of Trainers and carries with it much esteem due to its difficulty. . In order to accomplish Professor Oak's goal of a complete Pokémon database, the Pokédex is designed to find and record data on each Pokémon the Trainer meets. However, detailed entries are not recorded until the player catches the Pokémon, receives it as a prize/gift or acquires it in a trade. (Source: [Bulbapedia](#))

But recently, Pokédex started to be reported by users as ineffective in detecting all types of Pokémon. This happened after the phenomenon called "SkyEgg". According to reports, several egg-shaped "portals" appeared in the sky around the world. (Image Source: [r/NatureIsFuckingLit](#)).



After that day, new creatures unrecognizable by pokedex and impossible to capture by pokeballs were reported. The first photographic record showed a glitch in the image, which when released by the media resulted in the current name given to the creatures: Digimon.



The objective of this project is to give the pokedex the ability to differentiate between pokemon and digimon, improving the experience of our consumer that when correctly identifying a digimon will not waste pokeballs on these creatures, this will result in regaining the user's trust and reducing complaints calls.

The first phase of the project was the collection of data of digimon. For this was used the pokédex image capture system. Whenever a user tries to obtain information about a digimon, the pokédex will return "??? There are no records about this pokemon", at that moment the image of the creature will be registered in our databases with the label "???". The second phase of the project was the labeling by Professor Oak if the image was a pokemon. Otherwise, it was labeled as Digimon. The result was 1066 different species of digimon. The problem is that this process took 3 months and new species are still appearing. And Professor Oak's time is very expensive and essential for other research. In this way, we will use the data obtained to find patterns and correctly classify new unidentified species. Currently pokédex uses AI to identify different pokemons. But the pokédex classification model does not recognize digimon. Therefore, we will use ML / AI to retrain the pokédex classification model.

### **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.

Pokédex is recognized as the pokemon encyclopedia, but the delay in solving the cases of "??? There are no records about this pokemon" for our researchers and trainers is resulting in a confidence shake, which has already resulted in increased complaints for our support and the emergence of independent wikis with incoherent information and fake news that create panic and ignorance in the population.

### **Application of ML/AI**

What precise task will you use ML/AI to accomplish? What business outcome or objective will you achieve?

ML / AI will be used in the classification between pokemon and digimon. The resulting model will be implemented in pokédex. The user will be informed that the creature is a Digimon and should not attempt to capture it. Resulting in fewer frustrations by the coaches and maintaining the confidence of pokédex as the main information center.

# 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.

Reduction in the opening of daily calls alleging defect and complaints by the user. And decrease in cases classified as "???" by pokedex for 1% of cases

# 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 source of the data was and will be the data collected by user's pokedex. There is no additional cost for data collection, as we constantly acquire user data for research centers such as Professor Oak's laboratory. All personal data of coaches currently undergo treatment to ensure privacy. Pokémon research centers are only interested in Pokémon data.

## Data Source

Consider the size and source of your data; what biases are built into the data and how might the data be improved?

There are currently 898 Pokémons and we have obtained 1066 images of Digimon, which we do not have a defined number of species, but we estimate in more than 850 different ones. A pokedex image base will be used from all Pokémon to which we will extract 1066 random images. This way, we will have a reliable, random and balanced source of data. The bias was reduced by the classification of a Pokémon specialist when classifying the Digimon. Still, there is a chance that Pokémons that have never before been seen being classified as Pokémons are classified as Digimon. To resolve these issues, feedback from trainers will be essential, since if a trainer manages to capture a fake Digimon with a pokeball, it must be reevaluated as a Pokémon and the model must be retrained.

## 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 Pokémon and Digimon. We use Digimon because it is the most accepted popular name. The dataset used to train the model is publicly available on the platform [Kaggle](#).<sup>[OBJ]</sup>

## 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?

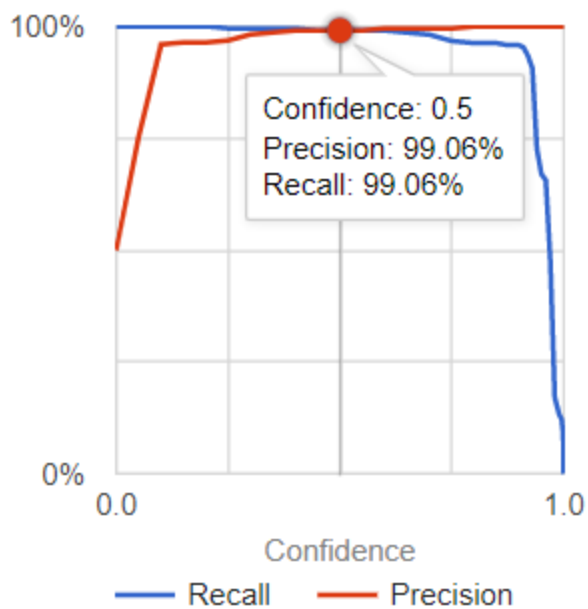
Due to the urgency, Google's AutoML Vision will be used to train and present the performance metrics of the model. The consumption of the model will be carried out via Rest Api, where the user sends the data of the creature to be analyzed, we send the json with the image link for the model to analyze, the return is handled by the Pokédex system that provides the classification and final information to the final user.

### Evaluating Results

Which model performance metrics are appropriate to measure the success of your model? What level of performance is required?

Precision and Recall was used as performance metrics to measure the success of the model. The Confidence threshold was 0.5. The results:

Total images	1,920
Test items	212
Precision ?	99.06%
Recall ?	99.06%



True Label	Predicted Label	digimon	pokemon
digimon	98%	2%	
pokemon	-	100%	

## Minimum Viable Product (MVP)

### Design

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

**Pokémon class:**



Digimon class:



## 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?

It's desing for Pokémon's trainers. The major epic-level use case is using a pokedex to obtain pokemon information and distinguish pokemon from digimon. They access this product via pokédex a handheld electronic encyclopedia device.

## Roll-out

How will this be adopted? What does the go-to-market plan look like?

The new version of the pokédex system will be adopted worldwide. The implementation plan will be to adopt the new version in some locations before making it available worldwide.



# 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?

As we still do not know all the species of Digimon and the success in identifying each special of this class of creatures has not been measured new versions of the necessary model. Before putting any new version of the model into production, A / B testing will be carried out as the region used candidate models while the increase will continue with the current model. The results in relation to the success metrics define the successor model.

## Monitor Bias

How do you plan to monitor or mitigate unwanted bias in your model?

As we have not yet discovered whether Digimon have sex or vary their physical attributes, such as size and color, our data may contain bias. Over time we will learn more about these creatures which will result in better classification models and even their possible capture. In the meantime we will collect data in order to monitor and better inform our users about the amazing Pokémon world.