

Capstone Project Proposal



Adalberto Gonzalez

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 challenge:

A soda bottler company (I will call it SodaCo) produces and dispatches thousands of products every day to stores to be consumed by people around the globe. In order for this to happen SodaCo follows several processes that involve human interaction in several steps.

One of those processes takes place inside their warehouses, we will call it, the Picking process, and goes like this:

- A list of products is handled to a worker (picker) who walks through the warehouse looking for each product listed in the purchase order.
- Identifies the product
- Takes the number of products required
- Place them in a cart/trolley
- take the cart to the ship area

This process presents a significant number of errors/losses in miscounted products.

Considering the numbers of pickers and orders per day some alternatives would be having someone who could check again however, the human error could be a possibility as well.

Why use ML/AI in solving this task?

Trough AI we believe it's possible to reduce the margin error and increase the control over all the carts in the warehouse. Today AI vision systems have proved to be more effective than a human in the product identification which is the core of this challenge.

<p>Business Case</p> <p>Why is this an important problem to solve?</p> <p>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.</p>	<p>By implementing this project, SodaCo will better track their products in their carts, thereby increasing the accuracy of the cart content over 95%. Therefore, reducing economic losses due to inventory discrepancy, order mismatch placing, customer claims and delivery time in the delivery chain.</p>
<p>Application of ML/AI</p> <p>What precise task will you use ML/AI to accomplish?</p> <p>What business outcome or objective will you achieve?</p>	<p>The core of this project will be the use of AI-Computer vision.</p> <p>Computer vision system will be placed at the carts to identify and count the total items in the cart in a dynamic way.</p> <p>By doing this the business outcome will be reducing the inventory mismatches and the direct impact in the economic losses.</p>

Success Metrics

<p>Success Metrics</p> <p>What business metrics will you apply to determine the success of your product?</p> <p>Good metrics are clearly defined and easily measurable.</p>	<p>The success metric of the product will be the correct detection of the products however, its expected that trough this project the following business outcomes can be impacted as well:</p> <ul style="list-style-type: none"> - Reducing client claims of mismatch orders by 20% based on their current reports. - Reducing warehouse inventory control (count) time by 90% per cart based on quality control times.
--	--

Specify how you will establish a baseline value to provide a point of comparison.

- Maintain average speed of the process at an average of 20 min per cart. The client expects the speed of the system to be at least as fast as the human validation.

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 data will be from the SodaCo company and we will get help from third parties that will work with the warehouse staff and us to photograph and label the products.

We expect to label the 85 different products with 200 pictures per product as an initial batch of images*.

The estimated cost is expected to be **2,571 USD****. And 8 days are expected to complete the task.

The cost is described as it follows:

Photography

The cost of getting the images will be approximately 1,000USD and it will take 3 days to acquire.

Labeling

- Project management: \$500
- Cost per 1000 image: \$63
- Objective: Bounding box

The cost will be 1,571USD and it will take around 5 days

Copyright

All the products are property of the SodaCo and they will grant the permission to get those images.

*Depending of the performance of the AI model we might require additional images.

**Final cost will be on actuals.

Data Source

Consider the size and source of your data; what biases are built

Biases

-Seasonal: Since this is a products company, they usually create versions of the products per season. We aim to work with the marketing team to have that

<p>into the data and how might the data be improved?</p>	<p>information in advance so we could prepare for those cases.</p> <ul style="list-style-type: none"> - Light: some warehouses have natural light and that could create unwanted noise, we would have either to create a larger data set with different light conditions or simulate those ones with data augmentation techniques.
<p>Choice of Data Labels</p> <p>What labels did you decide to add to your data?</p> <p>And why did you decide on these labels versus any other option?</p>	<p>Our labels will be associated through a dictionary, so each product will have a code that can be referenced to their commercial name.</p> <p>The code will be the barcode which follows this structure:</p> <pre>0000000000000 [commercial name] – [type] – [flavor] - [size]</pre> <p>In order to manage our images, they will be labeled and stored in folders as it follows:</p> <pre>[barcode_dir] --- [barcode]_[image_number]</pre> <p>Example:</p> <pre>--000000000000 --- 000000000000_0001 --- 000000000000_0002 --- 000000000000_0003 ... --- 000000000000_0004 --000000000001 --- 000000000001_0001 --- 000000000001_0002 --- 000000000001_0003 ... --- nnnnnnnnnnnn_yyyy</pre> <p>This way we will have each product with the presentation and number of images available.</p> <p>Using barcode is a very efficient alternative since its already a unique identifier and can be used to integrate the data to the inventory system.</p>

Model

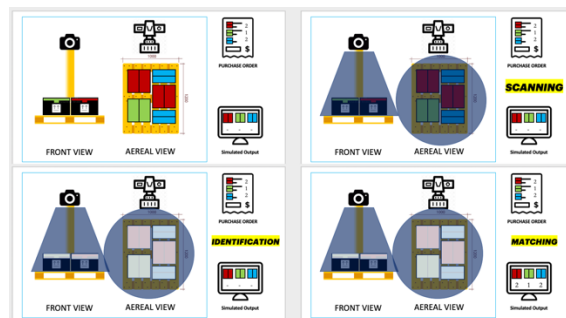
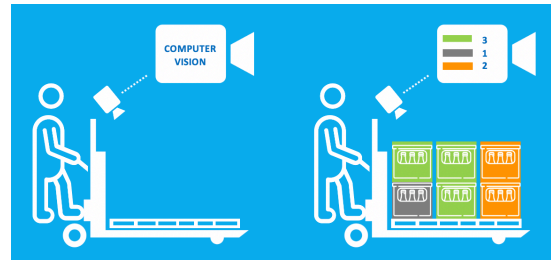
Model Building How will you resource building the model that you need? Will you outsource model training and/or hosting to an external platform, or will you build the model using an in-house team, and why?	<p>The project will require resources from different areas</p> <p>Human side:</p> <ul style="list-style-type: none">i) Clientii) Third partyiii) Developer team <ul style="list-style-type: none">- Product owner- Designer- AI engineer- Data science- Quality assurance- Devops <p>Equipment side:</p> <ul style="list-style-type: none">- Computers- Storage- API- Libraries <p>For the model the AI engineers will be the key role here, the model will start using available API and services that could complete the project as google cloud image service also the storage, if the results are not what we expect we would advance to use AI libraries like TensorFlow keeping the storage in the cloud service.</p>
Evaluating Results Which model performance metrics are appropriate to measure the success of your model? What level of performance is required?	<p>We will test several models to get the following metrics:</p> <ul style="list-style-type: none">- Accuracy- Precision- Recall <p>And from there we will compare them using the:</p> <ul style="list-style-type: none">- F1 score <p>Additional we will include:</p> <ul style="list-style-type: none">- Running time <p>The accuracy will have to be as close to 100% however if that number can't be reached, the difference has to provide a sufficient improvement margin to be feasible for the business.</p>

Minimum Viable Product (MVP)

Design


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

Our initial design includes equipping the cart/trolley with a camera that will be pointing to the cart from above. That camera will input the images so the model can identify and count the products placed in the cart.



The output will be a plain text file with the number of products identified with time stamp.

```
"order": "order-id",  
  
  "datasetId": "barcode-id",  
  
  "total": "3",  
  
  "updateTime": "2019-10-30T20:54:50.472328Z",  
  
  "datasetId": "barcode-id",  
  
  "total": "1",  
  
  "updateTime": "2019-10-30T20:54:50.472328Z",  
  
  "datasetId": "barcode-id",  
  
  "total": "2",  
  
  "updateTime": "2019-10-30T20:54:50.472328Z",
```

<p>Use Cases</p> <p>What persona are you designing for?</p> <p>Can you describe the major epic-level use cases your product addresses?</p> <p>How will users access this product?</p>	<p>We have two users from the input standpoint and the output.</p> <p>The input will be used by the picker who is moving the cart around the warehouse placing the items in the cart.</p> <p>The output will be consumed by the admin team who will use that output to manage the inventory and related process.</p> <p>Usage, The common process will be as followed</p> <ul style="list-style-type: none"> - The picker will have a purchase order - The system will be active when the picker takes the cart. - As the picker fills the cart the system will identify and count. 
<p>Roll-out</p> <p>How will this be adopted? What does the go-to-market plan look like?</p>	<p>Initial Milestones (in weeks):</p> <p>01-06 – Proof of concept with 25 products</p> <p>07-10 – Controlled test in one warehouse and with a family of products.</p> <p>Based on that information we will elaborate a plan to deploy the system to the regional warehouses and so on until we reach global deployment.</p>

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?	<p>The improving strategy will have a strong dependency on the business challenge we aim to solve. Regardless of that, we have to create a strategy to refresh the models based on new products and maybe escalate the project to different areas where could become an ally.</p> <p>The difference from real data vs training data should be minimal, we aim to get data in real circumstances. However, once the project is released we have to keep supervision and identifying opportunity areas for future improvements and releases.</p> <p>Part of the strategy is releasing the new features to a percentage of the warehouses. As we advance, we release the full version to a small segment and iterate as we gather feedback; we plan on introducing the system updates 20% of our users while 80% still use the system without updates. Our key criteria will be business impact.</p>
Monitor Bias How do you plan to monitor or mitigate unwanted bias in your model?	<p>Data bias will be addressed with a continuous strategy to refresh data, also applying corrective and preventive maintenance to the cameras and equipment to avoid noise like dust and degradation by usage.</p> <p>Other types of bias have to be addressed based on frequency and impact, keeping constant supervision of system performance.</p>