Project Experience Report

ARC - Artificial Recognition of Cannabis

ENSE 400/477 2022/23

Gregory Sveinbjornson

Feras Daghmoush

Introduction

The contents of this report includes the overall experience of our 4th year engineering capstone project. Our group designed a system in which given an image of cannabis packaging, the user would be returned all the relevant product information. Our project was created in collaboration with two companies, BudSense and Farmer Jane Cannabis. This report will cover the entirety of the project, from opening and planning through to closing.

Group Members/Roles

Greg Sveinbjornson

Roles - Team Lead, Back End Engineer, Front End Engineer, Data Collection

Feras Daghmoush

Roles - Machine Learning Engineer, Front End Engineer, Documentation

Problem Definition

When you walk into a cannabis store in Canada, you will always see a menu board similar to this:

Sativa	THC	CBD	1g	3.5 g	7 g
Headbanger Delta 9	22%	<1%	–	$32⁹⁹	$65⁹⁹
Chocolate Fondue DNA Genetics	21%	<1%	–	$37⁹⁹	–
Ultra Sour Namaste	18%	<1%	$11⁹⁹	$39⁹⁹	–
Lemon Z OGEN	19.8%	<1%	–	$38⁹⁹	–
OS Sativa Original Stash	20.3%	<1%	–	–	$57⁹⁹
Gather (Jack Herer) Solei LOW STOCK	17%	<1%	–	$33⁹⁹	$61⁹⁹
Blue Dream Spinach	22.2%	<1%	–	$35⁹⁹	–
Donegal TWD	21%	<1%	–	$34⁹⁹	–

Hybrid	THC	CBD	3.5 g	7 g
Electric Punch Delta 9	16.6%	<1%	$24⁹⁹	–
Grapefruit GG1 Emerald Health	17.9%	<1%	$32⁹⁹	–
Dealers Pick Good Supply	19%	<1%	$32⁹⁹	–
Diesel Spinach SALE	20%	<1%	$33⁹⁹	$66⁹⁹
White Widow Spinach SALE	19.2%	<1%	$33⁹⁹	$66⁹⁹
Unicorn Poop Tantalus	20.7%	<1%	$44⁹⁹	–

FARMER JANE
CANNABIS CO.

This is because under Canadian law, cannabis stores are not allowed to have products out on shelves for customers to buy. Products must be kept in a back room in a locked vault for safekeeping. Because of this, customers cannot see the products in stock, and must ask an employee to unlock the safe and get the

product they want. Menu boards are important for the shopping experience for customers.

Two of the most important indicators in a cannabis product are the THC and CBD percentages. THC and CBD determine the potency of the product, similar to alcohol percentage in alcoholic beverages. However, because cannabis is a living product, the THC and CBD percentages change from batch to batch. This means that identical products can have differences of up to 10%. This is a substantial difference that can go unnoticed by the consumer, as differences of this magnitude can alter a user's experience.

So far we have determined that;

1. Menu boards in cannabis shops are important for customers
2. THC and CBD change from batch to batch up to 10%

This leads us into the true problem this project aims to solve. As the THC and CBD percentages change with each new batch, the menu boards have to be updated to reflect the new values. But because of a combination of:

1. Rapidly changing values,
2. The volume of products,
3. The complexity and time to manually change the boards,

the THC and CBD values on menu boards are often incorrect.

To solve this problem, cannabis store managers have requested a system to update the menu boards as simply as possible, without having to leave the vault every time to update them. This is how, together with BudSense, we came up with the solution of using machine learning to gather the information needed with just a simple picture of the product.

One common question we received in response to our solution was why we didn't use the existing barcode infrastructure to change the menu boards. The reason we couldn't do this is that the Universal Product Code (UPC) only contains the product name and price, not the sufficient product information needed. This is not something that we can change, as all the labelling in the cannabis industry is heavily regulated by the Canadian Government, and not changeable by individual stores.

Golden Circle

Who

Stakeholders:

Team Members - Greg Sveinbjornson and Feras Daghmoush

BudSense Team - Andrew Cretin, Kevin Baker, and Paul Hewitt

Farmer Jane Cannabis - User Testers

University of Regina - Tim Maciag and Trevor Douglas

Target Audience:

The employees and managers of cannabis stores.

What

A web app to collect all the relevant product information from a picture of cannabis packaging.

When

September 2022 - April 2023 and beyond

Where

University of Regina and the Cultivator in Regina for development, and in any cannabis store that uses BudSense for deployment.

Why

To solve the issue of incorrect product information on cannabis store menus.

How

Using machine learning techniques, such as image classification and optical character recognition.

Project Planning

Project Management

To manage our project, we used several different tools. We wanted to stay in constant communication with the team at BudSense, so we opened a Slack channel dedicated to this project. In this slack channel, we discussed meeting times, bugs in code, presentation tips, and several other topics. This channel was at its most useful when BudSense was helping us in creating the user interface. Because our project was intended to be used with their existing systems, we needed the help of their team to build the framework that would best work with their existing infrastructure. The Slack channel allowed for fast communication and quick fixes.

Another project management tool we used was Trello. Our Trello board was used to track questions, tasks to do, our current work, bugs, bug fixes, and completed tasks. We also had a message board section for things to remember. Our workflow was as follows for tasks:

1. A list of tasks were placed on the to do column
2. Each group member selected a task to work on
3. Any bugs were reported
4. Once fixed, bugs were moved to bug fixes
5. The task was moved to completed tasks

6. Repeat

This workflow allowed us to constantly make progress while ensuring the other team member knew what was going on.

The final two tools used for project management were Google Colab and Github. We used Google Colab for machine learning as it provided two major benefits:

1. We could use Google's faster GPUs, drastically reducing training time.

2. We could simultaneously work on the machine learning code file

As Feras was our machine learning engineer, most of his work was done through Google Colab, then Greg would check for completeness and as the team lead would add it to our Github, which is why Greg has the majority of the GitHub commits. Github was used to track our progress and revision history, as well as a safety net in case our code was lost or we had to revert to a previous version. Github was also used as a display page for our project, including a working demo, a commercial, and all of our vlogs. Our Github was made as readable as possible to help anyone who read it to have a clear understanding of what our project does and how to replicate it.

Tech Stack

Our tech stack has 3 categories:

1. Front End

    For the user interface of our project we used Typescript in the Angular framework. This is because BudSense's infrastructure is in Angular, and this way our project could be easily imported into their systems.

2. Back End

    For the back end of our project, we had our machine learning code running on a Flask API in Python. Our API was running in the cloud on an AWS EC2 instance. Our back end also utilised Tesseract-OCR to read text from the images. Tesseract-OCR is the open-source version of Google's OCR.

3. Machine Learning

    For our custom machine learning model to classify cannabis packaging, we used the Python libraries Tensorflow and Keras in a Jupyter Notebook file to train our model.

Initial Timeline

This was our initial timeline made with BudSense, although we modified it to add more MVP's and never used AWS Sagemaker as it was too expensive, we did meet all the deadlines.

- **Project Tooling Configuration** - Oct 15, 2022
  - Setup Trello/Jira to manage task tracking
  - Create 1mo sprints for line of sight on tasks and to make reporting at the end of capstone easier.
    - See here for summary of Scrum Sprints
- **Project plan** - Oct 31, 2022
  - Build out documentation around project as required for Capstone
- **Data collection** - Nov 15, 2022
  - Begin collecting of images of cannabis product packages
  - Initial data set should be of Flower products
    - Can be extended to other product types if data availability is there
    - Primarily Flower products (and flower derivatives: ie: Pre-Roll) will be the only ones with consistently changing product attributes. Edibles, capsules, oils etc. will be lab manufactured and consistent from batch to batch
    - Want minimum of 1000 images to begin training the model
- **Begin Building ML model** - Dec 15, 2022
  - Use Git repo for project source control
  - Use Tensorflow platform and Python language
- **Milestone 1** - MVP 1 - Dec 24, 2022
- **Data validation, retraining, modifications** - Jan 1, 2023
  - Take findings from initial build, refine and retrain
  - Ongoing process but should have a first iteration of this complete by this date
- **Feedback loop for user validation** - Jan 15, 2023
  - Allow for human validation to further train the model and gain accuracy
- **Deploy model to AWS SageMaker** - Jan 31, 2023
  - Get the trained ML model into the AWS framework

- ○ All sorts of how-to docs on this, here, here and here, and more.
- **API to upload image and return results from ML** - Feb 28, 2023
  - ○ BudSense team to help build out API to interface with ML model in Sagemaker
- **Web app for capturing images / user validation** - Mar 30, 2023
  - ○ BudSense team to help build out Angular web app to capture/upload photo and provide real-time validation
  - ○ End results would be pushed into BudSense product database
- **MVP 2 - Final Product & Presentation** - April 1, 2023
- **Future State - Gamification** - TBD
  - ○ Gamification in web app to encourage Budtenders to import new product information
  - ○ Additional rewards if the item is new to the system, less if the incoming data has already been captured

Division of Tasks

We defined our roles in this project based on our preferences and by what we wanted to learn. Greg wanted to learn how to make an API so he was given the back end responsibilities. Feras was more comfortable with machine learning so he was given the machine learning tasks. The reasons were similar for the other tasks, except for the front end as we both wanted a say in the user interface.

These responsibilities were not stuck to absolutely. Thanks to our Trelloe board, we could always see what the other team member was up to and if they were ahead or behind schedule. Because of this, we would often work on each other's responsibilities depending on who was ahead or behind. However, the final say

on decisions was up to the person responsible for that section, and in the case of the front end the final say was given to the team lead.

Project Execution

This section will describe our progress at the specific dates of our class scrums.

Oct. 4, 2022

This was our first scrum after we decided on our project. At this point we had created our initial timeline with BudSense as well as started on our opening documentation. At this point we were in the yellow, as we had concerns about data collection. Feedback from Tim suggested that we add more MVP's

Oct. 18, 2022

At this point all of our initialization documentation was completed. We had made a demo machine learning application to ensure we could get our tech stack up and running. We were still in the yellow, as we knew how we were collecting data but were not sure how it would go. Feedback from Tim suggested our MVP's were ambitious and that we should reach out to our mentor Trevor Douglas.

Nov. 29, 2022

By the end of November, we had all of our data collected and our first version of our machine learning model in progress. At this point we felt as though we were in the Green, as our data collection had gone excellent and progress was strong. Our concern moving forward was computing power as it was taking 2 hours each time we trained our model.

Jan. 13, 2023

This time was a definite low point in our project. We had made nearly no progress on our machine learning over the break. We were lacking a high fidelity prototype or any user interface. All our time was being consumed by machine learning issues. At this time our target audience was not fully fleshed out, and would have to be straightened out with BudSens. Feedback from other students had lots of constructive criticism that helped us move forward and let us know what to focus on. We were in the yellow, and were closer to red than green. We knew we had a lot of work still to do.

Feb. 3, 2023

By this scrum we had solved several major issues. Our AI model was now working thanks to our new use of Google Colab. We had created a high-fidelity prototype that received lots of constructive criticism. Which ultimately resulted in us changing our design philosophy to KISS - Keep it Simple, Stupid. Our

API was up and running on an AWS EC2 instance.We had still yet to create our user interface, and although others were concerned, we felt confident that it could be done quickly over the February break with the help of BudSense. At this point we were still in the yellow but closer to green. Tim gave us feedback about straightening out our target user, which would be done the next day with BudSense.

Mar. 10, 2023

This scrum was when our project started to come together. We had fleshed out the target audience with BudSense and created a clean user interface. Our AI model had been updated to better adapt to new images and was now at 95% accuracy. At this point we were firmly in the green, the only thing left to do for our programming was to improve the user experience by making our project easier to use. Feedback from Tim said that things were looking good for project day. Our main upcoming work was preparing our presentation and closing documentation.

Apr. 1, 2023 - Project Day!

Eight months later and we were finally at Project Day! We were still firmly in the green with all our programming work done, our presentation prepared and our closing documentation nearly complete. We got good feedback from

professors during our presentation, who were pleased that we worked well with a real company trying to solve a real issue.

Project Reflection

What Went Well

We believe that our project ended in a good spot. We still have work with BudSense to build it out for industry, but the tasks as stated in the beginning of the project have all been finished. We felt that our presentation, poster, commercial, and tradeshow all went well. As well our progress from January to March was extremely efficient, as mid-January we were concerned about finishing our project.

What Could Have Gone Better

Much of our time was wasted figuring out two things: machine learning and who our target audience was. We lost nearly a month trying to train our machine learning model on our own computers. If we had known about Google Colab's GPUs earlier, we could have been much less stressed for the final months of this project. As for our target audience, we struggled to determine exactly why we were doing this with BudSense. BudSense was awesome to work with, but in this category it sometimes became more of a hassle as they are in a constantly changing industry that neither of us understood the complexities of.

If We Could Do it Again

In hindsight, we should have hammered out exactly who the target audience and what the use case was in the very first meeting. We would also use a different approach to our machine learning struggles, as we could have been still building out our API and user interface instead of running head first into a problem we hadn't figured out yet.

What We Learned

Both of us learned more about cannabis and cannabis stores than we ever thought we would know. We both learned about machine learning and the difficulties of managing large datasets. We both gained a deeper understanding of Github by using it for months on a single project. We both now have some experience with Angular, which is a popular framework in industry right now. We learned how to make an API, as well as researched all the security concerns that come with them.

The final thing we learned was how to work with a company. Neither of us had any co-op experience, so being able to work with a real company was extremely valuable in experience but also in building relationships with people doing the career that we want to learn.

Lifelong Learning

We plan on continuing this project with BudSense once we negotiate the exact terms under which they will continue. We don't know how to do much of the envisioned work to continue this project. But learning is not something only done in school. To have success in not only this project, but in our careers as software engineers, we must be constantly learning to conquer new challenges or we will soon be left behind in an industry that is always changing. It is impossible to know everything, but that won't stop us from trying.