Springboard Data Science Track – Capstone Project 2: Proposal

1. What is the problem you want to solve?

2. Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn't have otherwise?

3. What data are you going to use for this? How will you acquire this data?

4. In brief, outline your approach to solving this problem (knowing that this might change later).

5. What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

1. What is the problem you want to solve?

On a recent morning walk at a park, I came across a friendly husky dog. She was well-groomed and full of energy. But with neither a collar nor another person nearby, I presumed she was also lost. I also remembered that the Lunar New Year was the past weekend and fireworks are known to frighten dogs to the point where they tend to flee based on instinct. With some quick thinking and providing as much detail to the best of my knowledge to a Facebook post, I was able to reunite this lost dog with their owner in under 3 hours.

However, I couldn't help but wonder how many dogs and owners had not been as lucky. How many tragedies could have avoided and how much manpower could have been saved if only a passerby identified a lost dog situation and could correctly identify the dog's breed? What if there was a way to empower a clueless stranger to reunite with a lost dog using only by capturing a photo?

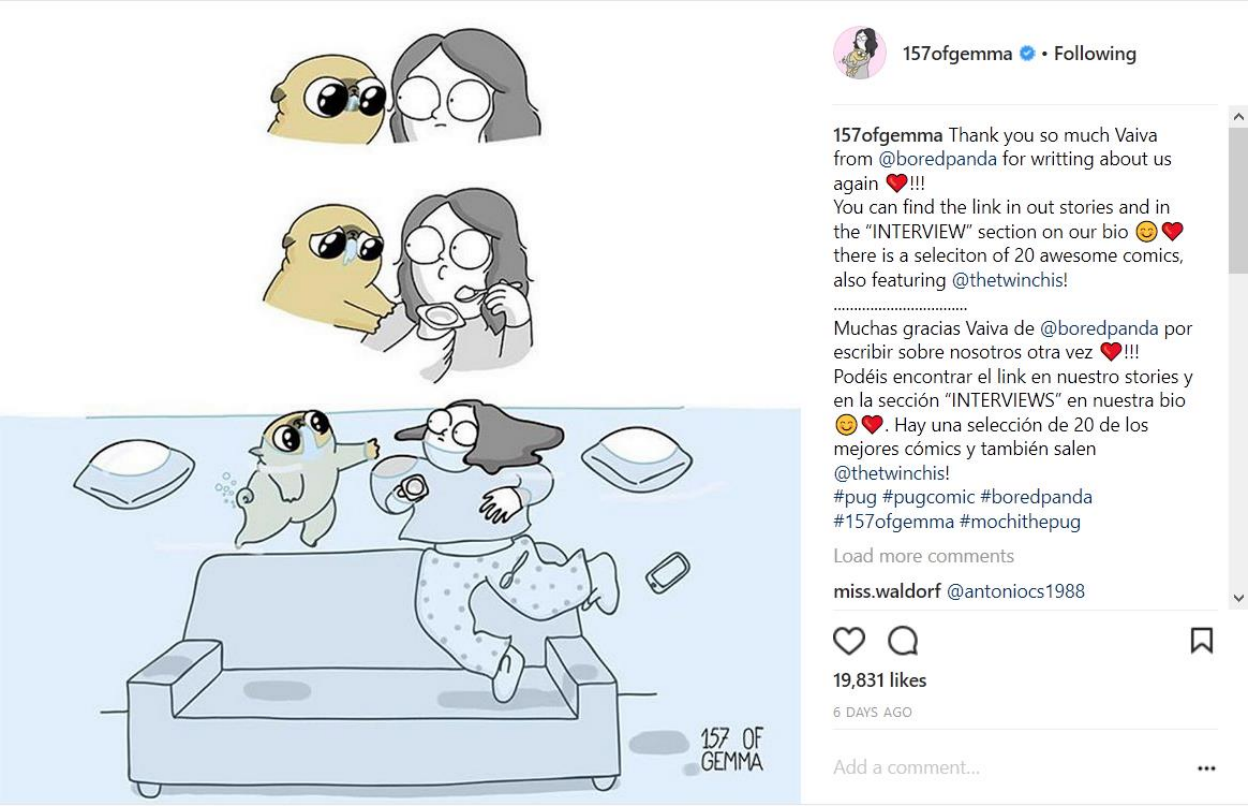
This hypothetical situation is the motivation for the problem I wish to solve. That is, I want to create a product/model which can identify and classify specific dog breeds better than a human. At 167 breeds recognized by the American Kennel club, even canine enthusiasts could struggle to correctly identify from an image.

2. Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn't have otherwise?

I can identify 3 main categories for clients who would be interested in this dog breed classification model/product. The first would be dog owners trying to locate a lost dog or strangers helping to reunite a lost dog with their owner. I envision the former client querying a website or social media with this model to find matching posts or pictures of their lost dog. I envision the latter situation much like the one described previously.

Another category would be for clients of marketing efforts such as pet food/toy brands, lifestyle brands and social media influences. For lifestyle brands, perhaps a brand is trying to market a new product to a demographic associated with ownership of certain dog breeds. In the last influencer category, they might be interested in reaching new audience members that may not have explicitly expressed interest

in certain breeds (via likes and followed accounts) but may have pictures of their pup on their Instagram? For example, if 157ofGemma is a Spanish-language Instagram Pug from Barcelona how could Instagram suggestion US-based Pug fanatics to follow?



@157ofgemma- 184K followers. Despite this account being in Spanish and based in Barcelona, much of the content isn't bounded by language. (By the way, she illustrates her comics in both Spanish and English).

Lastly, another category could be social media agencies which specialize in high-profile pet accounts who may be seeking out new clients.

3. What data are you going to use for this? How will you acquire this data?

I will primarily be using data compiled by Stanford Researchers from an online source, ImageNet. The data is hosted at http://vision.stanford.edu/aditya86/ImageNetDogs/. The data contains labeled data which consists of ~150 images of around 120 different dog breeds.

I may also supplement this dataset with photos from social media where dogs have different outfits may have articles of human clothing on (such as Dog the Pug or Menswear Dog).

4. In brief, outline your approach to solving this problem (knowing that this might change later).

I intend to utilize a Convolution Neural Network (ideal for image classification) and optimize for image classification of dog breeds (with the breeds as output classes). I also may choose to create an General Adversarial Network in which one NN will classify image inputs to an appropriate dog breed while the other NN will generate images from the N-Dimensional vectors generated by the first NN.

5. What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

For this project, I intend to write a research-style paper which outlines the process and annotates the optimizing/model-tuning steps along with a slide deck. If time allows, I also wish to explore Python libraries that enable me to productize and test on new unseen data or create an app.

@Mensweardog – 350K followers