My Skincare Assistant

about me

Quyen - 3 years in retail – 5 years in spa business.

I am pursuing Machine Learning thanks to Coder School.

Working in retail has bothered me with problems of over manufacturing thus over consuming which are among other factors that lead to enviromental issues.

For this reason, and my skincare knowledge gained from running spas, is why I decided to work on MySA for my Machine Learning final project.

about MySA

PROBLEM

Skincare has become indespensible to the majority of women and LGBT. There is no doubt the demand is only up for the rest of the population.

Yet almost no one really have the knowledge and patience to learn about skincare ingredients.

MySA – My Skincare Assistant

She receives photo of skincare product's front label and returns explanation for each ingredient.

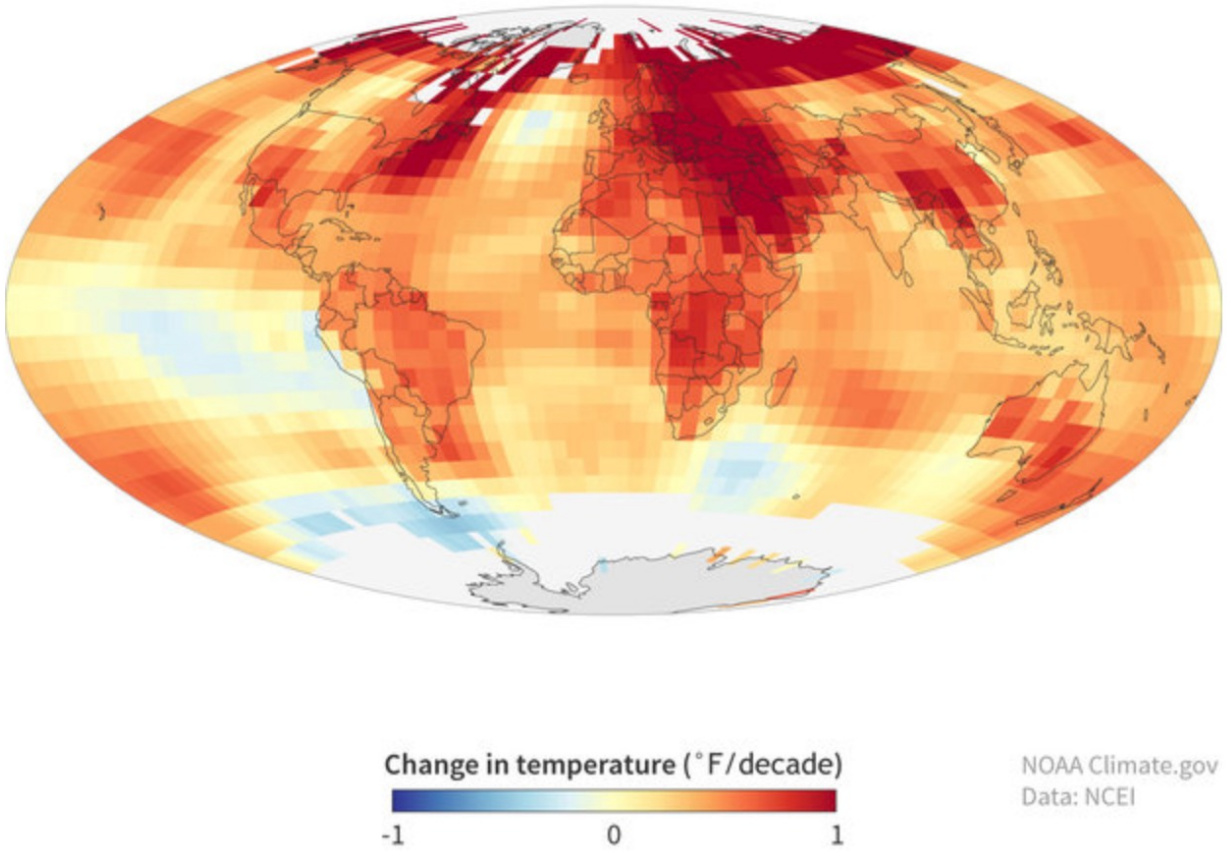
She is informative in order to save your time and money.

Finally, MySA promotes the habit:

BUY ONLY WHAT YOU REALLY NEED

Github

environment facts

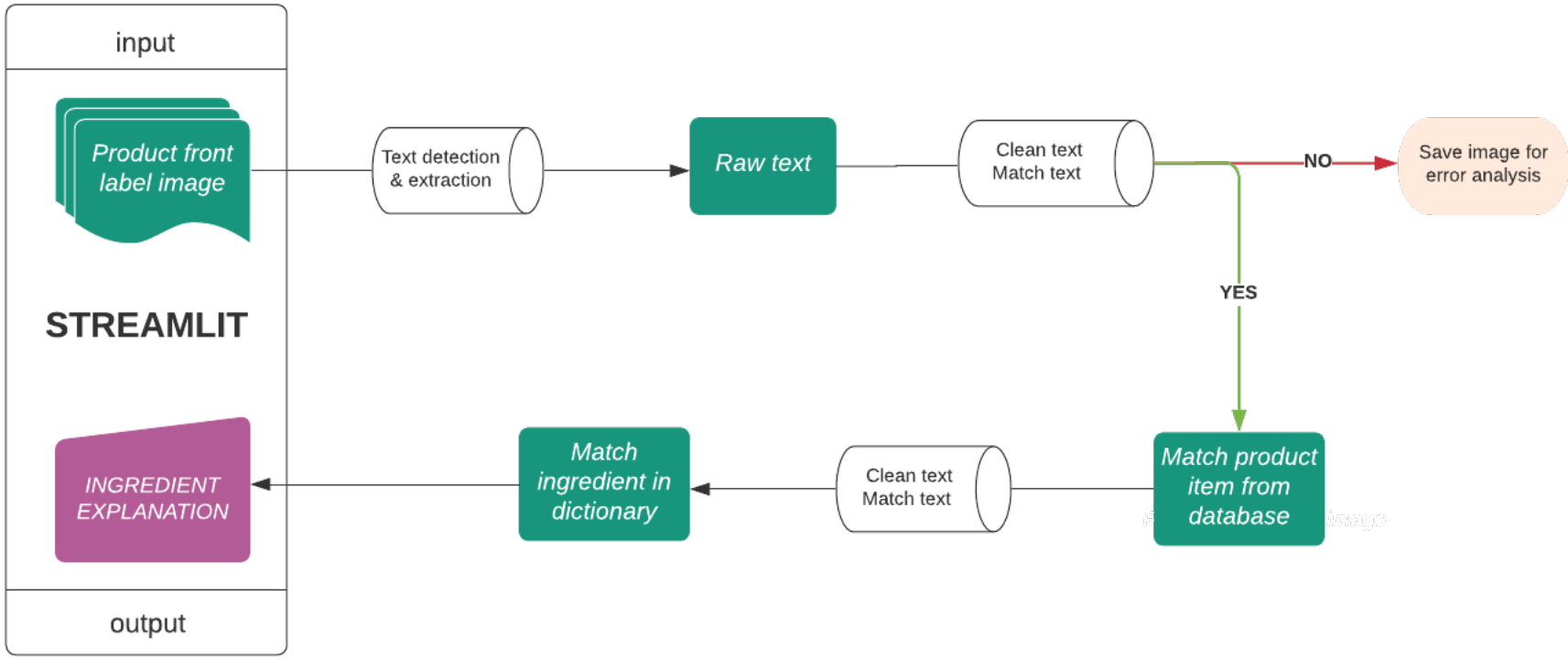


RECENT TEMPERATURE TRENDS 1990 - 2020

TOP CAUSES

- Power plants
- Transportation
- Farming, fertilizers
- Deforestation
- Permafost melting
- **GARGABE**

framework

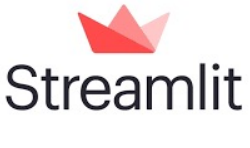

input
Product front label image
STREAMLIT
INGREDIENT EXPLANATION
output
Text detection & extraction
Raw text
Clean text Match text
NO
Save image for error analysis
YES
Match product item from database
Clean text Match text
Match ingredient in dictionary


Se


OpenCV


pandas


Google Vision API


Streamlit

dataset

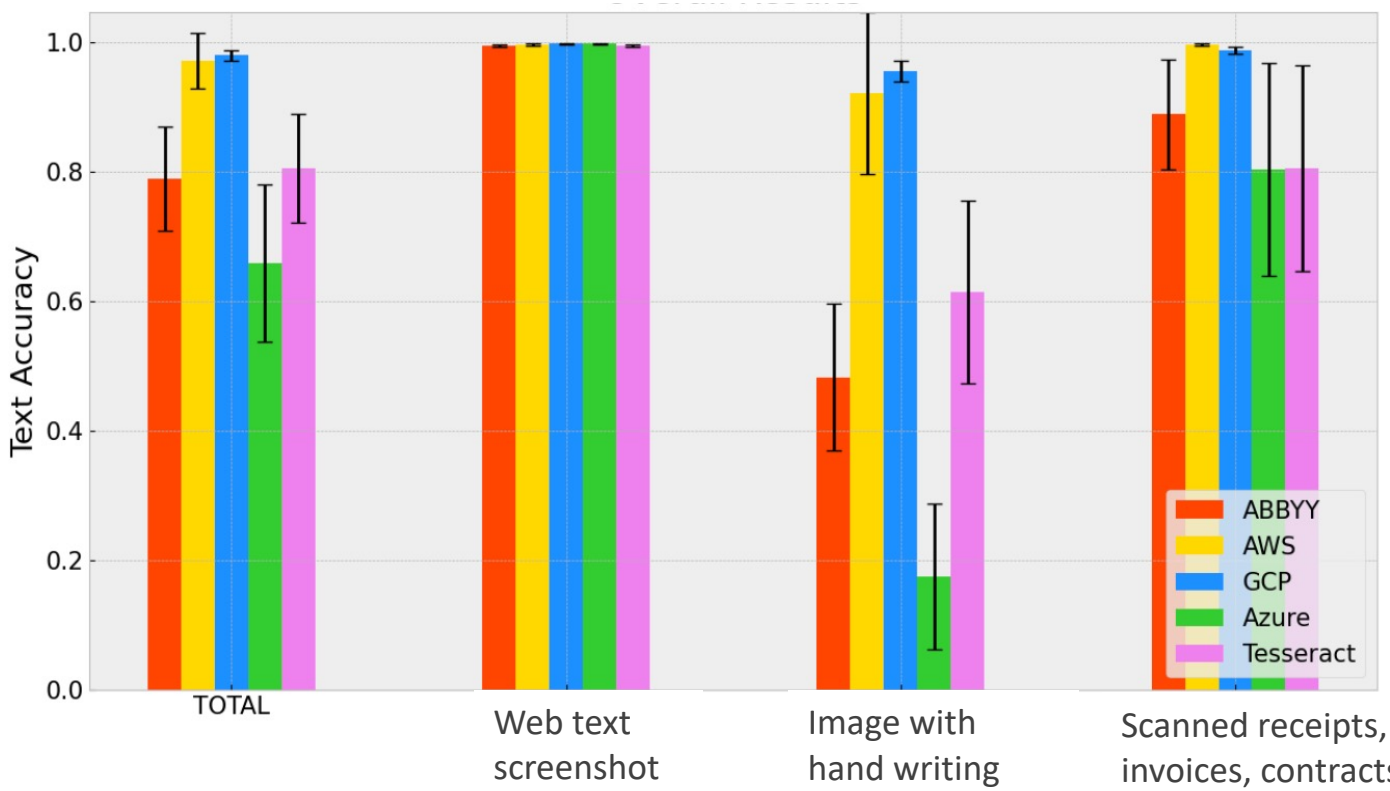
Skincare Product List:
From Sephora on Kaggle (link)

Missing Ingredient Update:
INCI Decoder

Ingredient Dictionary:
Web scraping from Paula Choice

OCR model

Optical Character Recognition to detect and extract text from image



ABBYY FineReader 15

AWS Amazon Textract

GCP Google Cloud Platform Vision API

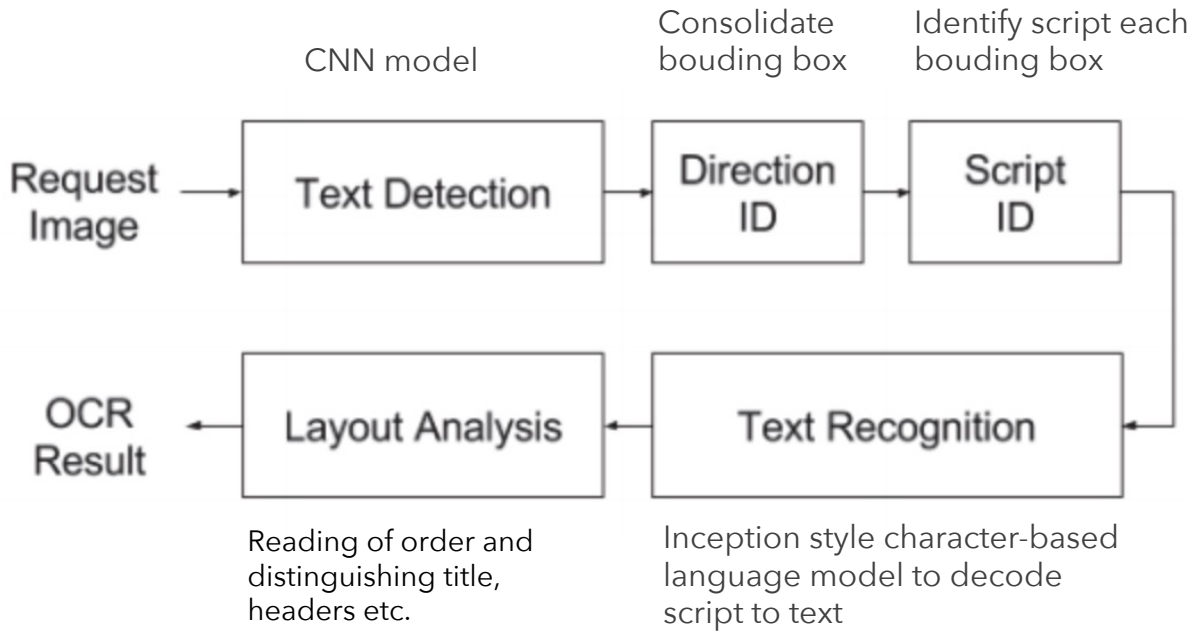
Azure Microsoft Azure Vision API

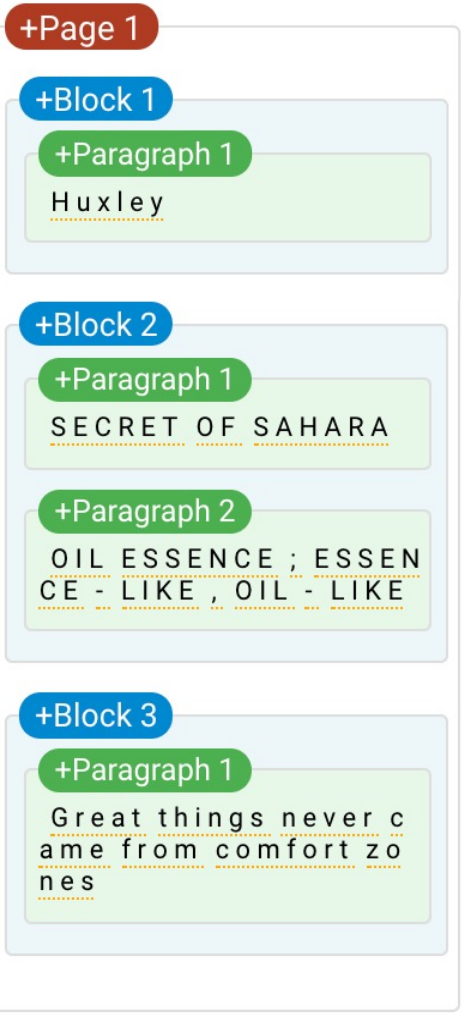
Tesseract OCR Engine

Source: [Best OCR by Text Extraction Accuracy](#) by Cem Dilmegani

OCR model

GOOGLE VISION OCR





Source: [Secret of Google OCR Service]() by Edward Ma

text matching metric

The Levenshtein distance measures how apart are two sequences of word, or the minimum number of edits done to make one sequence the same with the other.

$$\frac{(|a| + |b|) - lev_{a,b}(i,j)}{|a| + |b|}$$

$$\text{lev}_{a,b}(i,j) = \begin{cases} \max(i,j) \\ \min \begin{cases} \text{lev}_{a,b}(i-1,j) + 1 \\ \text{lev}_{a,b}(i,j-1) + 1 \\ \text{lev}_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} \end{cases} \end{cases}$$

if $\min(i,j) = 0$,

otherwise.

```
Str1 = "Apple Inc."

Str2 = "apple Inc"

Distance = levenshtein_ratio_and_distance(Str1,Str2)

print(Distance)

Ratio = levenshtein_ratio_and_distance(Str1,Str2,ratio_calc = True)

print(Ratio)
```

```
The strings are 2 edits away

0.8421052631578947
```

Source: [String Matching in Python](#) byFrancisco Javier Carrera Arias

demo

Top match item found: 83.0%

Brand: ESTÉE LAUDER

Description: Advanced Night Repair Synchronized Recovery Complex II

Price: $98

Rating: 4.3/5

Skin Type: ['Sensitive', 'Dry', 'Normal', 'Combination']

INGREDIENT DECODING:

	Ingredient	% Match	Match Item	Rating	Categories
0	Water	90	Aqua, Eau, Water	Average	Solvent
1	Bifida Ferment Lysate	100	Bifida Ferment Lysate	Best	Humectant
2	Methyl Gluceth-20	100	Methyl Gluceth-20	Good	Humectant
3	Peg-75	90	PEG	Good	Cleansing Ager
4	Bis-Peg-18 Methyl Ether D…	100	Bis-PEG-18 Methyl Ether …	Good	Silicone, Textu
5	Butylene Glycol	100	Butylene Glycol	Good	Humectant, Te
6	Propanediol	100	Propanediol	Good	Solvent, Textu
7	Cola Acuminata (Kola) Se…	95	Cola Acuminata Seed Extr…	Average	Plant Extracts,
8	Hydrolyzed Algin	90	Algin	Good	Antioxidant, Pl
9	Pantethine	100	Pantethine	Best	Humectant

[Streamlit]

future improvement

UPDATING DATASET

- Mid market products
- Japanese and Korean products
- New ingredients

PROMOTING INTERGRITY

- Machine learning product recommendation that also high light green/clean products and brands.

COMMERCIALIZATION

- Develope business plan with subscription revenue model.