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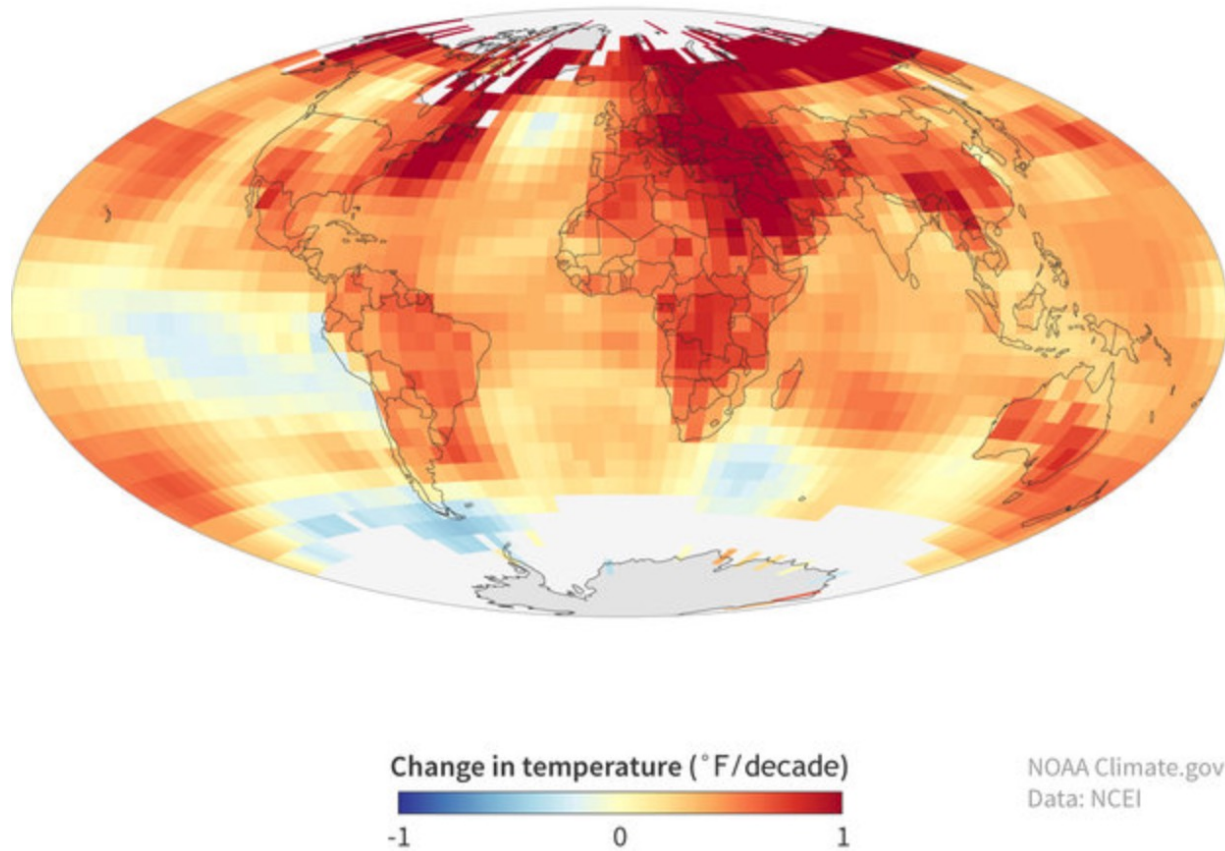
She is informative in order to save your time and money.

Finally, MySA promotes the habit:

BUY ONLY WHAT YOU REALLY NEED

Github

environment facts

Change in temperature (°F/decade)
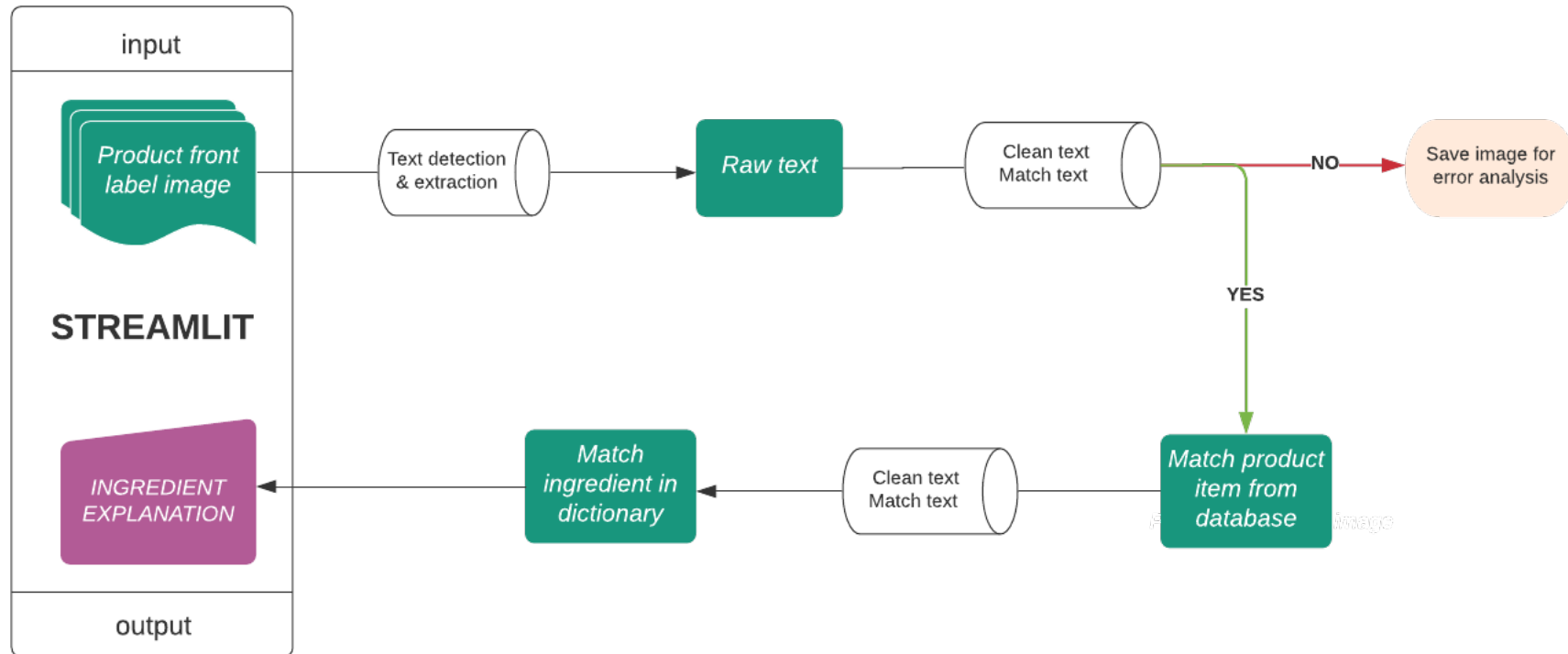
-1   0   1

NOAA Climate.gov
Data: NCEI

RECENT TEMPERATURE TRENDS 1990 - 2020

TOP CAUSES

Power plants

Transportation

Farming, fertilizers
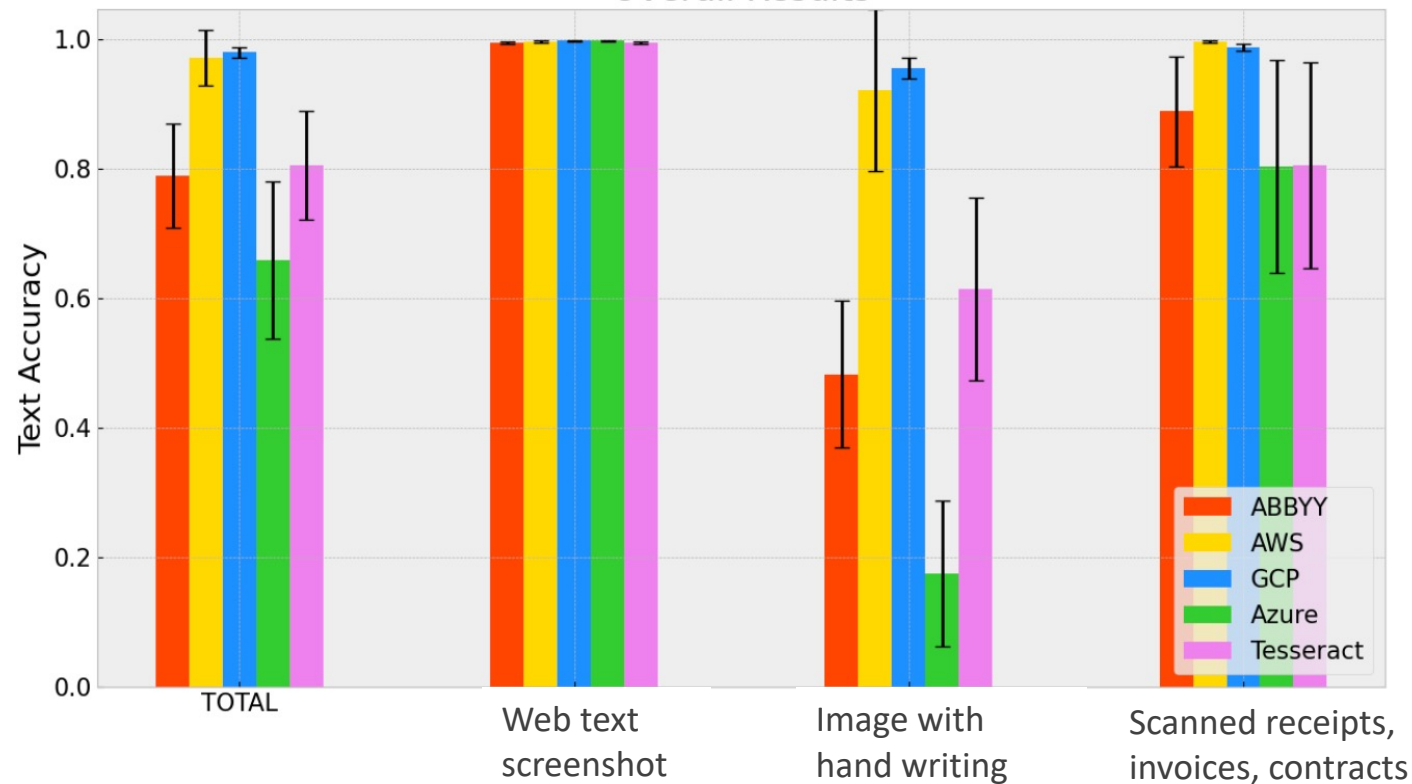
Deforestation

Permafost melting

GARGABE

framework

OCR model

Optical Character Recognition to detect and extract text from image
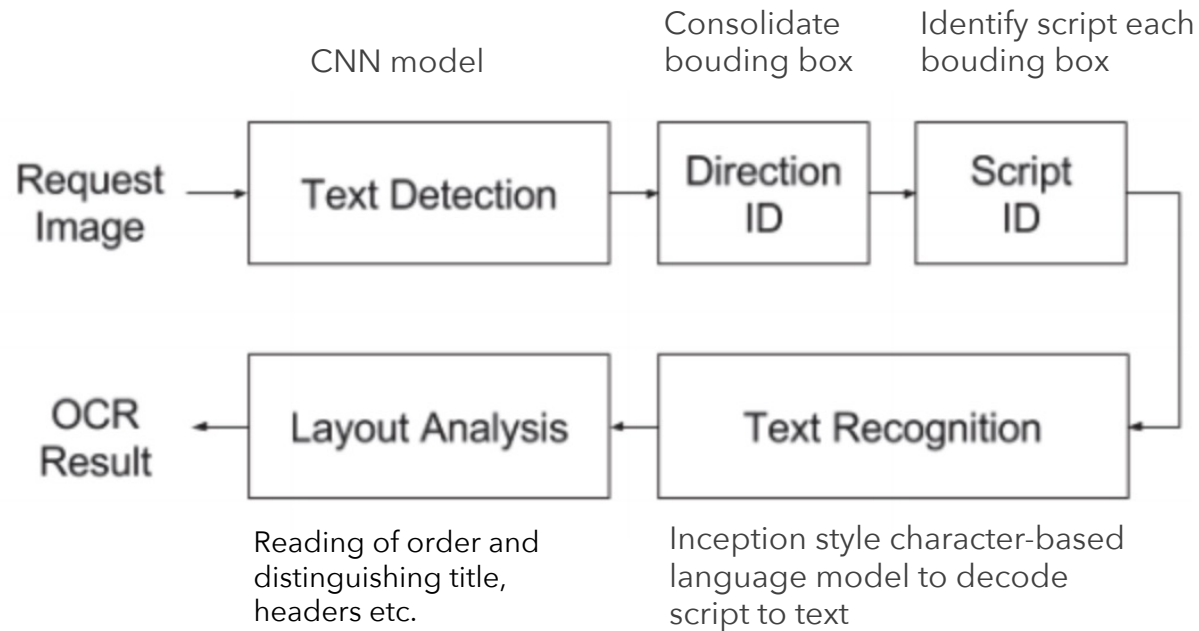

ABBYY FineReader 15

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

CNN model

Consolidate bouding box

Identify script each bouding box

Reading of order and distinguishing title, headers etc.

Inception style character-based language model to decode script to text

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.

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

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

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

otherwise.

```python
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

dataset

Sun protect	HATOMUGI	UV Milky Gel SPF 31 PA+++ 250ml	21	4.6	0	0	0	0	0	1	Water, Ethanol, Ethylhexyl Methoxycinnamate
Moisturizer	L'ORÉAL	Revitalift Day Cream SPF35 PA+++ 50ml	18	4.2	0	1	1	1	0	0	Aqua/Water, Cyclohexasiloxane, Glycerin, My
Treatment	HUXLEY	Secret of Sahara Oil Essence 30ml	19	4.6	1	1	1	1	0	0	Opuntia Ficus-Indica Stem Extract, Butylene
Eye cream	HUXLEY	Secret of Sahara Eye Cream 30ml	22	4.0	1	1	1	1	0	0	Opuntia Ficus-Indica Flower Extract, Aqua/W
Moisturizer	CURÉL	Intensive Moisture Face Milk 120ml	23	4.3	1	1	1	0	0	0	Water (Aqua), Glycerin, Allantoin ,Cetyl-Pg Hy
Moisturizer	HUXLEY	Secret of Sahara Anti Gravity Cream 50ml	21	4.1	1	1	1	1	0	0	Opuntia Ficus-Indica Stem Extract, Glycerin,
Treatment	DERMACTION PLUS	Alive Plankton repairing intensive serum 50ml	39	3.9	0	1	1	1	0	0	Aqua (Water), Butylene Glycol, Glycerin, Niaci

From text extracted match with one instance in skincare product dataset

From product's ingredient list match with multiple instances in ingredient dictionary

name	Rating	Categories	Benefits	Description
Aqua, Eau, Water	Average	Solvent	Hydration	Further, too much water can be a problem for skin because it can disrupt the
1, 2-Hexanediol	Good	Preservative, Humectant		A synthetic preservative and moisture-binding agent belonging to a class of
10-Hydroxydecanoic Acid	Good	Emollient, Occlusive/Opacifying Agent	Hydration	A synthetic ingredient that functions as a skin-conditioning and occlusive age
4-T-Butylcyclohexanol	Good	Emollient	Hydration	A synthetic fatty alcohol that functions as an emollient and soothing agent. It
Acacia Farnesiana Extract	Worst	Fragrance: Synthetic and Natural, Plant Extracts		A fragrant extract from a type of acacia tree. There is no research showing it
Acacia Senegal Gum	Good	Texture Enhancer, Emulsifier	Soothing	Acacia senegal gum is an herb used in various capacities to enhance the sm
Acai	Best	Antioxidant, Plant Extracts	Anti-Aging, Soothing	Acai (pronounced "ah-sigh-ee") is a small berry with a deep purple color. It is
Acetic Acid	Worst	Irritant, pH Adjuster/Stabilizer		Acid found in vinegar, some fruits, and human sweat. It can be sensitizing to

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 Agen
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 and low end products

  Japanese and Korean products

  New ingredients

PROMOTING INTERGRITY

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

COMMERCIALIZATION

  Expand solution's knowledge to related fields such as medicine, supplements, preprocessed food, etc.

  Develope business plan with subscription revenue model.