Sentiment Analysis with Supervised Machine Learning

A JOURNEY TO FIND A Positive Word FOR REVIEWING FOOD

6th Oct 2023

project by: Haruka Kozen
GitHub: https://github.com/haru-pyon

Problem:

Your passion for food has led you to create your **own brand** of food products.

For advertising, you would like to choose **THE POSITIVE WORDS** that make people think it's a great food !

But what is positive? Is it Delicious, Tasty or Affordable ?

what if you could draw inspiration from Amazon reviews?

Which words are commonly associated with positive food reviews?

Description of Data

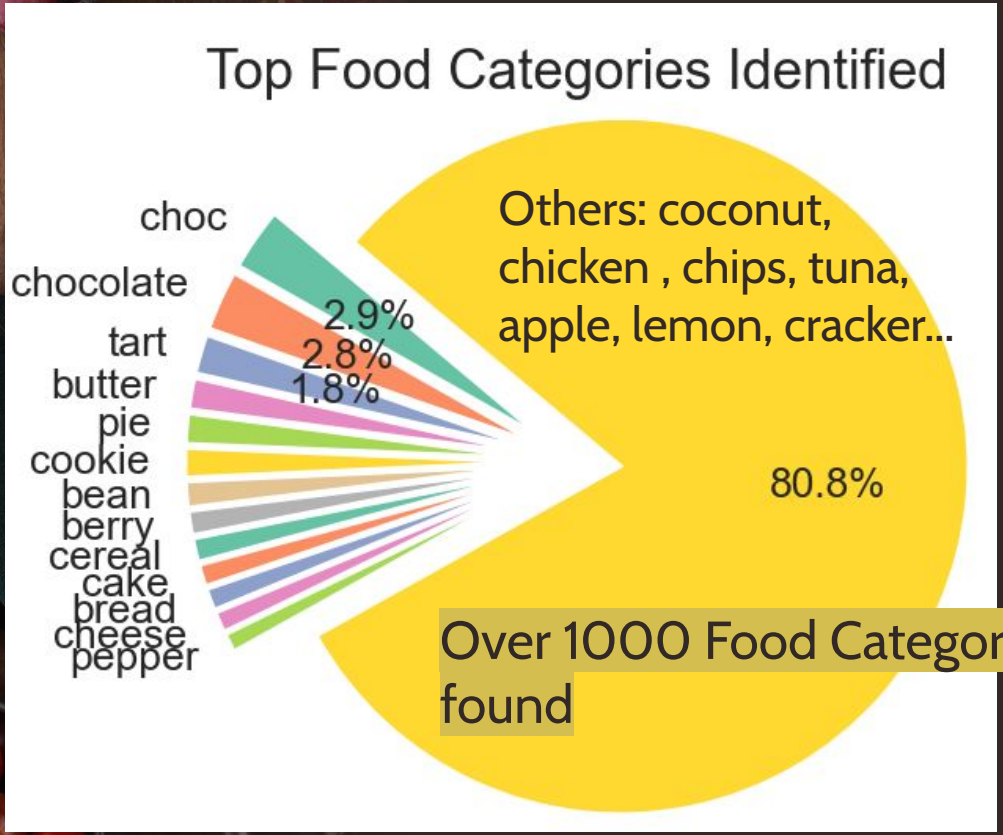
570,000 rows + 10 columns (90,000 rows used for final BERT)

ID	ID of the Review
Product ID	Unique identifier for the product
User ID	Unique identifier for the user
ProfileName	Profile name of the user
Helpfulness Numerator	Number of users who found the review helpful
Helpfulness Denominator	Number of users who indicated whether they found the review helpful or not
Score	Rating between 1 and 5
Time	Timestamp for the review
Summary	Brief summary of the review
Text	Text of the review

01

Word Cloud Analysis

Food Category Analysis



Over 1000 Food Categories found

Find Name of Food

Natural Language Toolkit

Library offering text resources

WordNet

Linguistic resource. Organize group of words with similar meanings

Made Food list from **food.n.02**

Pre-processing before Word Cloud

Remove Stop Words

SpaCy

Natural Language Processing (NLP) library

en_core_web_sm

Pre-trained model includes English vocabulary

Counting Words

(BoW)

Technique for creating a vocabulary of words and counts the occurrence

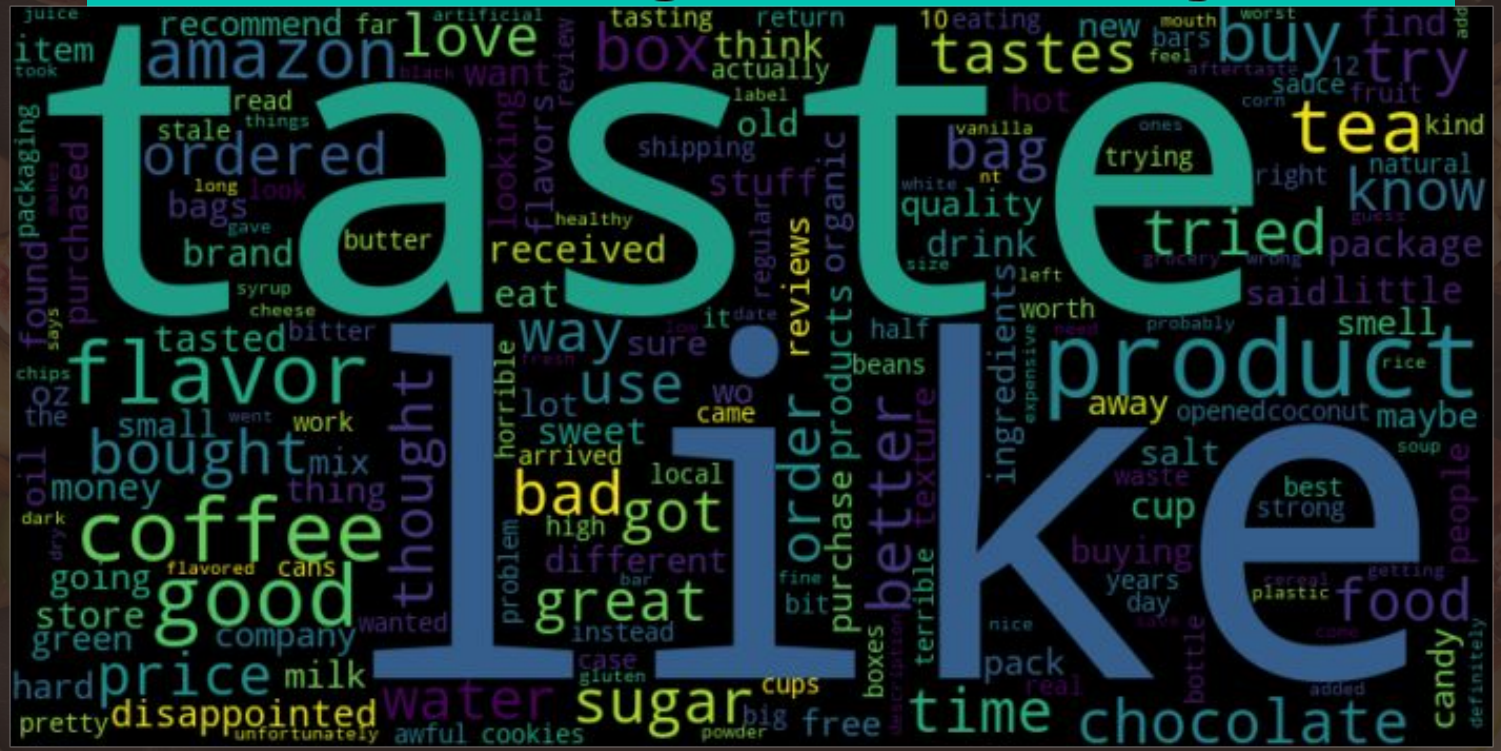
CountVectorizer

Specific implementation in Scikit-learn Library. Converts text into a matrix of token counts.

Word Cloud for Positive Scores (Unigrams)

like
great
good
taste
coffee
tea
flavor
product
price
amazon
use
tastes
bought
cup
delicious
sweet
time
stuff
right
definitely
recommend
bit
family
health
milk
years
flavors
hot
favorite
makes
nice
order
free
eat
easy
water
find
know
tried
better
love
buy
mix
add
try
way
store
little
want
fresh
salt
chocolate
drink
sugar
found
best
think
food
bag
tasty
box
loves
high calories
butter
green
size
perfect
shipping
day
buying
products
strong
tasted
brands
kids
honey
pasta
fat
10
blend
smooth
cream
black
actually
worth
peanut
chips
hard
fruit
foods
bread
drinking
dark
getting
said
it
months
brand
excellent
grocery
sure
year
pack
oil
quality
tasting
probably
trying
bar
chicken
stores
texture
work
husband
flavored
came
days
making
added
especially
eating
go
usually
thought
breakfast
treat
enjoy
far
bars
having
extra
bad
feel
new
rice
package
candy
looking
oz
rich
expensive
amazing
regular
half
people
snack
ordered
need
old
bags
natural
purchased
ingredients
organic
different
cookies
wonderful
cereal
sauce
local
beans
available
cheese
juice
going
low
comes
bottle
gluten
coconut
long
thing
small
lot
highly
healthy
purchase
pretty
quick
teas
morning
loved
things
deal
real
soup
the
big
home
protein
diet
cups
arrived
light
case
vanilla
happy
package
far
bars
extra
bad
feel
new
rice

Word Cloud for Negative Scores (Unigrams)



Neutral word "taste" appears while positive words such as"like" and "love" also appears. "coffee", "tea", "chocolate".

Word Cloud for Positive Scores (Bigrams)

"gluten free" , "green tea" and "dark chocolate"
gets interests. "good" and "great" comes after "taste".
What "grocery store" means is unclear at this level.

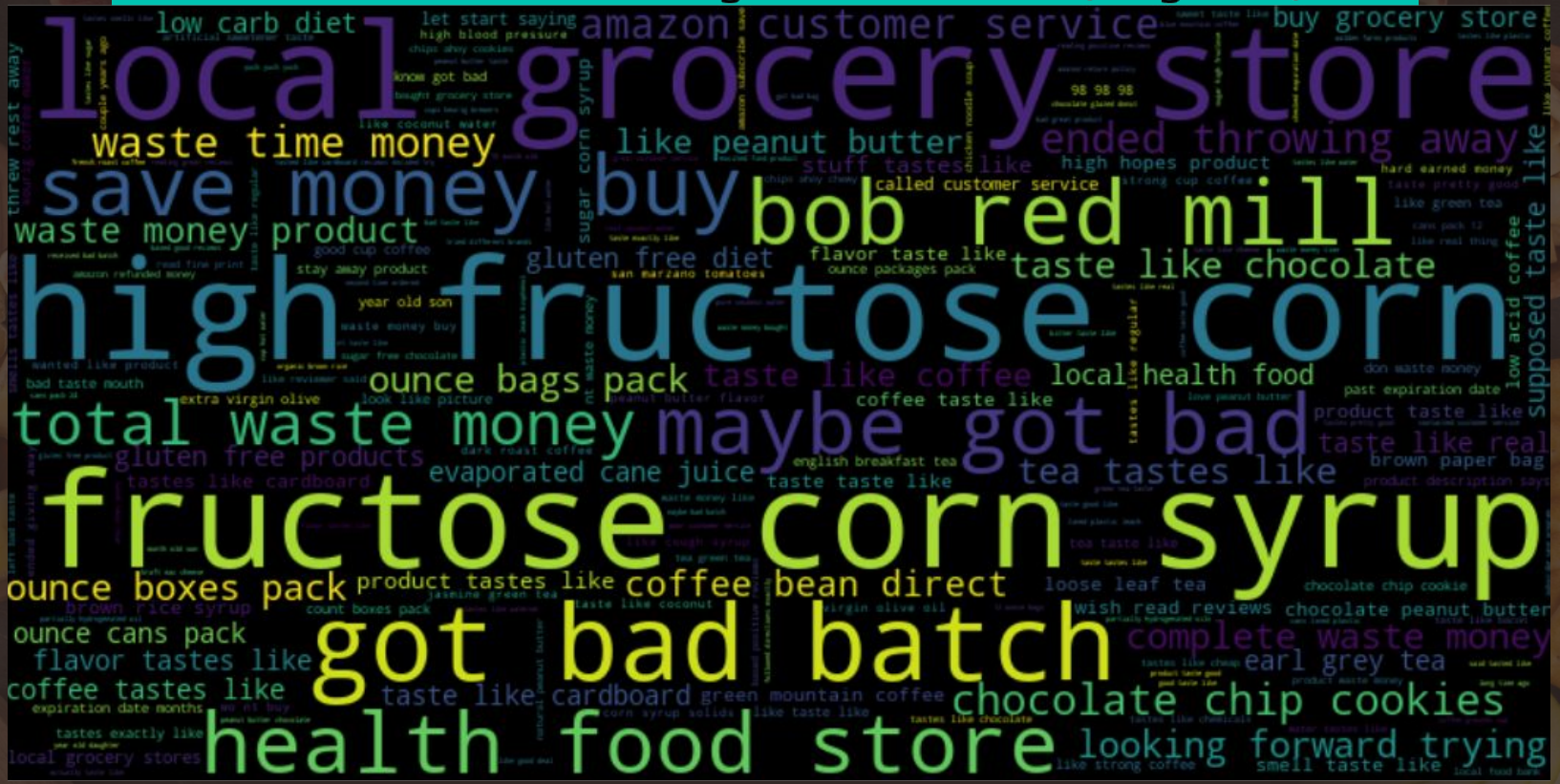
Word Cloud for Negative Scores (Bigrams)

Here you see "taste like" is often used for a bad (score 1 or 2) reviews. "coffee" meant "instant coffee"

Word Cloud for Positive Scores (Trigram)

On Trigram level, there are suggestions that "local grocery store" may associate with "hard find", "happy find".

Word Cloud for Negative Scores (Trigram)



Notice "fructose corn syrup" appears both on Positive and Negative reviews.

fructose corn syrup?

"... It uses cane sugar instead of high fructose corn syrup" (4 Score review)

"... Most have too much sugar or worst, corn syrup. This mixer is all natural" (5 Score review)

"read label discovered contains ... high fructose corn syrup ." (2 Score review)

People using a fructose corn syrup as a bad unhealthy example. Therefore it appears on both Positive / Negative reviews.

local store ?

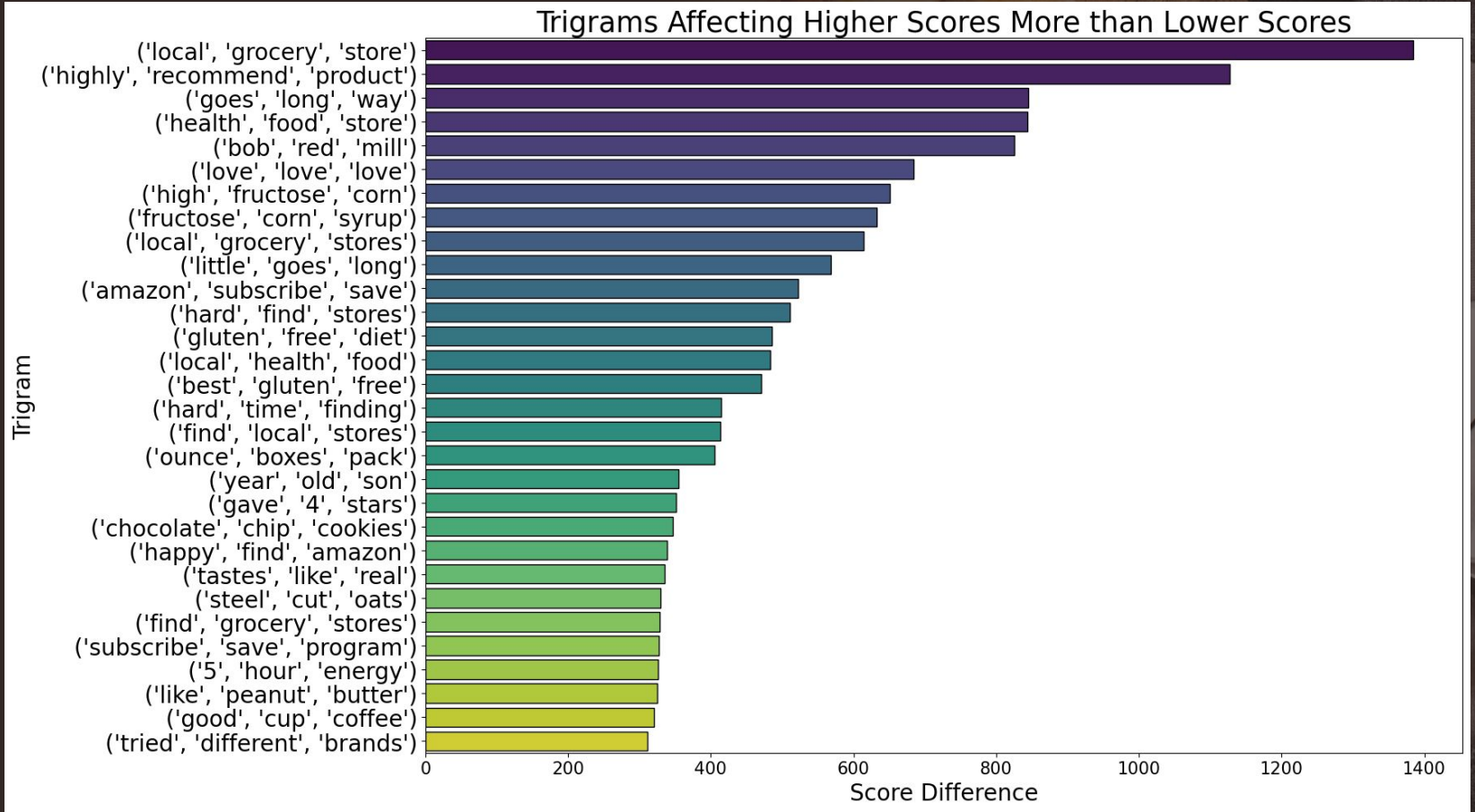
"... my local store went out of business ....thanks amazon!" (5 Score review)

"... I found them by accident on amazon after hunting all over my local stores" (5 Score review)

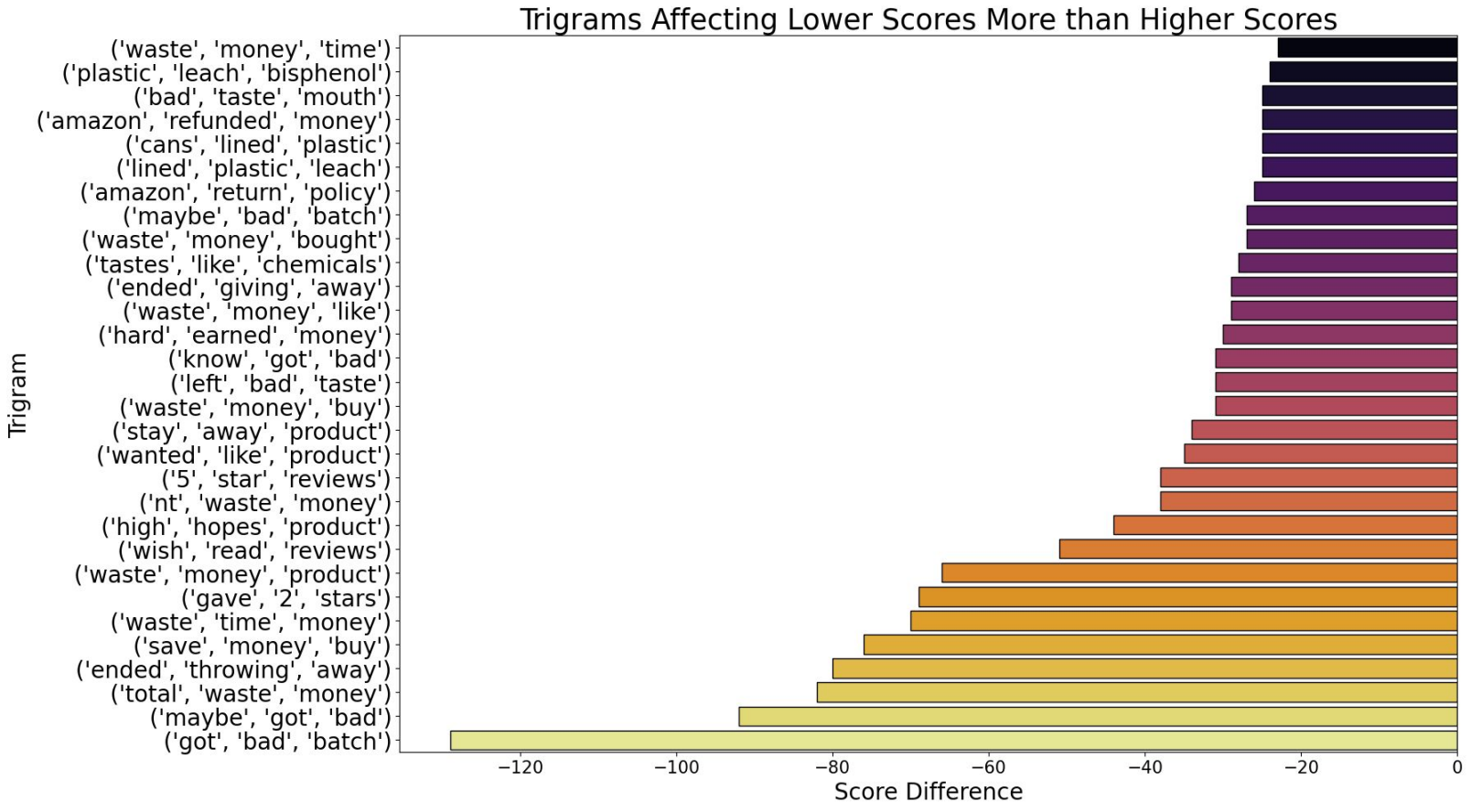
" I was disappointed though when i went to a local store and saw it for less than what i paid just for shipping. " (2 Score review)

People appreciate when they could not find the product locally. While disappointed when local store sells cheaper price than Amazon.

Trigram Analysis of Positive Review



Trigram Analysis of Negative Review



Summary

Avoid Neutral Words

Don't need to use "Tastes like" if the food taste is good!

Appeal Healthiness

Healthy ingredients must be clearly mentioned

You can buy this only here

People don't need your product if they can buy similar products at their local store

Advertisement Suggestion for Positive Impression:

" Discover the Organic Difference – No Corn Syrup! Tastes great, sold Exclusively Here! "

02

Deep Learn with NLP

Albert BERT

Deep learning model for natural language processing (NLP).

Lighter version of BERT

01

Tokenize text

did not remove stop words

02

Hyperparameter Tune

learning_rates 2e-5

num_epochs_ 1

03

Train with train data

Baseline Accuracy: 0.19

Train data Accuracy: 0.44

04

Evaluate test score

Baseline Accuracy: 0.19

Test data Accuracy: 0.53

Predicting Score with text!

5

This chocolate tastes like happiness.

4

I enjoyed this banana. It was fresh and tasty.

3

Pasta is OK. Not super delicious but tastes good.

1

The cookie was awful. I had to throw it away.

2

I found the same product at half the price at a local store. I liked the taste but was disappointed.

Feature Importance Analysis

```
Enter text: This chocolate tastes like happiness.

Word: [SEP], Importance Score: 3.9458
Word: ., Importance Score: 3.7869
Word: _this, Importance Score: 3.7131
Word: _taste, Importance Score: 3.7085
Word: _like, Importance Score: 3.6330
Word: s, Importance Score: 3.3688
Word: _chocolate, Importance Score: 3.3556
Word: _happiness, Importance Score: 3.2476
Word: [CLS], Importance Score: 2.1146

Predicted Score: 5
```

Although the score seems reasonable, the evaluation of the importance of each word by the model may not align with human intuition.

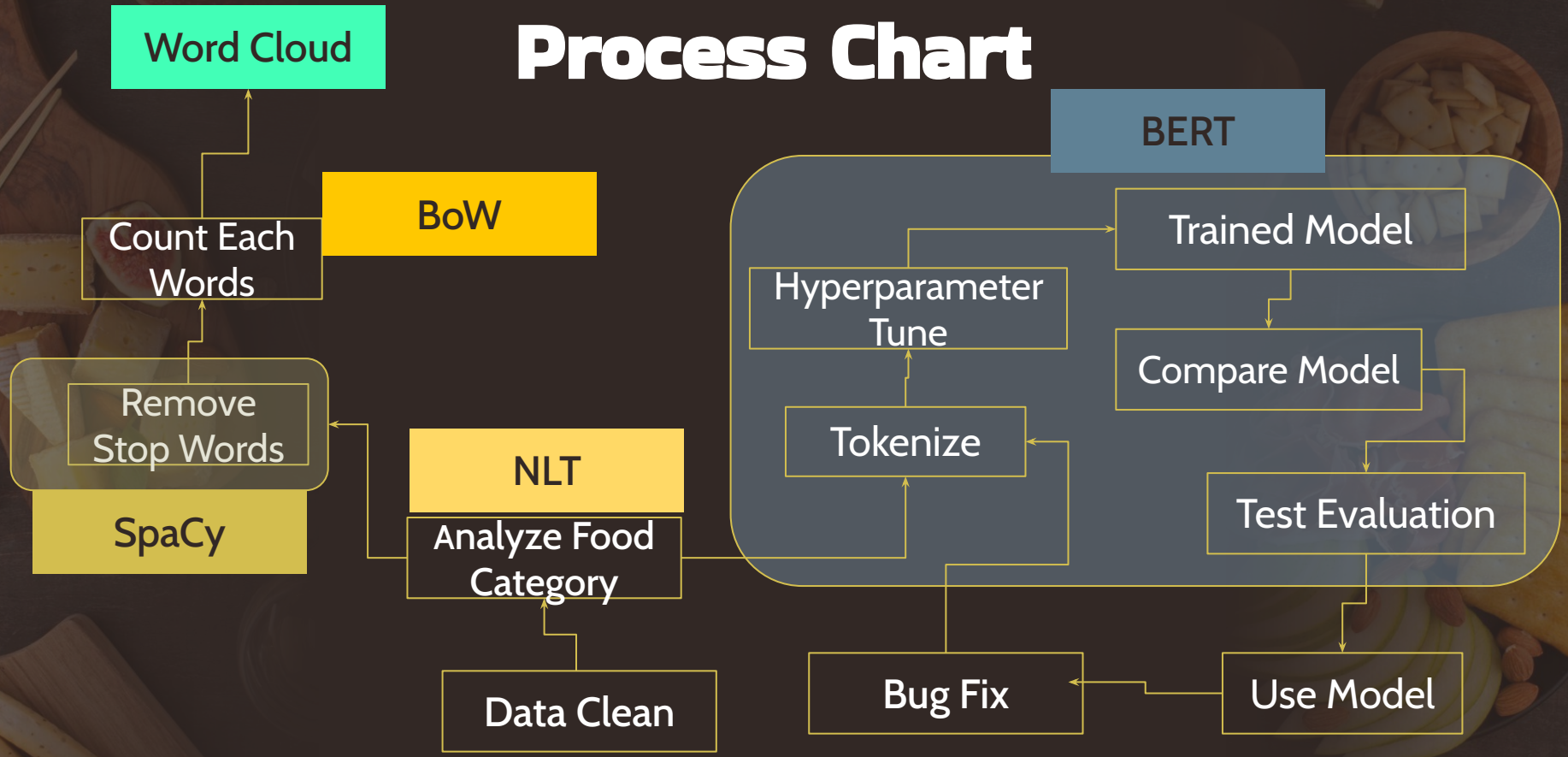
Advertisement Suggestion for Positive Impression:

Discover the Organic Difference – No Corn Syrup! Tastes great, sold Exclusively Here!

```
Word: ,, Importance Score: 4.0177
Word: [SEP], Importance Score: 3.9458
Word: !, Importance Score: 3.7842
Word: !, Importance Score: 3.7842
Word: _taste, Importance Score: 3.7085
Word: _syrup, Importance Score: 3.7007
Word: _the, Importance Score: 3.6994
Word: _exclusively, Importance Score: 3.6269
Word: _difference, Importance Score: 3.3883
Word: s, Importance Score: 3.3688

Predicted Score: 5
```

Process Chart



Learning

- Pre-progressing and Training of NLP is time-consuming.
  Could not finish this project without Tqdm (Python Library)

```
Epoch 2 Training:   2%|▌                    | 121/5900 [12:45<3:15:3
```

  Which made me downgrade from BERT bert-base-uncased → distilbert-base-uncased → Albert BERT for optimization.
  It made me learn about various models and parameters.

- What learned during this Bootcamp made me autonomously fix bugs of NLP.

Thanks!

Amazon Fine Food Reviews on Kaggle

https://www.kaggle.com/datasets/snap/amazon-fine-food-reviews

CREDITS: This presentation template was created by **Slidesgo,** including icons by **Flaticon** and infographics & images by **Freepik**

slidesgo