

Cloud Engineering Project Sentiment Analysis

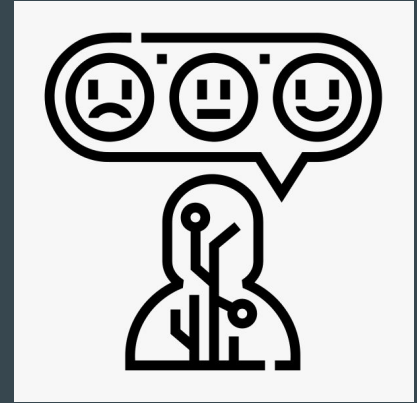
...

Xiyi Lin, Stella Wang, Yan Wang, Grace Xie, Yumin Zhang

Business Problem

The primary goal of this project is to develop a real-time sentiment analysis system that evaluates the sentiment of customer reviews on Amazon.

This system aims to classify each review as *positive, neutral, or negative*, providing immediate feedback to both sellers and buyers.



The analysis will help sellers improve their products and services and give prospective buyers insights into product quality and customer satisfaction.

Data: Amazon Fine Food Reviews

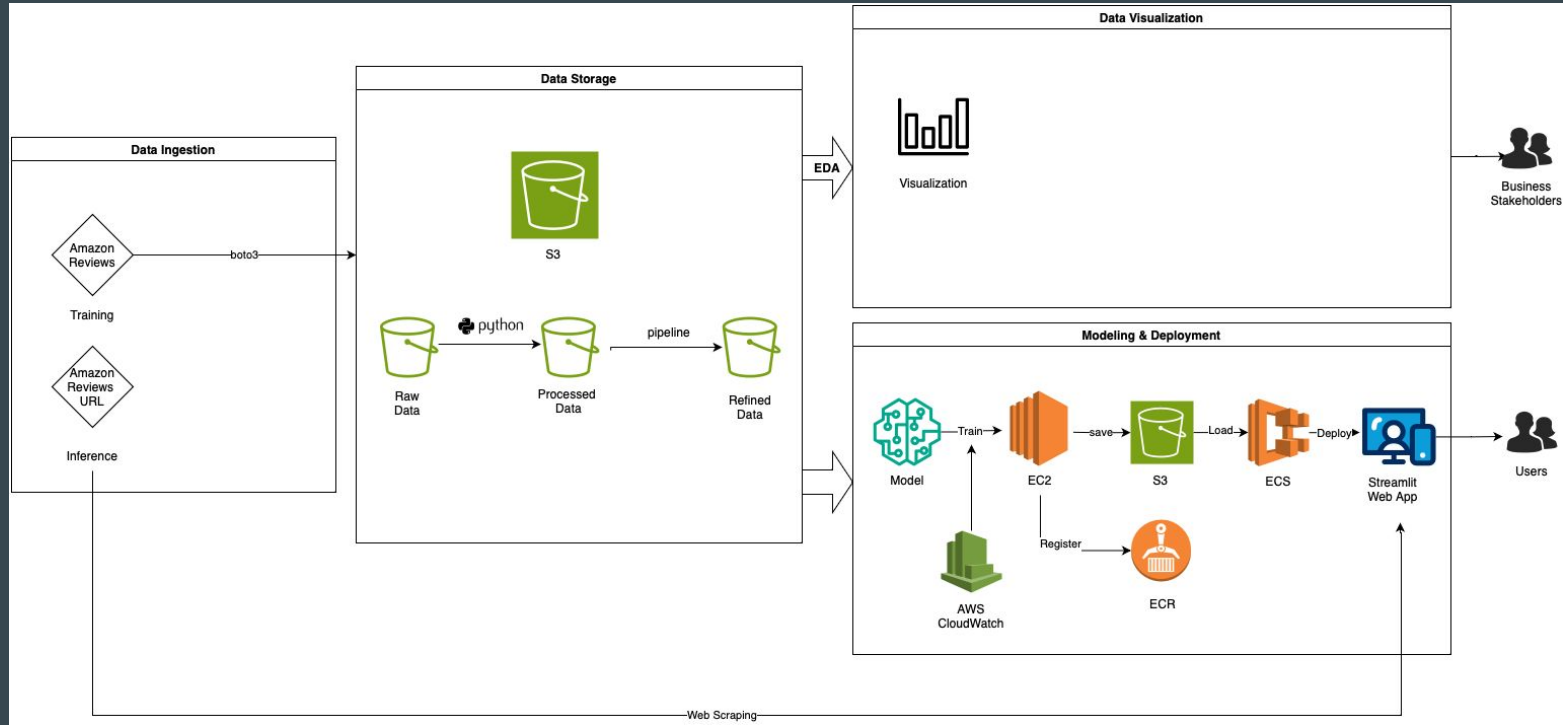
Source: <https://www.kaggle.com/datasets/snap/amazon-fine-food-reviews>

- Reviews from Oct 1999 - Oct 2012
- 568,454 reviews
- 256,059 users
- 74,258 products
- 260 users with > 50 reviews



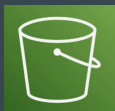
	ProductId	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summary	Text
89	B0019CW0HE	0	0	5	2012-09-12	Great Food	Great food! I love the idea of one food for al...
528	B000G6RYNE	20	27	1	2009-09-27	They changed the Chips now they taste horrible	I once loved these chips and they were the onl...
571	B000G6RYNE	2	2	5	2009-09-05	Great Chip!	With Kettle Chips, you really have to be caref...
633	B000G6RYNE	0	0	5	2012-09-08	I made a mistake.....	This may sound rediculous, but I actually orde...
1497	B002KGN4LE	23	23	5	2010-06-08	Detailed reply from company over infant feedin...	MY QUESTION FOLLOWED BY THEIR REPLY - VERY GOO...
...

Architecture



https://app.diagrams.net/#G1d7NA_R85oRzC4rppq43iS9rdR6t-2z_F%23%7B%22pageId%22%3A%22GrQoB2HVvH_tSdHMKaNm%22%7D

Data Sizing



S3:

- **Storage** : 1 GB per month (S3 Standard)
- **Requests** :
 - PUT, COPY, POST, LIST: 1000 requests
 - GET, SELECT, other requests: 1000 requests



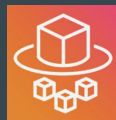
EC2:

- **Tenancy** : Shared Instances
- **Operating System** : Linux
- **Workload** : Consistent
- **Instances** : 1 (t2.micro)
- **Pricing** : Compute Savings Plans 3yr No Upfront
- **Monitoring** : Disabled



ECS:

- **Data Stored** : 0.5 GB per month



Fargate:

- **Operating System** : Linux
- **CPU Architecture** : x86
- **Average Duration** : 10 minutes
- **Tasks/Pods** : 1 per day
- **Ephemeral Storage** : 20 GB



CloudWatch:

- **Metrics** : 4 (includes detailed and custom metrics)

AWS Pricing Calculator

Upfront cost

0.00 USD

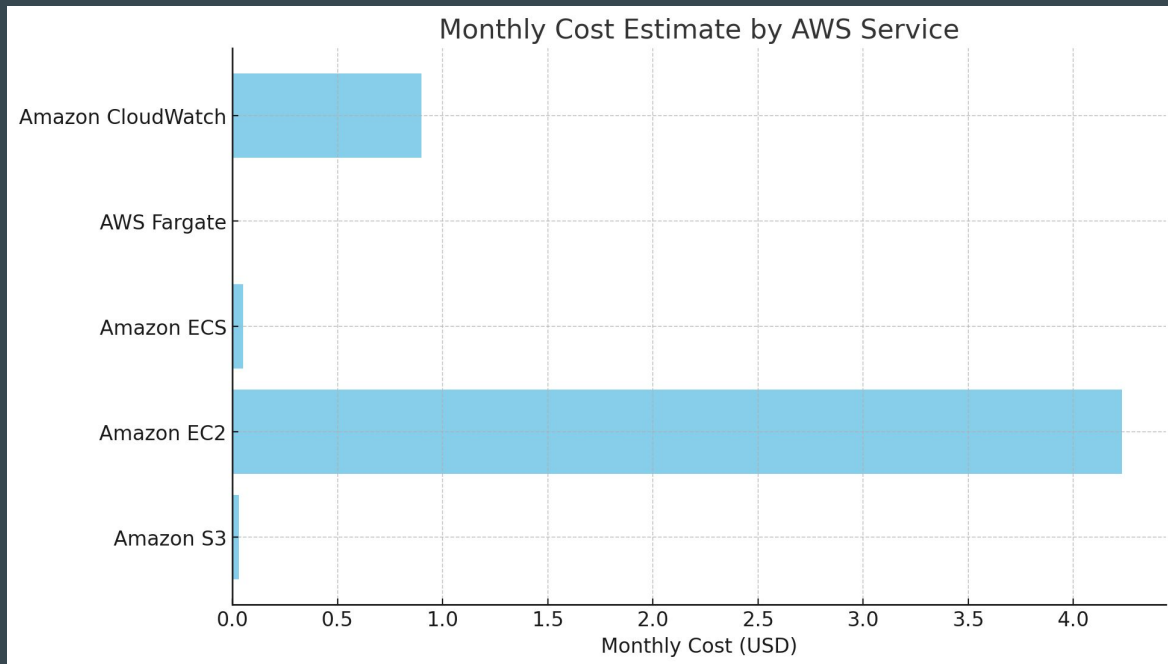
Monthly cost

5.21 USD

Total 12 months cost

62.52 USD

(Includes upfront cost)



<https://calculator.aws/#/estimate?id=81e86d18ac3dc3807e14aed8c299d83affec24c8>

Text Processing

- Remove HTML tags
- Lowercase
- Remove non-alphabetic characters and numbers
- Normalize informal terms and abbreviations
- Tokenize
- Remove Stop words
- Lemmatization

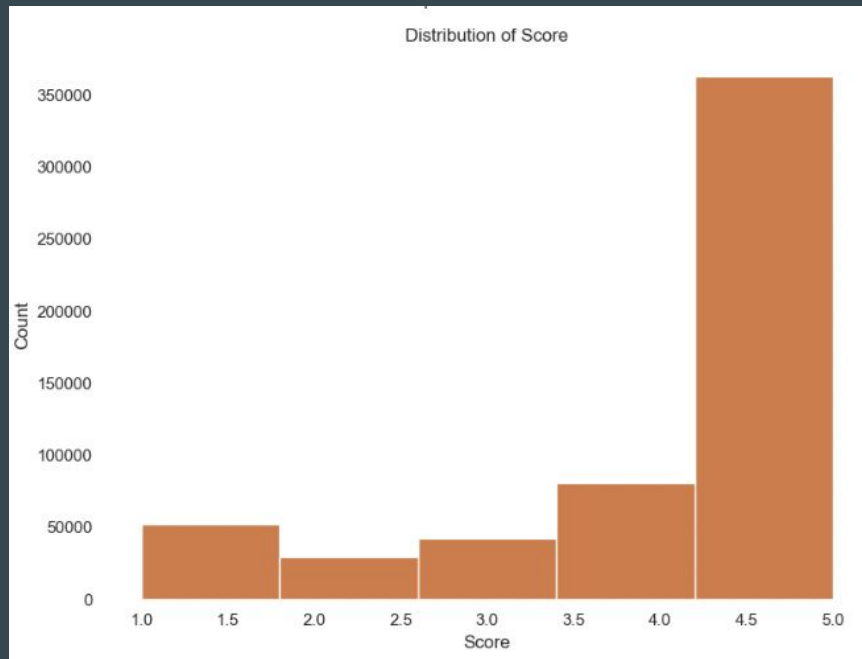
—
Embedding:

pre-trained GloVe embeddings

	converted_text	processed_text
0	I have bought several of the Vitality canned d...	bought several vitality canned dog food produc...
1	Product arrived labeled as Jumbo Salted Peanut...	product arrived labeled jumbo salted peanutsth...
2	This is a confection that has been around a fe...	confection around century light pillowy citrus...
3	If you are looking for the secret ingredient i...	looking secret ingredient robtussin believe f...
4	Great taffy at a great price. There was a wid...	great taffy great price wide assortment yummy ...
...
568449	Great for sesame chicken..this is a good if no...	great sesame chickenthis good better resturant...
568450	I'm disappointed with the flavor. The chocolat...	disappointed flavor chocolate note especially ...
568451	These stars are small, so you can give 10-15 o...	star small give one training session tried tra...
568452	These are the BEST treats for training and rew...	best treat training rewarding dog good groomin...
568453	I am very satisfied ,product is as advertised,...	satisfied product advertised use cereal raw vi...

Figure. Text before & after being processed

EDA



Problem: Class imbalance

Define the true labels based on *Score* :

1.0 - 2.0: Negative

3.0: Neutral

4.0 - 5.0: Positive

Fix the imbalance:

Downsample to ensure each class has the same number of data points.

Positive Reviews



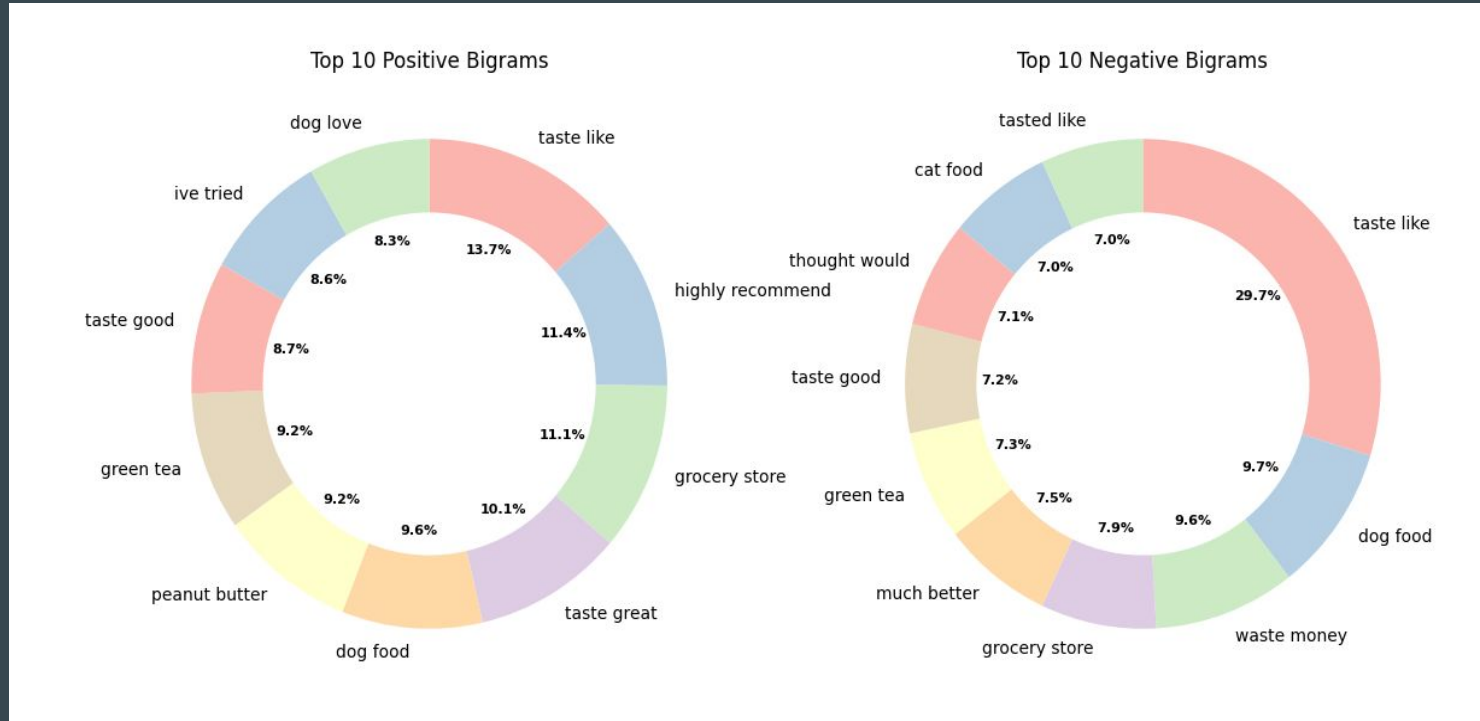
Negative Reviews



Neutral Reviews



Words like ‘taste’, ‘product’, ‘use’ , and ‘good’ are common across different sentiments, but their context changes.



Positive bigrams: 'taste great', 'highly recommend'

Negative bigrams: 'thought would', 'waste money'

Models

1. Simple Neural Network (SNN)

- Layers: Embedding (GloVe) → Flatten → Dense (softmax)
- Use Case: Quick, efficient, and interpretable for small to moderate datasets.

2. Convolutional Neural Network (CNN)

- Layers: Embedding (GloVe) → Conv1D → GlobalMaxPooling1D → Dense (softmax)
- Use Case: Captures local dependencies; suitable for moderately complex patterns.

3. Long Short-Term Memory (LSTM)

- Layers: Embedding (GloVe) → LSTM → Dense (softmax)
- Use Case: Handles long-term dependencies; ideal for complex text sequences.

Models (cont.)

	Model Name	Test Accuracy
1	<u>Simple Neural Network (SNN)</u>	0.720
2	<u>Convolutional Neural Network (CNN)</u>	0.738
3	<u>Long Short-Term Memory (LSTM)</u>	0.730

Why doesn't the LSTM model achieve the highest accuracy?

Deployment on Cloud

✓ config

! config.yaml

✓ models

≡ cnn_model.h5

≡ lstm_model.h5

≡ snn_model.h5

≡ tokenizer.pkl

✓ src

> venv

🔗 analysis.py

🔗 app.py

≡ glove.6B.100d.txt

🔗 nn_modeling.py

🚢 dockerfile

≡ dockerfile-app

🔗 pipeline.py

≡ pylint.txt

≡ requirements.txt

Services (1) Info

[Manage tags](#)[Update](#)[Delete service](#)[Create](#)

Filter launch type

Any launch type ▼

Filter service type

Any service type ▼

< 1 >



<input type="checkbox"/>	Service name ▼	ARN	Status ▼	Service... ▼	Deployments and tasks ▼
<input type="checkbox"/>	sentiment-app-3	arn:aws:ec...	Active	REPLICA	<div><div></div></div> 1/1 Tasks running

May 21, 2024 at 21:54 (UTC-5:00)

You can now view your Streamlit app in your browser.

May 21, 2024 at 21:54 (UTC-5:00)

Network URL: <http://172.31.46.98:8501>

May 21, 2024 at 21:54 (UTC-5:00)

External URL: <http://18.188.77.206:8501>

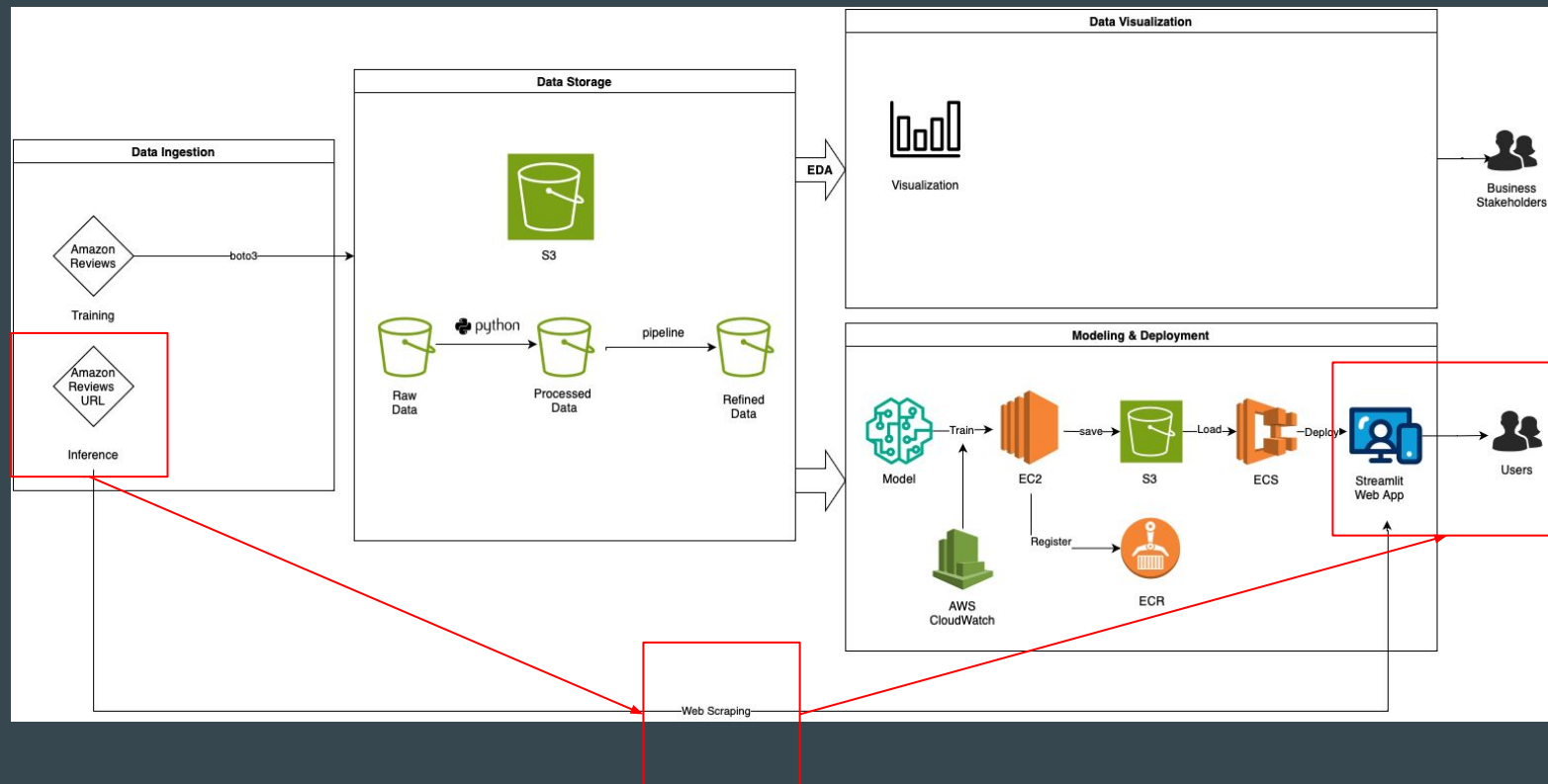
May 21, 2024 at 21:54 (UTC-5:00)

Collecting usage statistics. To deactivate, set `browser.gatherUsageStats` to `false`.

May 21, 2024 at 21:54 (UTC-5:00)

[1m1/1 [0m [32m ————— [0m [37m [0m [1m0s [0m 497ms/step
[1m1/1 [0m [32m ————— [0m [37m [0m [1m1s [0m 501ms/step

Now...



Inference Using Web Scrapping

Amazon.com : Entenmann's L × +

URL

https://www.amazon.com/Entenmanns-Little-Bites-Chocolate-Muffins/dp/B0068L1QYY/ref=...?crid=ZOZ7I2FB4QG1&dib=eyJ2Ij...

amazon Delivering to Evanston 60201 Update location

All Medical Care Groceries Best Sellers Amazon Basics New Releases Music Prime Customer Service Today's Deals Amazon Home Books Registry Pharmacy Gift Cards

Grocery Deals Snacks Breakfast Warm Beverages Cold Beverages Cooking Staples Baby Food Candy & Chocolate Subscribe & Save International Foods SNAP-eligible Groceries

OREO Double Stuf Chocolate Sandwich Cookies, Family Size, 3 Packs ★★★★★ 16,493 \$12.20 \$12.84 prime Save 5% more with Subscribe & Save

Grocery & Gourmet Food > Breads & Bakery > Pastries & Bakery > Muffins

Entenmann's Little Bites 20 Pouch Value Pack

Food Name

Entenmann's Little Bites Chocolate Chip Muffins, 20 Pouch Value Pack

Brand: Entenmann's

4.5 ★★★★★ 156 ratings | Search this page

#1 Best Seller in Muffins

2K+ bought in past month

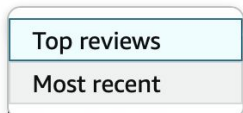
\$13.99 (\$0.44 / Ounce)

Get \$10 off instantly: Pay \$3.99 \$43.99 upon approval for the Amazon Store Card. No annual fee.

Style: 20 Pouch Chocolate Chip | 1 Pack

\$13.99 (\$0.44 / Ounce) \$27.85 (\$0.42 / Ounce) 4 options from \$54.95

Inference Using Web Scrapping



Top reviews from the United States



★★★★★ **So delicious**

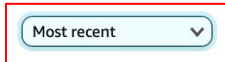
Reviewed in the United States on March 3, 2024

Style: 20 Pouch Chocolate Chip | 1 Pack | **Verified Purchase**

These are sooo delicious 😊

Helpful

| Report



Select

From the United States



DSCC

Stars



Expired!

Reviewed in the United States on May 8, 2024

Style: 20 Pouch | Variety 4 Pack | **Verified Purchase**

Text

Beware of short dates!!

We received 3 of the (4pk) 80ct muffins on May 8th, 2024. Expiration date on all of the brownies is May 12th, 2024. The blueberry and chocolate chip muffins expiration is June 9th, and the birthday cake is June 16th!!

How terrible. We have less than a month to consume these 240 packages of muffins,

Paste an URL:

Find the 'Most Recent' Review Section -> Grab the top reviews shown on the current page

Inference Using Web Scrapping

Most recent

From the United States

john e oneil

★★★★★

Expired Chips!!

Reviewed in the United States on May 26, 2024

Flavor Name: Classic Mix | **Verified Purchase**

Every bag in box is expired and I just received them today!!!

Helpful

Report

Susan Collier

★★★★★

Expired product

Reviewed in the United States on May 26, 2024

Flavor Name: Baked & Popped Mix | **Verified Purchase**

EVERY package of the chips expired in January 2024!

Helpful

Report



Predicted Sentiments:		
	Review Text	Predicted Sentiment
0	every package chip expired january read	negative
1	chip gotten expensive buying single serving size worth money started buying big bag	positive
2	chip lot read	positive
3	delivery fast chip problem chip taste odd old taste read	neutral
4	get exactly say doesnt love big box chip even say variety may vary received pretty nic	positive
5	frito lay company ashamed amount chip one bag reduced lowest denominator bag fl	negative
6	perfect quick snack definitely worth read	positive
7	two packet missing box otherwise fresh good kid snack box read	negative

There are 4 positive, 3 negative, and 1 neutral reviews, so we would recommend this product.

Use Text Processing Pipeline:
Automatically run when scraping the reviews

Inference Using Web Scrapping

Predicted Sentiments:

	Review Text	Predicted Sentiment
0	every package chip expired january read	negative
1	chip gotten expensive buying single serving size worth money started buying big bag	positive
2	chip lot read	positive
3	delivery fast chip problem chip taste odd old taste read	neutral
4	get exactly say doesnt love big box chip even say variety may vary received pretty nic	positive
5	frito lay company ashamed amount chip one bag reduced lowest denominator bag fl	negative
6	perfect quick snack definitely worth read	positive
7	two packet missing box otherwise fresh good kid snack box read	negative

There are 4 positive, 3 negative, and 1 neutral reviews, so we would recommend this product.

Make a recommendation conclusion based on *Majority Vote*!

Deployment: Streamlit

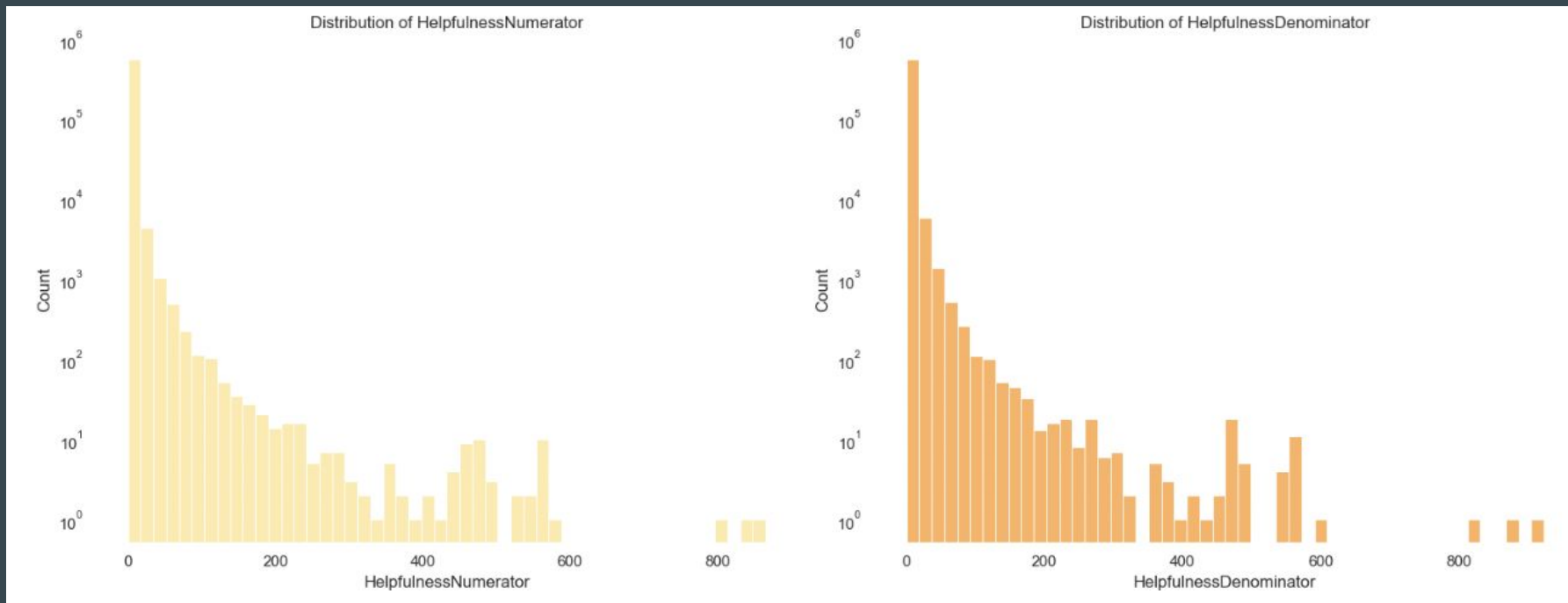


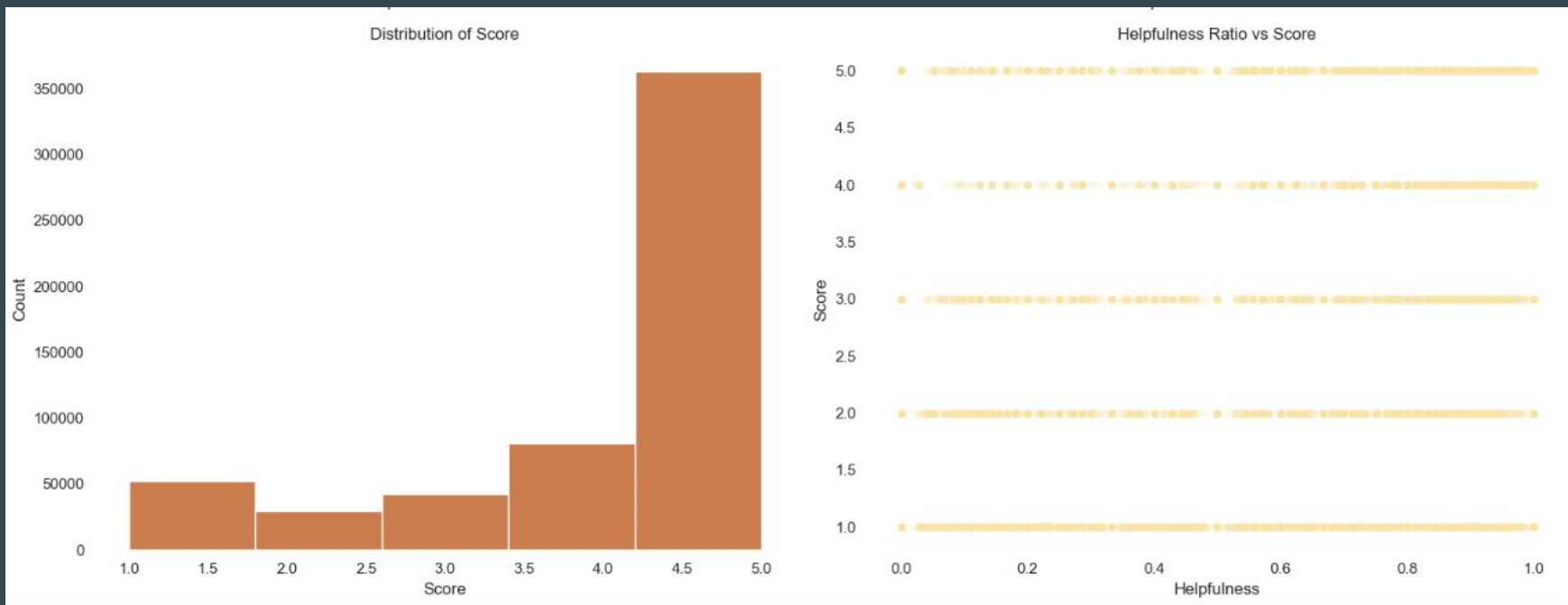
Demo Time



Thank You For Your Listening

Appendix





Deployment: Model Training and Artifact Storage (ECS → S3)

```
2024-05-23 23:23:56,603 - INFO - AWS login successful.
2024-05-23 23:23:56,609 - INFO - AWS session created.
2024-05-23 23:23:56,702 - INFO - Loading cached SSO token for https://d-9a6775d2
ac.awsapps.com/start/#
2024-05-23 23:23:57,218 - INFO - Uploaded nn_modeling.py to s3://cloudengineerin
gproject-s3/nn_modeling.py
2024-05-23 23:24:11,765 - INFO - Uploaded models/cnn_model.h5 to s3://cloudengin
eeringproject-s3/cnn_model.h5
2024-05-23 23:24:27,101 - INFO - Uploaded models/lstm_model.h5 to s3://cloudengi
neeringproject-s3/lstm_model.h5
2024-05-23 23:24:42,255 - INFO - Uploaded models/snn_model.h5 to s3://cloudengin
eeringproject-s3/snn_model.h5
2024-05-23 23:24:43,595 - INFO - Uploaded models/tokenizer.pkl to s3://cloudengi
neeringproject-s3/tokenizer.pkl
Pipeline execution completed successfully.
```