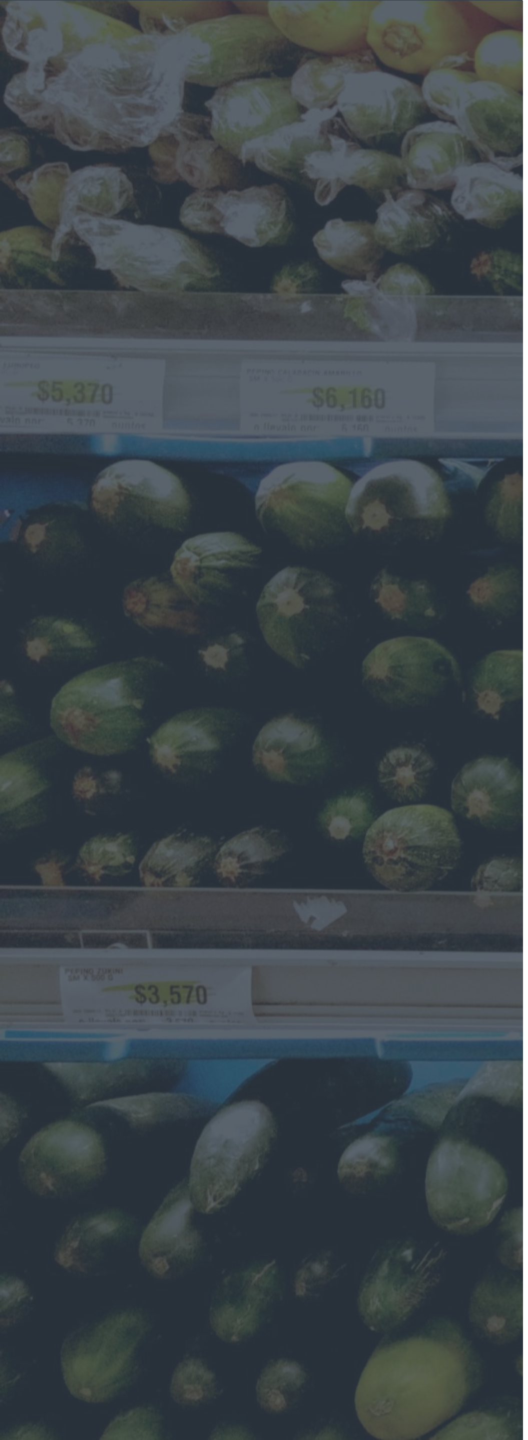
Customer sentiment analysis for Amazon Fine Food products



Outline of this presentation

01
Importance of customer sentiment analysis

02
Objective of our project

03
Overview of methodology

04
Deep-dive into dataset & unstructured data

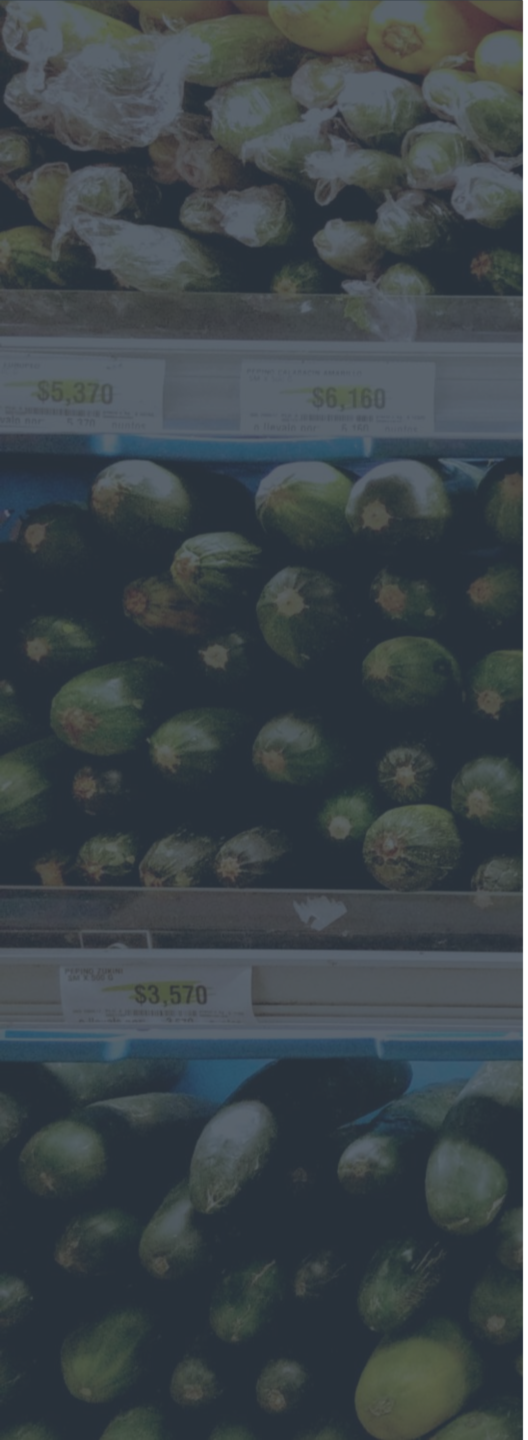
05
Data cleaning and preprocessing

06
Overall customer sentiment analysis

07
Aspect-based sentiment analysis

08
Discussion of insights and model performances

09
Conclusion and future direction



Outline of this presentation

01
Importance of customer sentiment analysis

02
Objective of our project

03
Overview of methodology

04
Deep-dive into dataset & unstructured data

05
Data cleaning and preprocessing

06
Overall customer sentiment analysis

07
Aspect-based sentiment analysis

08
Discussion of insights and model performances

09
Conclusion and future direction

Insights from customer reviews are business-critical in today's markets

Enhanced Customer Insights

Improved Product Development

Customer-centric Marketing

Competitive Advantage

Risk Management

Enhanced Customer Support

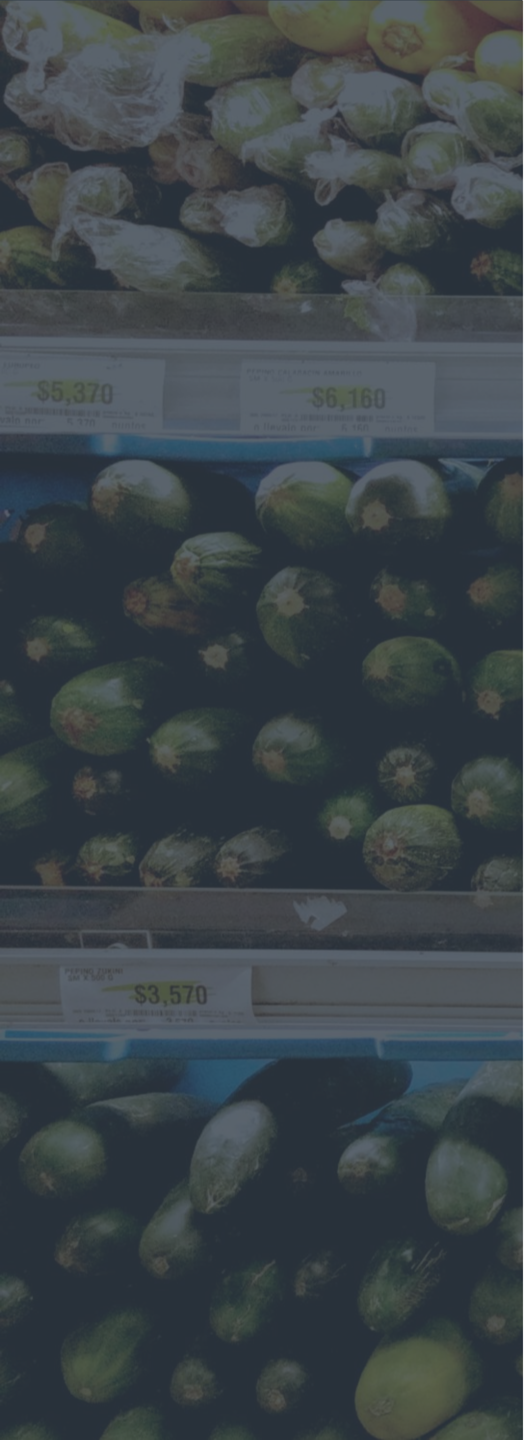
Data-driven Decision Making

Increased Customer Retention

Objective of our project

Exploration of different models to conduct **customer sentiment analysis** while **increasing complexity** and **level of detail**

and evaluating each method's **performance** in extracting valuable business insights

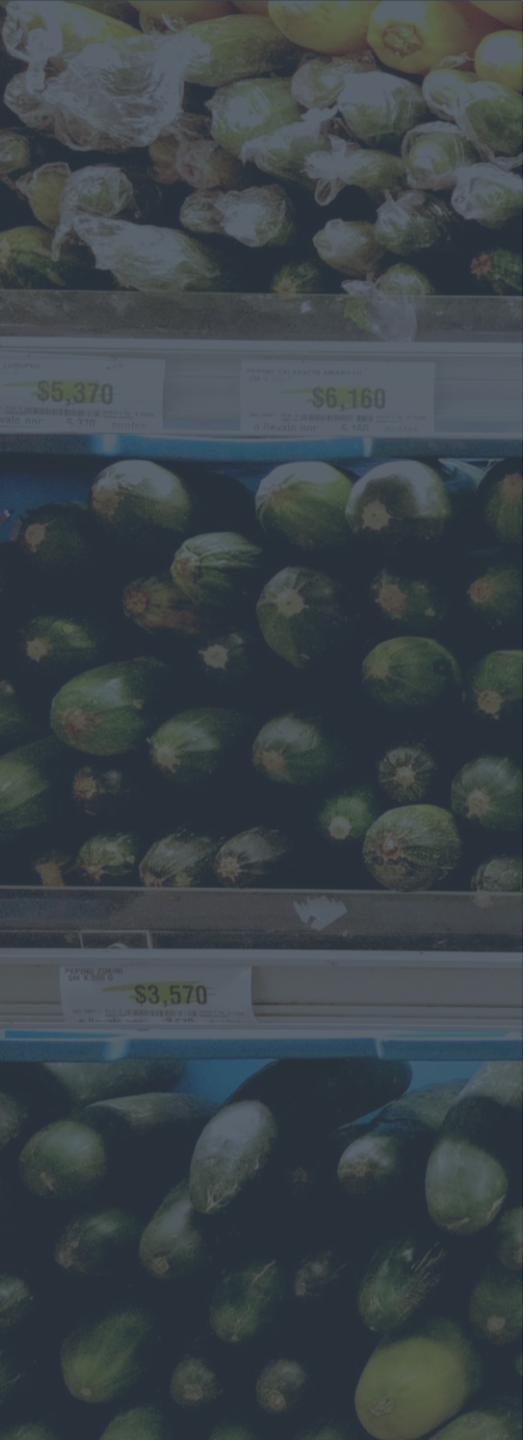


Outline of this presentation

For our initial customer sentiment analysis, we tested five different models

Overall sentiment analysis			Aspect-based sentiment analysis	
Logistic Regression	RoBERTa Transformer	DistilBERT Transfomer	Spacy Aspect Classifier	PyABSA Transformer
Naïve Bayes	ChatGPT 4	Multi-head Attention		

Increasing complexity level →



Outline of this presentation

01
Importance of customer sentiment analysis

02
Objective of our project

03
Overview of methodology

04
Deep-dive into dataset & unstructured data

05
Data cleaning and preprocessing

06
Overall customer sentiment analysis

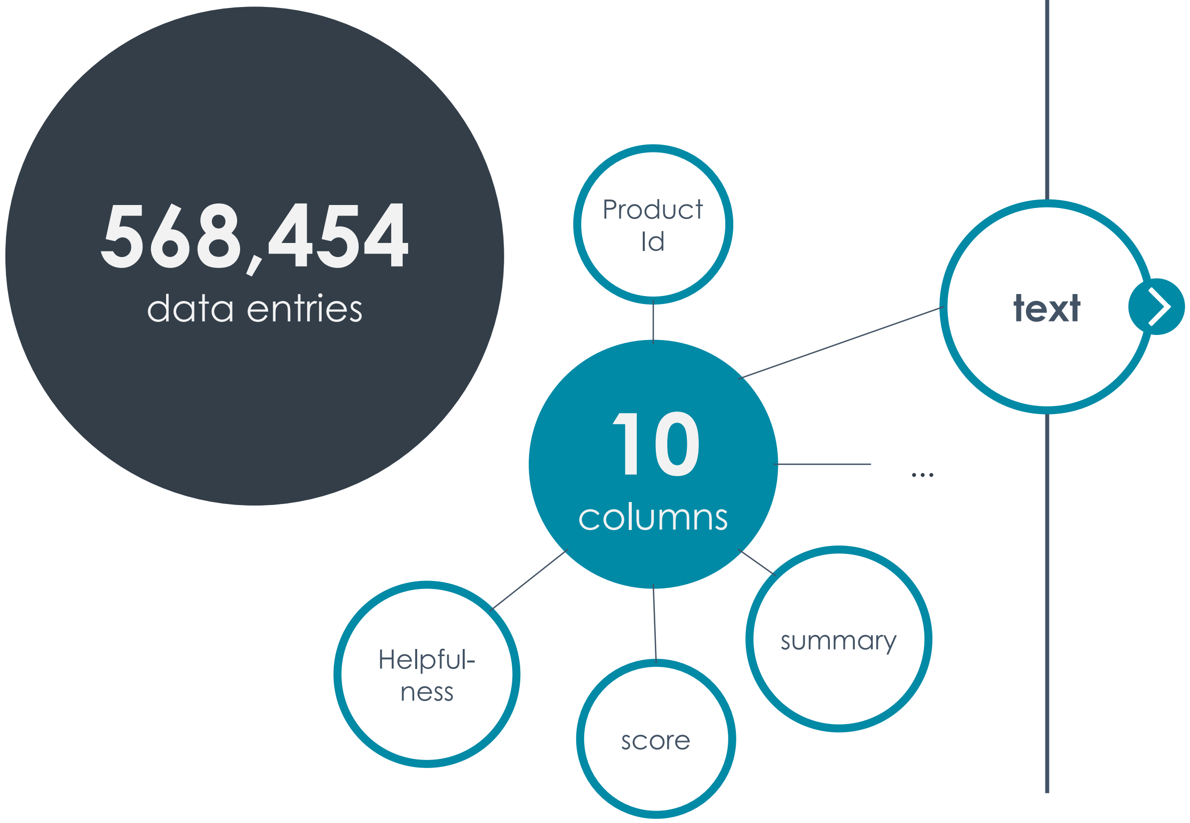
07
Aspect-based sentiment analysis

08
Discussion of insights and model performances

09
Conclusion and future direction

Unstructured review text is Amazon Food Review's main feature

Large dataset containing Amazon customers' review about food ordered



We are taking the unstructured review text data to conduct our customer sentiment analysis

*"I have never been a huge coffee fan. However, my mother purchased this little machine and talked me into trying the Latte Macciato. No Coffee Shop has a better one and I like most of the other products, too (as a usually non-coffee drinker!).
The little Dolche Guesto Machine is super easy to use and prepares a really good Coffee/Latte/Cappuccino/etc in less than a minute (if water is heated up). I would recommend the Dolce Gusto to anyone. Too good for the price and I'am getting one myself! :)"*

We conducted extensive data cleaning and preprocessing

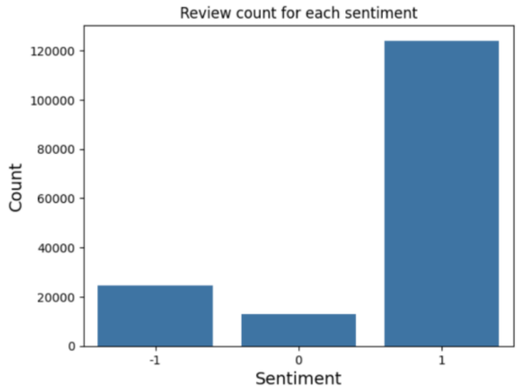
Data cleaning

- Removing unwanted columns as UserID and Profile Name
- Dropping Helpfulness Numerator and Helpfulness-Denominator
- Checking the number of duplicate records by grouping the data on UserId and ProductId, and dropping duplicate records and keeping the latest one per group

Data preprocessing

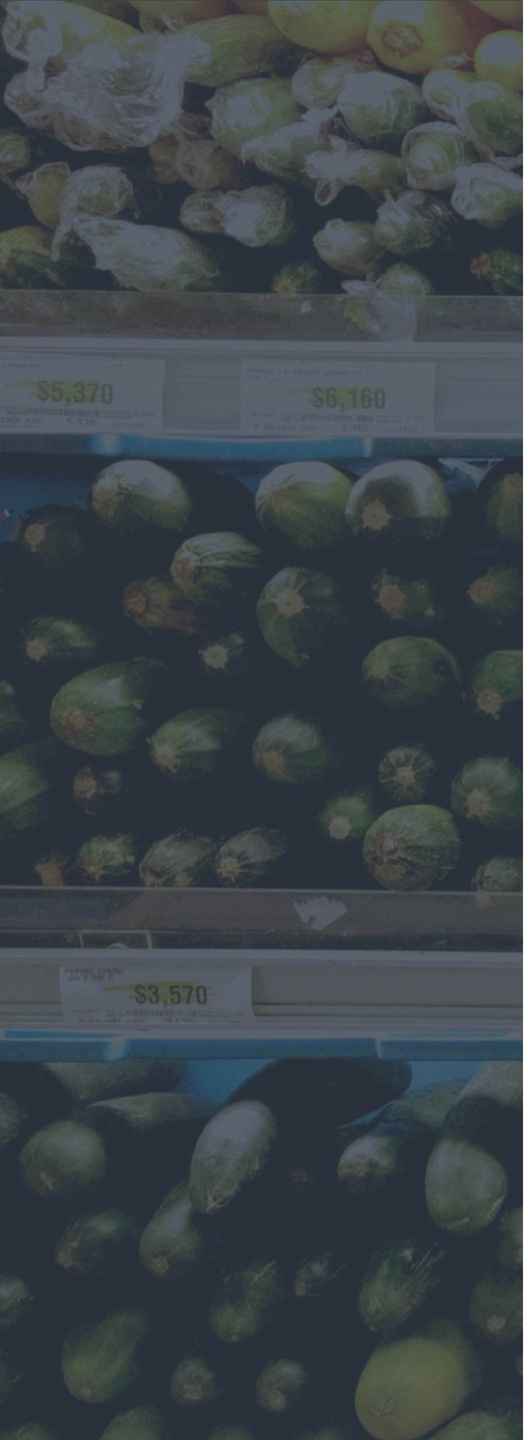
- Combining the summary and text column
- Cleaning text data by removing irrelevant elements like URLs, mentions, punctuation, and stopwords, while also tokenizing the text for further analysis → Clean_text column
- Converting the customer-given score from 1-5 into positive (+1), neutral (0) and negative (-1) scores

Imbalanced dataset for 3 sentiments



Feature engineering

- Calculating a 'Helpfulness Ratio' where 'Helpfulness Denominator' > 0, otherwise put -1
- Computing the length of each review in the original 'Summary_Text' column
- Computing the length of each review in the Clean_Text' column
- Polarity and Subjecitvity of the review using TextBlob
- Compute number of unique products purchased by each user



Outline of this presentation

Logistic Regression

Overall sentiment

Model performance
- performs extremely well in predicting the positive sentiment (= 1) of the reviews with a precision score of around 90%
- struggles with accurately classifying reviews with neutral sentiment
- demonstrates good performance in identifying negative sentiment

Advantages
- computational efficient (low training and prediction costs) underscores suitability for rapid prototyping and building a good for minimum viable product for proof of concept

Limitations
- main limitation is the ability to handle the complexity of large texts and the overall context of the reviews

Model	Metric	1	2	3	4
Stopwords and punctuation		Retained	Retained	Removed	Removed
Additional Features		Included	Excluded	Included	Excluded
Naive Bayes	Precision	0.87	0.87	0.86	0.86
Logistic Regression		0.90	0.90	0.88	0.88
Naive Bayes	Recall	0.79	0.79	0.79	0.79
Logistic Regression		0.87	0.87	0.84	0.84
Naive Bayes	F1 Score	0.82	0.82	0.82	0.82
Logistic Regression		0.88	0.88	0.85	0.85

Naïve Bayes Model

Overall sentiment

How Naïve Bayes Model predicts Overall Sentiment

- Probabilistic algorithm by measuring the relative importance of each word
- Naïve refers to the assumption that the features in the dataset and independent of each other

Model performance and Advantage/Disadvantages

- Performs slightly worse than Logistic Regression model in all the performance metrics
- Similar advantages (simplicity) as the logistic regression model
- However, model is not able to analyze complex nuances and context within the data
- Naïve Bayes cannot measure the relationship between keywords, e.g. The word "good" would add a positive sentiment prediction value even if it has the word "not" in front

Model	Metric	1	2	3	4
Stopwords and punctuation		Retained	Retained	Removed	Removed
Additional Features		Included	Excluded	Included	Excluded
Naive Bayes	Precision	0.87	0.87	0.86	0.86
Logistic Regression		0.90	0.90	0.88	0.88
Naive Bayes	Recall	0.79	0.79	0.79	0.79
Logistic Regression		0.87	0.87	0.84	0.84
Naive Bayes	F1 Score	0.82	0.82	0.82	0.82
Logistic Regression		0.88	0.88	0.85	0.85

BERT Models

Overall sentiment

Advantages
- Transformer models can address the limitations of RNNs, such as difficulty in capturing long-term dependencies
- Transformers can analyze the sequence of words because of the attention mechanism, which considers inputs before and after each word, enabling better context comprehension
- These are pre-trained models on extensive corpus

Limitations
- Intensive computing requirement for training and validation of transformer models, even when using a relatively faster pre-trained model such as DilstilBERT and DistilRoBERTa, which results in difficulties for hyperparameter optimization
- Specifically, the training and validation of these models took around 1.5 hours with extra computational units purchased on google colab

Model Training
- Target: Sentiment Column
- Feature: Summary Text
- Train-Validation-Test Split: 80-10-10
- Tokenizers: Tokenizers from the respective Hugging Face Transformer Library were used
- Max Length for Tokens: 512
- Finetuning: Last 2 layers of the models

BERT Models

Overall sentiment

Model performance

- RoBERTa model performing slightly better
- scores ranging from 86-89% across all metrics and data partitions also suggest the models adjust efficiently to the imbalance of the dataset

	DistilBERT	RoBERTa	DistilRoBERTa
Precision	0.86	0.88	0.88
Recall	0.87	0.89	0.89
F1 Score	0.87	0.88	0.88

Performance on Validation Data

	DistilBERT as Tokenizer (Logistic Regression)	DistilBERT	RoBERTa	DistilRoBERTa
Precision	0.85	0.88	0.89	0.86
Recall	0.87	0.86	0.89	0.88
F1 Score	0.85	0.87	0.87	0.88

Performance on Test Data

Multi-head Self-Attention

Overall sentiment

Advantages

- Ability to gain a deeper insight into the text by analyzing reviews holistically compared to the other transformer models
- Model can analyze different nuances of the same word simultaneously, providing deeper context on the possible sentiments of the text

Limitations

- Complexity and training times of the model increase quadratically with the size of the data and took around 1 hour with extra computational units on google colab
- Like the previous transformer models, computation power is a main challenge as well as figuring out the optimal number of hyperparameters and epochs due to limited GPU resources

Model Training

- Target: Sentiment Column
- Feature: Summary Text
- Train-Validation-Test Split: 80-10-10
- Tokenizers: Keras Tokenizer
- Number of Heads: 2
- Max Length: 562

Multi-head Self-Attention

Overall sentiment

Model performance

- Clear improvement over all other transformer models, yielding the best precision, recall and F1 score

Predictions on 2012 Data

- Tested the model performance on top two most popular products from the 2012 data: Quaker Oatmeal Cookie and Energy Drink
- Model's performance decreased notably in scenarios where there was a higher prevalence of negative (-1) and neutral (0) classes in Energy Drink dataset

	Test Data
Precision	0.90
Recall	0.91
F1 Score	0.91

Performance on 2011 Test Data

	Oatmeal Cookies	Energy Drinks
Precision	0.92	0.67
Recall	0.93	0.66
F1 Score	0.92	0.66

Performance on 2012 Data

Selecting Multi-head Attention as best model for overall sentiment analysis

Multi-head attention model is the superior model

due to its ability to analyze data from different perspectives,

allowing the model to assign different weights depending on the relevance of the relationship and filtering out noisy data

Comparison of Multi-head Attention model and ChatGPT 4 showed ...

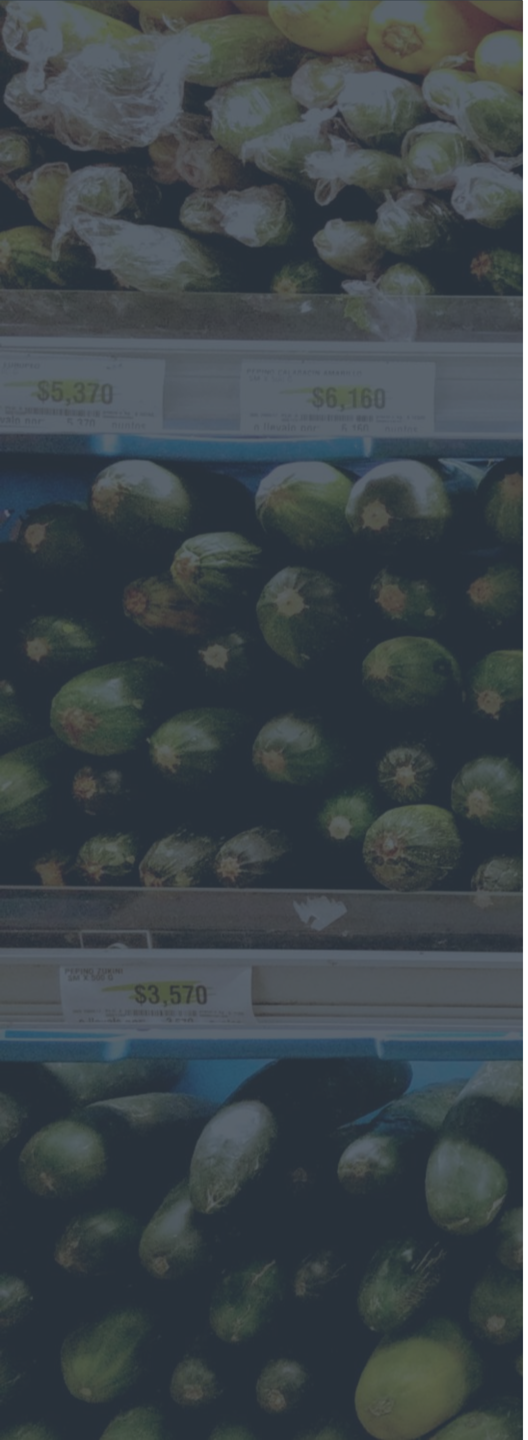
Making predictions on the 2011 Test Data, Multi-head attention model outperformed ChatGPT4, achieving higher precision, recall, and F1 score

Input the review data into our well-trained multi-head attention model

Imported a csv file containing a column of reviews without ground-truth labels to ChatGPT4 and the prompt used was:: *Pretend you are a data scientist manually labeling data points for sentiment analysis based on text reviews. For each review in the attached file, determine if the sentiment of the review is positive, negative or neutral. Return the output file with the sentiment column filled in.*

	Multi-head Attention	ChatGPT 4
Precision	0.90	0.68
Recall	0.91	0.84
F1 Score	0.91	0.72

Performance of Multi-head Attention Model vs ChatGPT 4 Sentiment Analysis



Outline of this presentation

Spacy Aspect Classifier

Aspect-based sentiment

Methodology:
- Utilize SpaCy package for extracting aspect terms from reviews
- Apply classifier to determine aspect class based on predefined Key Performance Criteria (KPC) for each review

→ Offer valuable insights into specific KPCs mentioned within reviews

→ Provide insights across different sentiment classes, aiding in understanding customer feedback

Advantage: Ability to predefine aspect class based on relevant KPCs for a product

Limitation: Inability to predict sentiment for each KPC, especially when multiple KPCs are identified within the same review

*Leverage multi-head attention model for **overall sentiment prediction***

↓

*Identify **improvement areas** within negative sentiment class **for top 2 products purchased in 2012***

aspect	-1	0	1
delivery	4	5	148
health	1	1	29
packaging	8	13	274
portion	7	14	150
price	1	3	69
quality	10	10	193
taste	18	18	589
texture	11	20	335

Fig. 1 Number 1 Product - Oatmeal Cookie

aspect	-1	0	1
delivery	3	16	23
health	13	13	24
packaging	37	98	147
price	1	6	14
quality	41	83	128
taste	57	123	182

Fig. 2. Number 2 Product - Energy Drink

PyABSA Transformer

Aspect-based sentiment

What is PyABSA?

- Modularized framework built on PyTorch for reproducible Aspect-Based Sentiment Analysis (ABSA)
- Offers granular insights into sentiment related to specific aspects of a product

Key Features:

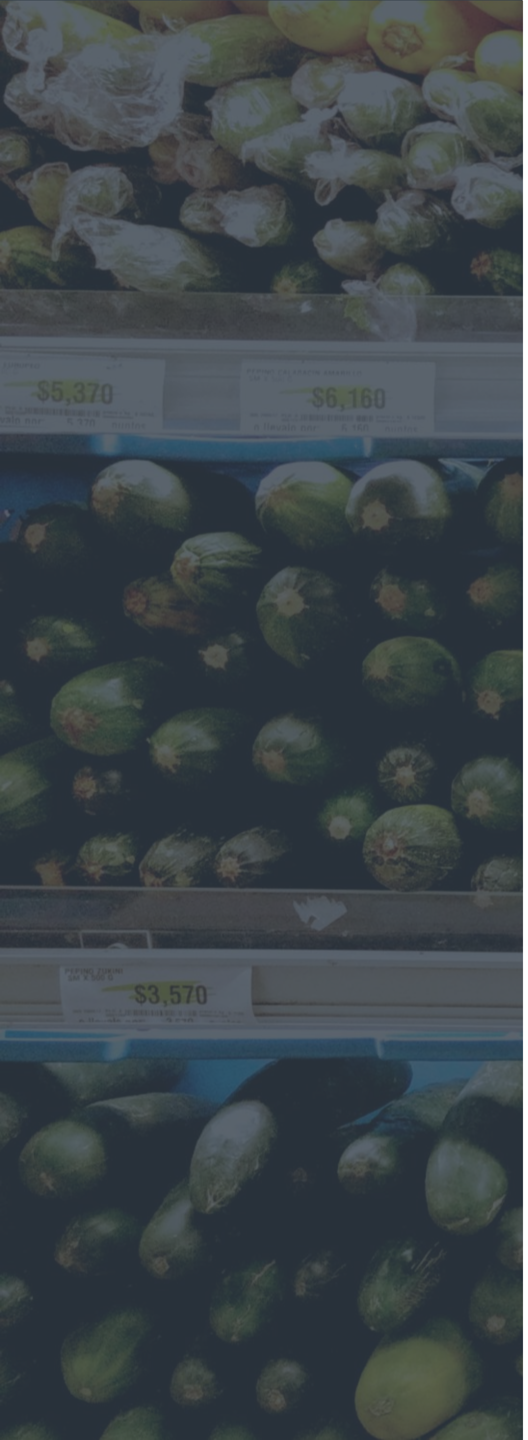
- **Reproducible:** incorporates various models incl. attention-based, graph-based, and BERT-based models and utilizes diverse datasets such as Yelp, MMAs, Restaurant14, and MOOCs data for robust ABSA tasks
- **Data Scarcity Handling in Training:** eliminates concerns about data scarcity, especially regarding absent ground-truth labels for aspects in the dataset

Benefits

- Offers developers a robust solution for handling various ABSA tasks effectively
- Enhances developers' understanding of ABSA performance through unsupervised learning approaches

*Example: PyABSA **extracts aspects** from sentences and performs **sentiment analysis for each***

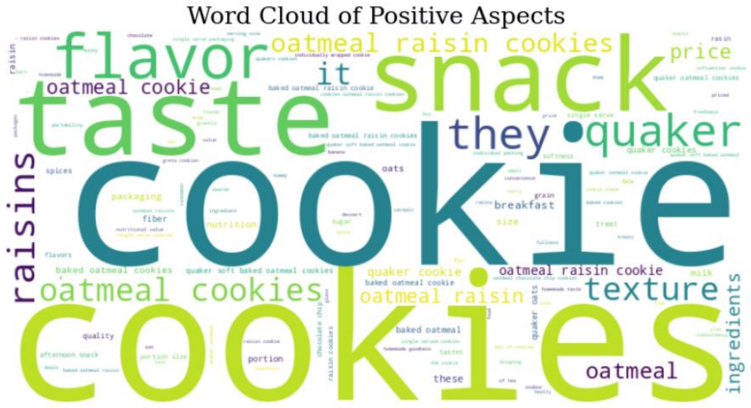
	sentence	Predictions	tokens	aspect	sentiment	probs	confidence
86	not a big fan . . . flavor for kids , but too ...	0	['not', 'a', 'big', 'fan', '.', '.', '.', 'fla...	[mix, water, mix, caffeine]	['Negative', 'Negative', 'Negative', 'Negative']	[[0.9807977080345154, 0.011564690619707108, 0....	[0.9808, 0.9938, 0.9808, 0.9945]
428	too confusing " a squeeze " ? how do i know ho...	-1	['too', 'confusing', '"', 'a', 'squeeze', '"',...	[bottle, cup, taste, flavor]	['Negative', 'Negative', 'Negative', 'Negative']	[[0.9970003962516785, 0.002317150356248021, 0....	[0.997, 0.9953, 0.9925, 0.7394]
51	surprisingly wonderful drink additive . . . as...	1	['surprisingly', 'wonderful', 'drink', 'additi...	[drink]	['Positive']	[[0.005293132737278938, 0.0003364187723491341,...	[0.9944]
120	average taste and pep vitamin squeeze isn ' t ...	0	['average', 'taste', 'and', 'pep', 'vitamin', ...	[water, caffeine]	['Negative', 'Positive']	[[0.9904863238334656, 0.008949346840381622, 0....	[0.9905, 0.9698]

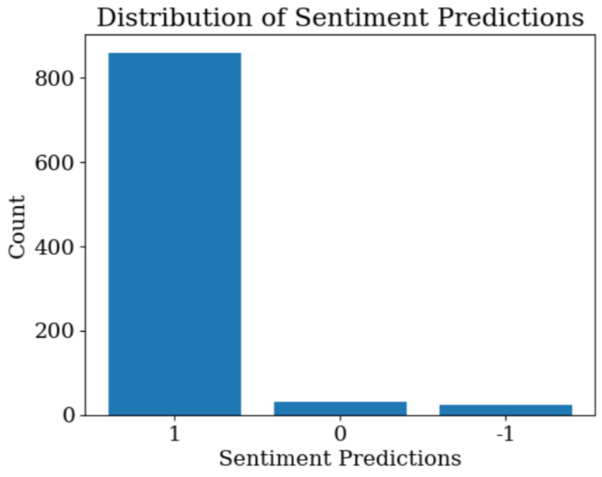


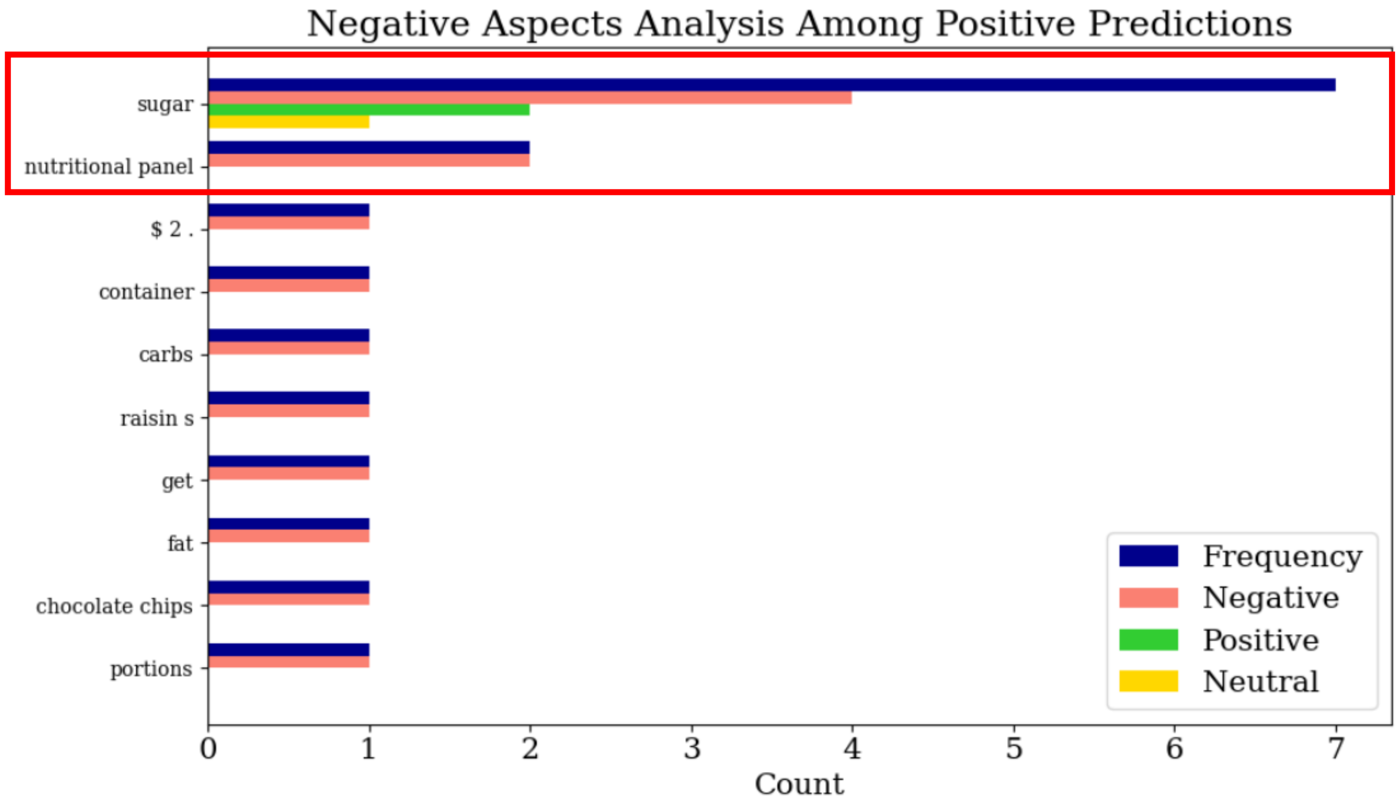
Outline of this presentation

Cookie’s negative aspects regarding sugar and nutritional information

Negative aspects within **positive** sentiment from PyABSA

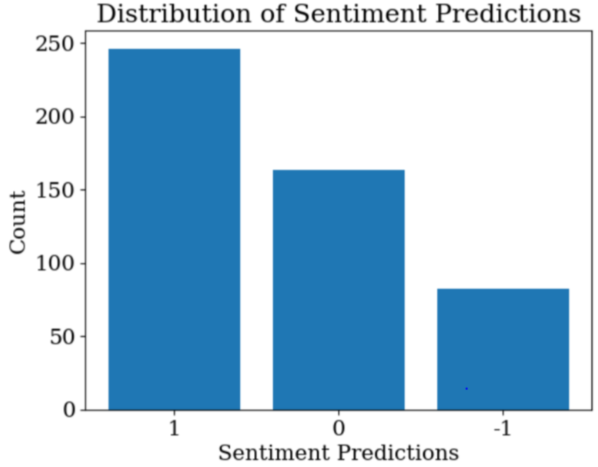


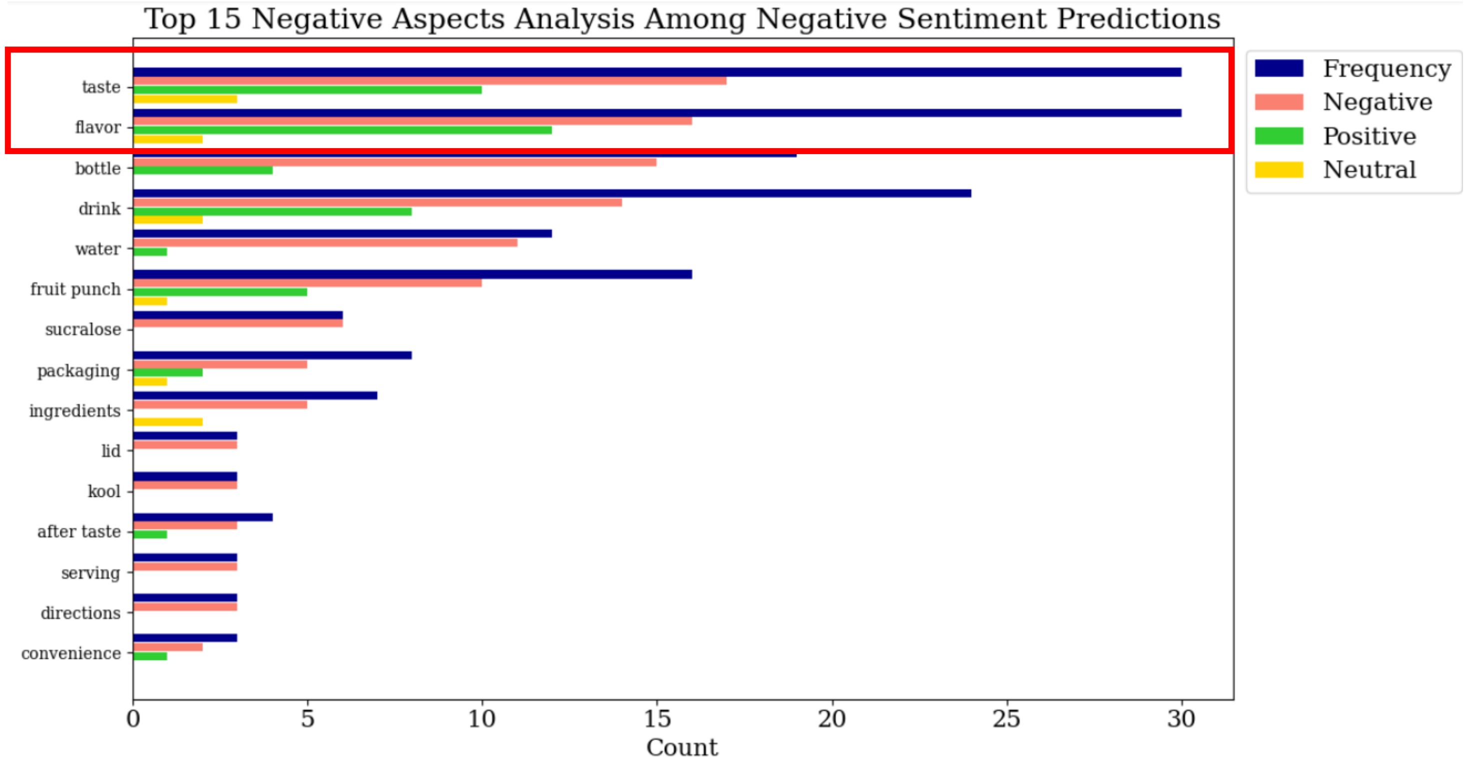




Flavoured drinks negative aspects regarding taste and flavour

Negative aspects within **negative** sentiment from PyABSA





Model suitability by Resource Constraint

Presenting different models as menu options for businesses to pick and choose based on sentiment requirements and resource availability

Overall sentiment analysis			Aspect-based sentiment analysis	
Logistic Regression	RoBERTa Transformer	DistilBERT Transfomer	Spacy Aspect Classifier	PyABSA Transformer
Naïve Bayes	ChatGPT 4	Multi-head Attention		
Simple models	**Zero-code**	**Fine-tuned**	**KPC defined**	**KPC not defined**

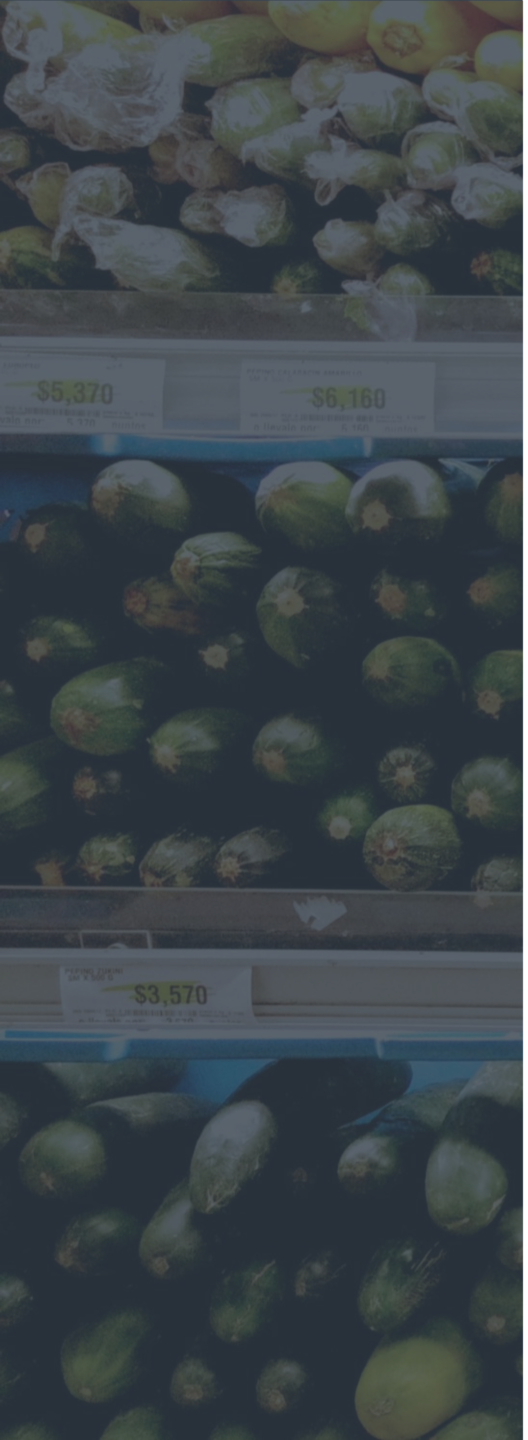
Increasing complexity level →

Model suitability by Resource Constraint

Presenting different models as menu options for businesses to pick and choose based on sentiment requirements and resource availability

Overall sentiment analysis			Aspect-based sentiment analysis	
Logistic Regression	RoBERTa Transformer	DistilBERT Transfomer	Spacy Aspect Classifier	PyABSA Transformer
Naïve Bayes	ChatGPT 4	Multi-head Attention		
3-hour solution	**30-minute solution**	**3-day solution**	**Business-led**	**Unsupervised**

Increasing complexity level →



Outline of this presentation

01
Importance of customer sentiment analysis

02
Objective of our project

03
Overview of methodology

04
Deep-dive into dataset & unstructured data

05
Data cleaning and preprocessing

06
Overall customer sentiment analysis

07
Aspect-based sentiment analysis

08
Discussion of insights and model performances

09
Conclusion and future direction

Future works – Building the best model for ABSA

For important use cases that warrant large resource allocation

01 Data labeling

- KPCs differ by business use case
- Ground truth needed for each KPC

02 Finetuning

- Train model based on defined KPCs
- Ensures model output is relevant and KPC exhaustive

03 Iteration

- AB testing on users or user labeling
- Model drift and refreshing

Conclusion – Reccomendations for Implementation

Overall sentiment analysis			Aspect-based sentiment analysis	
1	2	3	4	5
Zero code solution 30-minute solution	**Simple model 3-hour solution**	**Fine tuned 3-day solution**	**Pre-trained with KPC defined**	**Finetune with KPC defined**

Thank you!

Team members:

Alex Chen Chen (E1148800)
Laura Ngoc Ha Do (E1148768)
Lim Ciwen Brendan (E0540070)
Toh Yu Qi Chermaine (E0201966)
Wong Cheuk Wah (E1148763)