Aspect Based Review Analysis

Chattoraj, Ankuran chattoraj.a@northeastern.edu

Abstract

With the current and growing importance of E-commerce, this project conducts sentiments with Topic modeling in aspect analysis using the Amazon product review dataset.
This project contains key steps: preprocessing of JSON formatted review data, sentimental analysis with supervised methods; BERT and LSTM, then sentiment polarity analysis with topic modeling, and comprehensive evaluation. As a result of our research we were able to identify sentiments of overall text reviews and extract topics from each review to identify sentiments of individual topics. This approach assists marketing research needs in identifying the overall attitude of customers to each product along with attitude towards individual aspects of a product.

Introduction

Compared to the past when people went to the store and bought products, people spent more money on online shopping. As of 2023, the research E-commerce market size is estimated to be USD 16.27 Trillion and expected to be USD 57.22 Trillion by 2032 [1]. In this era of E-commerce, companies should catch customer reactions fast to make more products or distribute them efficiently to maximize profit. More importantly, to lower the cost of loss, catching customers' negative reactions and reflecting them - such as customer service and product improvement - is key to the market strategy. Therefore, we decided to work on a NLP project related to reviews in E-commerce.

In our project, we have multi-step objectives. First, we will analyze customers' sentiments toward the products with Amazon customer reviews. Then, we will extract aspects from product reviews and identify customers' sentiments based on these aspects.

Our approach is centered around leveraging advanced aspect-based sentiment analysis to understand the average rating of products, enhancing the depth and accuracy of rating predictions. We begin by grouping reviews for each product and creating product-specific datasets to ensure a targeted analysis. Next, we employ techniques like dependency parsing and named entity recognition to extract the key aspects mentioned in these reviews. These aspects represent the specific attributes or features considered by reviewers, offering a comprehensive view of product evaluation.

Subsequently, a sentiment analysis phase is implemented for each aspect, determining the sentiment—whether positive, negative, or neutral—associated with it. This approach ensures that we capture both the strengths and weaknesses of products, providing a balanced

perspective. The sentiments are then numerically encoded, forming the foundation for constructing a predictive model. This model is designed to estimate the average rating of a product based on the numeric sentiment representations, offering a more nuanced and informative rating prediction process.
By delving into the aspects and sentiments within product reviews, our approach enriches traditional rating predictions. It empowers consumers with detailed insights to make well-informed decisions and provides businesses with valuable feedback to enhance their products and customer satisfaction.

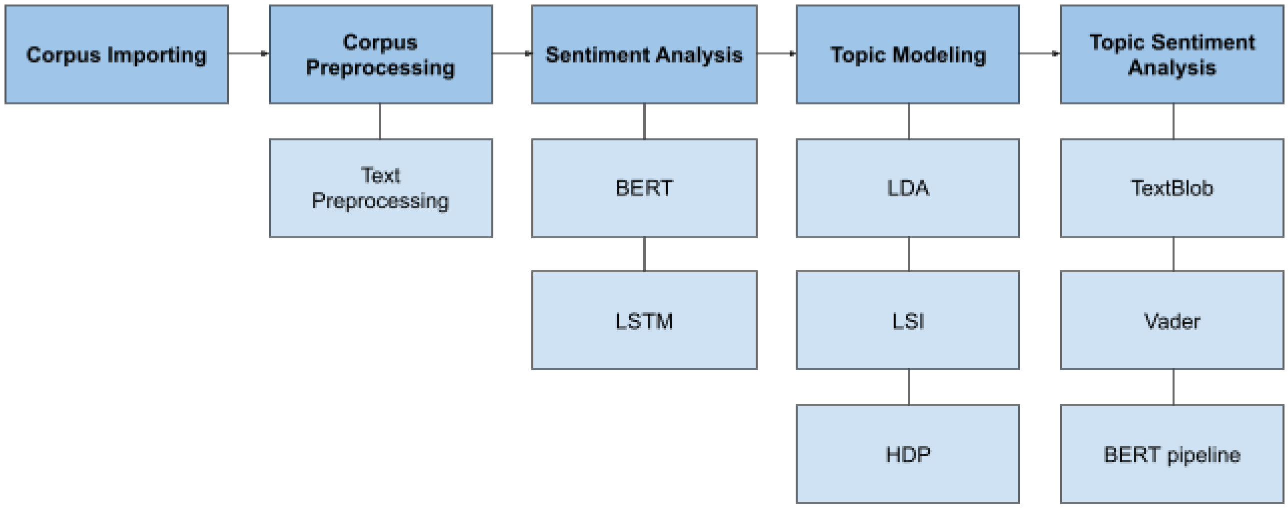
The Amazon Review Data [6], features Amazon reviews from May 1996 to October 2018. It's balanced across 5-star ratings, with 233.1 million reviews overall. Records include review text, descriptions, category information, price, brand, and image features. Data is separated across 29 categories including Clothing, Appliance, Beauty, Software, etc.

Background

Sentiment analysis has been used for analyzing reviews for a long time. Xing Fang and Justin Zhan[2] used sentence-level categorization and review-level categorization using online product review corpus from Amazon. Their research focused on sentiment category polarization and obtained promising results in both levels of categorization but had limitations in relying on sentiment tokens for reviews that contain implicit sentiments. The categorization of sentiment polarity is a fundamental problem in sentiment analysis [3]. Ligthart, Catal and Tekinerdogan[4] observed different machine learning techniques for sentiment analysis and came to a conclusion that LSTM and CNN “are the most used deep learning algorithms for sentiment analysis”. In aspect-based sentiment analysis (ABSA), data is categorized by aspect and sentiment is assigned to each. By associating specific sentiments with different aspects of a product or service, aspect-based sentiment analysis can be used to analyze customer feedback. ABSA using deep learning techniques is still in its initial stage and further exploration of reviews and information from external sources is needed to improve sentiment analysis.
With the popularity of social platforms, the focus of sentiment analysis has shifted from English to multilingualism[5] and many researchers applied deep-learning as well as machine learning to extract sentiments.

Approach

In our project we wanted to build a system that would assist the marketing team in identifying what products customers like or dislike and what aspects specifically they liked or disliked. This approach helps to understand target audience: what is important for your customer, what needs to be improved in products that customers did not favor etc. To approach this problem we firstly start with sentiment analysis on full text of reviews left by customers and then, by separating each review in corpus sentence by sentence, we apply topic extraction techniques to identify topics within each category and align each sentence with topic. After that each sentence and topic is assigned a sentiment using unsupervised modeling technique.



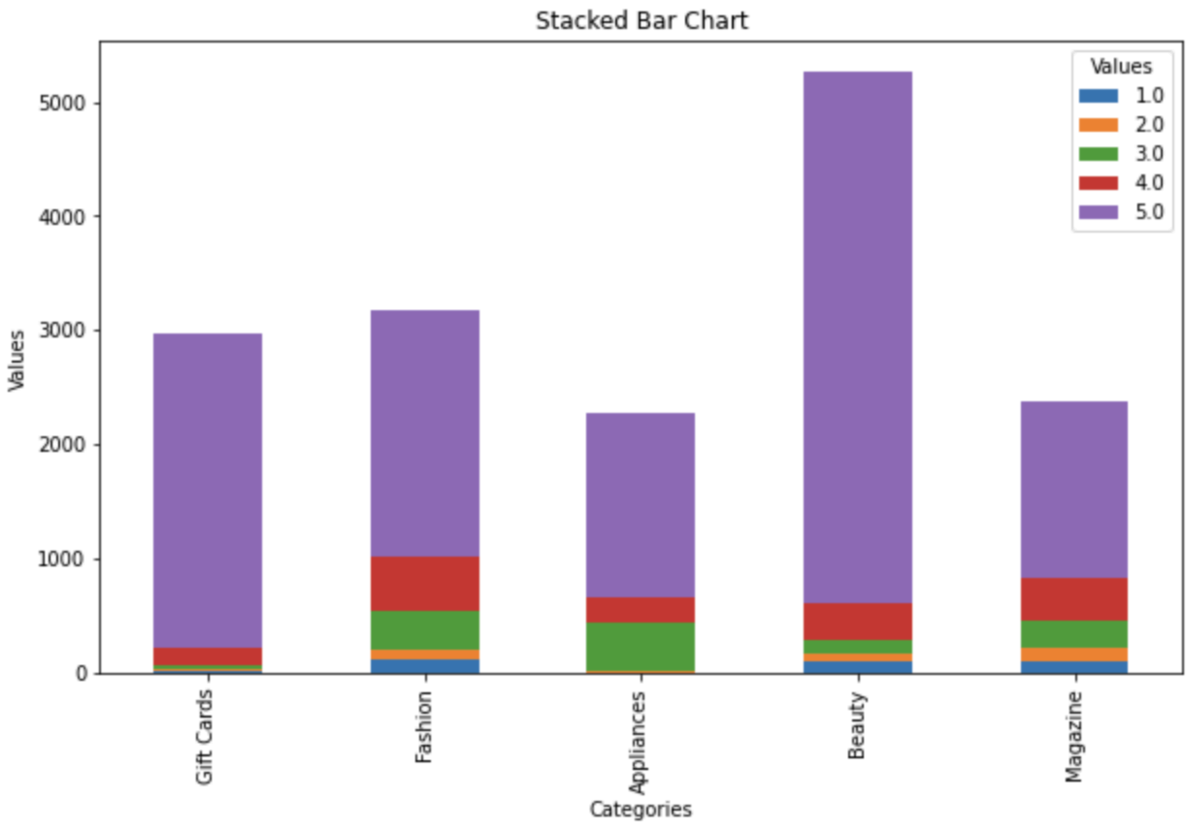
Corpus Importing

Our corpus is stored in JSON format, therefore after uploading our dataset in Python notebook, we defined a function to read each line in the document and store it as a list of dictionaries, which was later converted to dataframe format.

—---

Corpus Preprocessing

We firstly worked with the dataset itself and then started working on text reviews. Full Amazon Reviews dataset consists of around 30 categories of products, but in our project we only focused on categories: Appliances, Fashion, Beauty, Gift Cards and Magazine Subscriptions. Each category consists of around 2~3 thousand rows with columns for overall rating of product, product ID, reviewer ID, review text, review title, style of product, summary and review time. As you can observe on the graph most reviews in all categories are rated as 5.



Corpus preprocessing:

We removed duplicates from our dataset and dropped all rows that contained missing text in the review column. After that unnecessary columns were dropped and only necessary columns remained for further analysis: ‘overall’ (1-5 star rating column), ‘reviewerID’, ‘asin’ (unique product id), ‘reviewText’, ‘summary’. For our topic analysis we extracted sentences in each review in a separate row. This way it is easier to extract topics as each sentence is focused on one main idea.

Review text preprocessing:

Using RegEx we removed punctuation, converted text to lowercase characters, and removed non-alphabetic characters. To remove stop words from our text review, we needed to modify stop word list as simply applying full list would remove ‘no’ and ‘not’ words which mean negative sentiment We removed these negative particles from the stopword list and corrected spelling of reviews using Levenshtein distance since reviews are being written by customers all over the world, spelling of words in English might sometimes be grammatically incorrect (‘love’ can be written as ‘luv’, ‘comfortable’ written as ‘comfy’, etc.). After that we applied lemmatizer for our text reviews.

—---

Supervised Learning

To approach the sentiment analysis we implemented two models: Bidirectional Encoder Representations from Transformers (BERT) and Long Short Term Memory Recurrent Neural Network.

BERT:

BERT model is a natural language processing model introduced by Google in 2018. It is a transformer-based model designed to understand the context of words in a sentence by considering the surrounding words, both to the left and right. BERT has significantly advanced the state-of-the-art in various NLP tasks, such as question answering, sentiment analysis, and named entity recognition.

Before applying the model, we converted our target variable from 1 to 5 star rating to 2 class sentiments: positive and negative. To apply BERT we used Tensorflow, tokenized our text using BertTokenizer, and converted reviews to tensors. Adam optimizer was used along with 2 epochs batch size of 16.

BERT model showed better convergence for some categories to higher accuracy scores but takes a lot of computation time.

LSTM:

LSTM(Long Short-Term Memory) is a deep learning system that avoids the vanishing gradient problem as a type of recurrent neural network (RNN). LSTM allows us to learn and remember over long sequences, making it well-suited for tasks such as natural language processing (NLP).

Before applying the model, in the same way, we used star rating as the target variable and changed it into 2 classes of sentiments: positive and negative. Before applying the LSTM model we applied Tokenizer in the Keras preprocessing library.

For the model implementation, we used the Keras sequential model in TensorFlow, with 64 units, 0.2 rate of drop-out, and 5 epochs batch size of 32 with Adam optimizer.

LSTM model showed better convergence for other categories and it was computationally more efficient than BERT model.

—---

Topic Modeling

Within the domain of Natural Language Processing (NLP), topic modeling emerges as a robust methodology for revealing latent thematic structures inherent in a textual corpus. Our primary research objective during this phase was to conduct an in-depth examination of customer reviews, with the overarching aim of discerning the underlying factors influencing product ratings.

Techniques for Topic Extraction: LDA, LSI, HDP

The application of unsupervised topic modeling techniques, notably Latent Dirichlet Allocation (LDA), Latent Semantic Indexing (LSI), and Hierarchical Dirichlet Process (HDP), played a pivotal role in extracting substantial topics from the expansive collection of reviews. These techniques enable the identification of latent themes without relying on labeled data, fostering an unbiased exploration of customer sentiments.

To optimize the effectiveness of our models, we implemented Bigrams and executed the removal of low-value words through Term Frequency-Inverse Document Frequency (TF-idf). This preprocessing step aimed to refine the topics generated by capturing nuanced word relationships and eliminating less informative terms.

Evaluation of our topic models utilized coherence scores, providing a quantitative measure of interpretability and clarity in the identified topics. Noteworthy was the HDP model, demonstrating a proclivity for returning to a higher number of topics while maintaining notable coherence. Strategic parameter adjustments in the LDA model yielded coherence levels comparable to our LSI model. Interestingly, within specific product categories, our meticulously tuned LDA model outperformed LSI, underscoring its versatility in capturing nuanced topic structures.

Similarity Analysis

Within our investigative framework, we employed similarity analysis as a methodological approach to identify prevailing topics discussed within sentences. This analytical technique facilitated the discernment of intricate relationships and connections between different topics, thereby illuminating the complex thematic tapestry woven within customer reviews. The application of similarity analysis served as a conduit for unraveling inherent connections between diverse subjects, ultimately providing nuanced insights into the expansive thematic landscape embedded in customer feedback.

Sentence-Level Sentiment Prediction

Expanding our analytical scope, our inquiry extended to include the prediction of sentiments at the sentence level for each identified topic within the reviews. This involved deciphering the emotional context associated with the topics discussed in customer feedback. By predicting sentiments at the sentence level, our objective was to uncover how sentiments are distributed across different facets of products. This approach contributed to a more holistic understanding of customer reactions, offering a nuanced perspective on the sentiments encapsulated within the topics discussed in the reviews.

For sentence-level sentiment prediction we applied 3 various techniques: TextBlob, Vader and roBERTa. As our sentence level corpus in an unlabeled data, we stuck with unsupervised learning models in this part.

Text Blob

TextBlob is a library that analyzes the sentiment of the given text using its pre-trained sentiment analysis model.
It measures polarity (-1 to 1 range; -1 is most negative, 0 is neutral, and 1 is most positive) and subjectivity (0 to 1 range; 0 is very objective and 1 is very subjective).

VADER

VADER (Valence Aware Dictionary and sEntiment Reasoner) tool uses bag-of-words approach and takes into account negation words as ‘not’ (which we intentionally tried to keep) to assign a negative sentiment, while also taking into account emphasize words like ‘very’, ‘little’ to adjust sentiment scores accordingly.

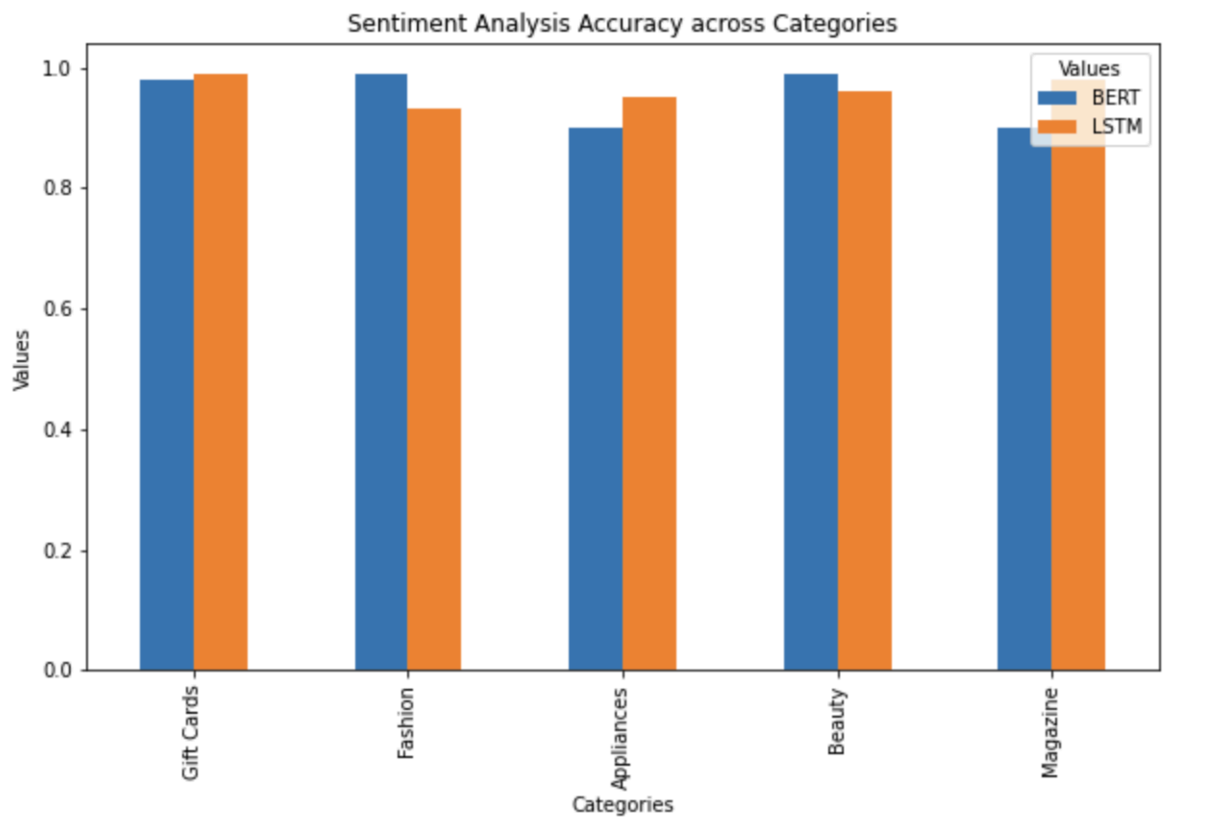
Pipeline BERT model

This is a pretrained sentiment analysis model built by ‘hugging face’. ‘This is a RoBERTa-base model trained on ~124M tweets from January 2018 to December 2021, and fine tuned for sentiment analysis with the TweetEval benchmark’.

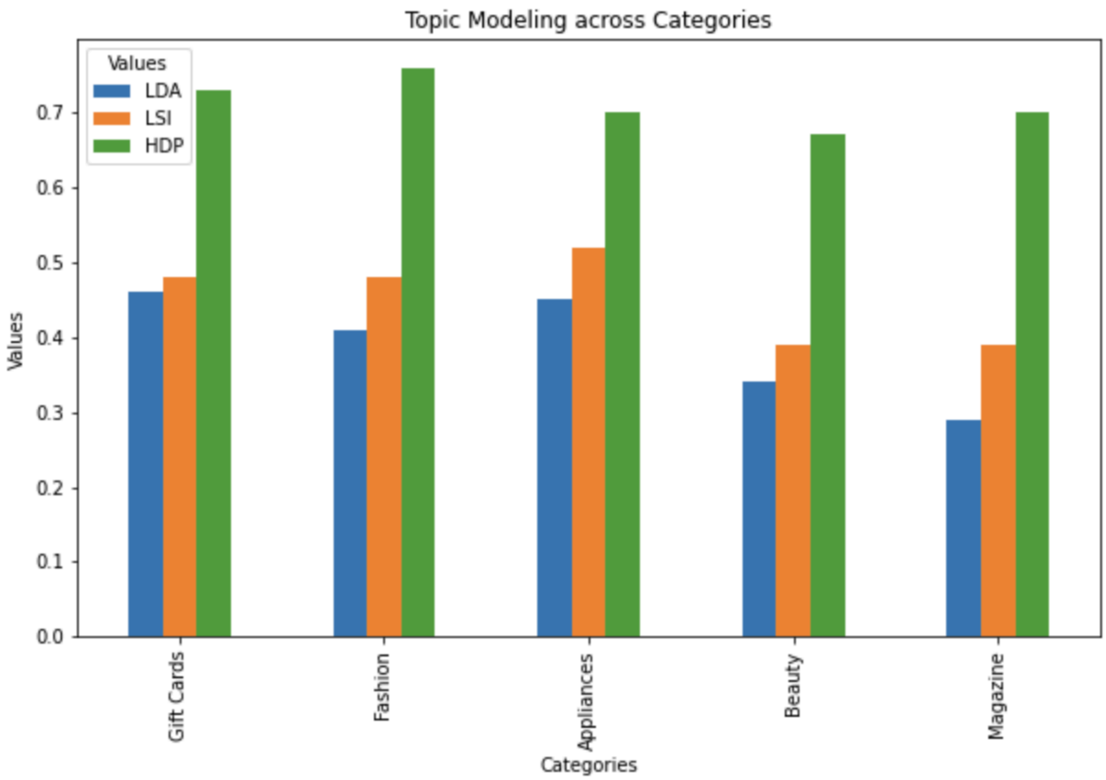
We have tried all three unsupervised sentiment analysis models, but came to the conclusion that Text Blob generates more accurate sentiments while taking smaller time to generate them.

Results

As a result of BERT and LSTM models on overall review text we have identified that LSTM provides better results for 3 out of 5 categories (Gift Cards, Appliances, Magazine Subscriptions) and takes shorter computation time, thus making LSTM better model for sentiment analysis of our corpus.



As a result of LDA, HDP and LSI models, we have decided to proceed with LDA model since it takes shorter computation time and provides clearer topics, even though it shows lower coherence score.



As a result of the LDA application, we have generated the following topics for each category and presented some examples of generated topics and sentiment analysis on sentences.

Appliances Category

Topics:

0: "Brush Kit and Usage",
1: "Product, Washer, and Shipping",
2: "Rod, Part, and Drill",
3: "Filter and Replacement",
4: "Time, Coil, and Brand",
5: "Rod, Time, and House"

Examples:

overall	reviewerID	asin	sentence	Highest_Similarity_Topic	Sentiment	Sentiment_Category
5	A1T1YSCDW0PD25	B0013DN4NI	This Hitachi may be a bit slow for drilling.	Time, Coil, and Brand	-0.3	Negative

5	A3VYM3NY7W4MX4	B0014CN8Y8	To get this done for now, I had to do half from inside the house and half from outside.	Rod, Part, and Drill	-0.111111	Negative
5	AMY6O4Z9HINO0	B0014CN8Y8	The exterior cover on the vent had been gone for some time, and the amount of leaves and pine needles plus lint wads that I got out were shocking!	Rod, Time, and House	-1	Negative
1	A3SM7H70QI3TY8	B000N6302Q	My best estimate is that the pump was still working (the water was getting pumped through) but the cooling mechanism was not.	Filter and Replacement	1	Positive
1	A3SM7H70QI3TY8	B000N6302Q	Since this product is still under warranty, I am hoping to acquire a new unit since this particular one has essentially never been used and is obviously faulty in some capacity.	Filter and Replacement	0.075758	Positive

----

Gift Cards

Topics:

0: "Gift Card Purchase",
1: "Gift Card Review and Balance",
2: "Amazon Box and Purchase Experience",
3: "Box, Amazon, and Pizza",
4: "Amazon, Like, and Number",
5: "Pizza, Box, and Coupons",
6: "Food, Like, and Perfect",
7: "Preferences, Pizza, and Number",
8: "Christmas, Tin, and Deal"

Examples:

overall	reviewerID	asin	sentence	Highest_Similarity_Topic	Sentiment	Sentiment_Category
5	A2IGSUZ4FEJWFV	B00FTGF3P2	With limited choices of restraints in there area.	Preferences, Pizza, and Number	-0.07143	Negative
5	A2CM3SWOMP3A9C	B01E4QPDV6	Who but Donald Trump could hate anything Amazon?	Preferences, Pizza, and Number	-0.8	Negative

5	A1INHPEXMKFOGR	B01FERR9FW	Books are so expensive.	Preferences, Pizza, and Number	-0.5	Negative
4	A1ISQ4BFCCGKVY	B00BXLT1UA	Every day she asks if she can go build her bear, but it's just been too cold.	Gift Card Review and Balance	-0.6	Negative
4	A3T3BD40FHQFYK	B00PGOMSU0	The "convenience fee" is annoying if you're not buying tickets for a movie you actually expect to be sold out, like on a premier night or something.	Gift Card Review and Balance	-0.4	Negative
1	A1U1G73EI5IRZF	B00PG8502O	Upon arrival at a local Olive Garden branch in October 2014, my friends and I saw only a few customers sit there so we should be assigned a seat quickly.	Preferences, Pizza, and Number	0.16666666666666666	Positive
2	A3RUNGX8PZ8VYO	B00BXLVE6Y	Once I managed to enter it right, after having to call it back twice, I was able to check the balance and see that the card was loaded.	Preferences, Pizza, and Number	0.2619047619047619	Positive

----

Magazine Subscriptions

Topics:

0: "Magazine Content and Reading Experience",
1: "Recipes and Cooking",
2: "Money, Home, and Information",
3: "Reading, Weekly, and Star",
4: "Magazine Pages and Content",
5: "Esquire Magazine and Content",
6: "Money, Esquire, and Property",
7: "Magazine Subscription and Advertising",
8: "Home Ideas and Dwell Magazine",
9: "Automobile and Car Magazines",

Examples:

overall	**reviewerID**	**asin**	**sentence**	**Highest_Similarity_Topic**	**Sentiment**	**Sentiment_Category**

5	A281NPSIMI1C2R	B00005N7RD	I buy this for my husband because he never thinks to buy it for himself since he hates shopping.	Reading, Weekly, and Star	-0.8	Negative
5	A24ZEH6X4AJNR4	B00005N7PG	I almost gave my subscription up because with physical disabilities it's unlikely I can do a lot of travel.	Money, Home, and Information	-0.25	Negative
5	A309PFW8T6DMMZ	B00005N7QW	This stuff is not great for clipping out and saving because it is so thin.	Home Ideas and Dwell Magazine	-0.4	Negative
5	A2877WXAPQ7T50	B00005N7TL	One complaint would be the reveiws and advertisements are too pricey for 95% of americans.	Automobile and Car Magazines	-0.15	Negative
5	AMVC9WTXYKNJ1	B00005N7SC	Unfortunately, I would bet that if you looked at the circulation figures, half the title is wrong.	Magazine Content and Reading Experience	-0.38889	Negative
1	A30JPZ9TZ7I61U	B00005NIOH	But the last year or two they have been on track with solid stories and spectacular photography.	Automobile and Car Magazines	0.1999999999999998	Positive
1	A16PDB0CIYNAMC	B00007AVYI	The staff worked well together and they put out a consistently good magazine.	Automobile and Car Magazines	0.7	Positive
1	A1JXRXYVSUAW6I	B00005N7TL	It's a diabolically entertaining techie/consumer delight.	Recipes and Cooking	0.5	Positive
2	A2N1WT1EBTWTBV	B00005N7QN	But great price.	Money, Esquire, and Property	0.8	Positive

----

Fashion

Topics:

0: "Running Shoes Comfort and Size",
1: "Size and Comfort in Footwear",
2: "Sneaker Size, Running, and Comfort",
3: "Comfort, Lightness, and Foot Pain",
4: "Sneaker Characteristics and Feel",
5: "Footwear Pair and Training",
6: "Feel and Pair",
7: "Lightness, Feel, and Running",
8: "Running, Toe, and Color",
9: "Lightness, Weight, and Training",

Examples:

overall	reviewerID	asin	sentence	Highest_Similarity_Topic	Sentiment	Sentiment_Category
5	AW8UBYMNJ894V	B000V0IBDM	Remember to replace them every 3 months if you wear them every day, especially if you weigh more than average and/or run.	Comfort, Lightness, and Foot Pain	-0.075	Negative
5	AXVYVUK058AM3	B0058YEJ5K	(except my home workouts, they drag pretty hard on the carpet)	Sneaker Characteristics and Feel	-0.04722	Negative
5	AOFQAZVA6Q6E7	B001IKJOLW	Considering the fact that I have wide feet, the shoes are slightly tight.	Footwear Pair and Training	-0.13929	Negative
4	A3T1O294HJJTUH	B0058YEJ5K	Not a lot of cushion so I wouldn't run long distances in them.	Feel and Pair	-0.05	Negative
2	A1UVWOQTPDSYGY	B001IKJOLW	I **loved** these shoes when I got them.	Sneaker Size, Running, and Comfort	0.7	Positive
2	A3M69Q5SNV7U0X	B0014F7B98	I bought these for gym training - weight class and dance class (body jam, Zumba) - and really wanted to like them, since I loved the color and the light weight feel.	Sneaker Size, Running, and Comfort	0.4333333333333333	Positive
2	A3M69Q5SNV7U0X	B014IBJKNO	These might be fine for someone with a narrow foot.	Sneaker Size, Running, and Comfort	0.1083333333333334	Positive

----

Beauty

Topics:

0: "Scented Soap and Hair Care",
1: "Product, Scent, and Price",
2: "Soap, Bar, and Fragrance",
3: "Nail Polish and Soap Review",
4: "Hair Product and Toothpaste",
5: "Hair Care, Shampoo, and Dryness",
6: "Product, Scent, and Brush",

7: "Hair, Shampoo, and Scalp",
8: "Product, Scent, and Skin"

Examples:

overall	reviewerID	asin	sentence	Highest_Similarity_Topic	Sentiment	Sentiment_Category
5	A3JMIBCH90HGWG	B00006L9LC	My hair was falling out due to a very dry scalp.	Hair, Shampoo, and Scalp	-0.09583	Negative
5	A2N4OUX2ORC859	B00006L9LC	Hard to get.	Soap, Bar, and Fragrance	-0.29167	Negative
5	ATCQ2K4QG03T4	B00006L9LC	So my hair always feels dry.	Hair Care, Shampoo, and Dryness	-0.06667	Negative
4	A3E5V5TSTAY3R9	B0006O10P4	Unfortunately, they stain our white washcloths.	Nail Polish and Soap Review	-0.25	Negative
2	A24W4W9E62FZP2	B00006L9LC	It does smell good and my hair feels clean.	Hair Care, Shampoo, and Dryness	0.53333333333333333	Positive
2	A3R9H6OKZHHRJD	B000LIBUBY	I was really hoping this scent would be true to it's name, and that it would smell of redcurrant and basil.	Product, Scent, and Brush	0.275	Positive
1	A3JQUR31CLU4VK	B000URXP6E	According to the rate and review I thought this would be awesome.	Hair Product and Toothpaste	1	Positive
1	A3VLTT45T236B	B019809F9Y	Williams Lectirc Shave is a great product that has served men's electric shaving needs for several decades!	Soap, Bar, and Fragrance	0.4	Positive

Conclusion

In this project, we explored the natural language processing methods in the category of supervised methods, Topic modeling, and unsupervised methods.

With supervised methods, BERT and LSTM, the result showed that while both models generate high accuracy scores (above 90%), LSTM is more efficient in computation time.
With topic modeling the LDA model was considered the best for our project due to accuracy of generated topics.
The Unsupervised methods with Textblob, roBERTa, we discovered that Text Blob generated more accurate sentiments for sentences more efficiently than other models..

In summary, our exploration into topic modeling, enriched by advanced preprocessing techniques and sentiment analysis, provided a comprehensive, multidimensional analysis of customer reviews. This comprehensive study uncovered latent themes and sentiments, furnishing valuable insights for businesses navigating the dynamic landscape of E-commerce.

In this project we aimed at understanding text reviews and how different parts of reviews can influence the rating and sentiment of customers. While we fulfilled our goal to generate topic modeling on sentence and sentiment analysis, there are some points that are still to be considered for further research:

- Class imbalance: our corpus consists mostly of positive reviews, which might result in inconsistent results towards negative reviews. Imbalance should be addressed with respective preprocessing techniques for more accurate results.
- Bag-of-words usage in embedding for topic modeling: other methods could have been used that might generate clearer topics.
- Sentence extraction for topic modeling: since reviews are being written by humans, text and style of writing is inconsistent from one customer's review to another, so one sentence might consist of more than one topic, which we have not explored in our research.
- Broader category research: in our project we were able to process five categories while corpus consists of around 30 categories

References:

[1] *E-commerce Market Size, Trends, Growth, Report By 2032*. (n.d.). https://www.precedenceresearch.com/e-commerce-market#:~:text=The%20global%20e%2Dcommerce%20market%20size%20exceeded%20USD%2014.14%20trillion,15%25%20between%202023%20and%202032.
[2] Fang, X., Zhan, J. Sentiment analysis using product review data. Journal of Big Data 2, 5 (2015). https://doi.org/10.1186/s40537-015-0015-2
[3] Pang B, Lee L (2008) Opinion mining and sentiment analysis. Found Trends Inf Retr2(1-2): 1–135.
[4] Ligthart, A., Catal, C. & Tekinerdogan, B. Systematic reviews in sentiment analysis: a tertiary study. Artif Intell Rev 54, 4997–5053 (2021). https://doi.org/10.1007/s10462-021-09973-3
[5] Zhu L, Xu M, Bao Y, Xu Y, Kong X. 2022. Deep learning for aspect-based sentiment analysis: a review. PeerJ Computer Science 8:e1044https://doi.org/10.7717/peerj-cs.1044
[6] Justifying recommendations using distantly-labeled reviews and fined-grained aspects.

Jianmo Ni, Jiacheng Li, Julian McAuley. Empirical Methods in Natural Language Processing (EMNLP), 2019
[7] cardiffnlp/twitter-roberta-base-sentiment-latest · Hugging Face. (2001, June 1). https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment-latest