Explore Women's Clothing Preferences and Sentiment to Drive Customer Satisfaction

Executive summary

In this project, we use text mining methods to analyze customer review data. From topic modeling, the most common topic across clothing departments is "Fit", and in different departments, customers have different focuses. In sentiment analysis, SVM model has a better performance with accuracy rate of 0.79. After combining the two methods, results show that for most of the topics in each department, they obtain a "positive" score higher than 0.5, except for "Overall Look and Back Design" from "Dress" department and "Quality and Design Feedback" under "Tops" department. Our project can help e-commerce sellers to refine their products, targeted at in the crucial aspects mentioned above and improve customer satisfaction.

Project objectives

The goal of this project is to help sellers of clothes to improve customer satisfaction by mining customer review topics and sentiments. Focus on clothing departments to specifically analyze detailed customer opinions on products, we plan to analyze factors which drive customers' satisfaction. Also, the project aims to identify emotional feedback and whether a comment is positive/negative/neutral, digging into the reasons for negative comments.

Data description

The dataset used in the research is the real world and anonymous data from Kaggle.
Source: https://www.kaggle.com/datasets/nicapotato/womens-ecommerce-clothing-reviews.
The dataset contains 23486 rows and 10 feature variables. Every row represents a client evaluation and contains the variables, and the description of each attribute. The variables are Clothing ID, Age, Title, Review Text, Rating, Recommended or not, Positive Feedback Count, Division Name, Department Name and Class Name. For more detailed introduction, descriptive analysis and visuals of the dataset, see Appendix Table 1, and Figure 1-4.

Methodology

1 Topic Modeling

1) Data Splitting: We imported libraries, loaded women's clothing review data, and grouped reviews by department for separate modeling.
2) Data Preprocessing: Next, we preprocessed the text data by tokenizing, removing stop words, and stemming the words. This helped to simplify the text data and improve the quality of the models.
3) Find Optimal Number of Topics: We then trained LDA models for each department using a fixed number of topics, and evaluated the models using the perplexity and coherence scores with the limit 5. These scores helped us to determine the optimal number of topics for each department.
4) Retraining LDA Model: We then retrained the LDA models using the optimal number of topics, and printed out the top words for each topic in each department. This helped us to understand the main themes and topics within each department. and we used ChatGPT to name each topic.

2 Sentiment Analysis

We applied sentiment analysis using an unsupervised model: VADER method; and a supervised model: SVM to gain insights of customers' sentiment towards products based on the reviews of different topics in different groups.

2.1 VADER

1) Create "polarity": First, we create a new column called "polarity" based on the 'rating' column. We labeled the 'Rating' with 1 or 2 as negative, 3 are labeled as neutral, and 4 or 5 are labeled as positive.
2) Split Datasets: Next, we split the data into 80% training data and 20% test datasets.
3) VADER lexicon model: Then we use the VADER lexicon method, which assigns the sentiment scores to each of the words in the text with 'positive', 'negative', 'netural', and 'compound' scores.
4) Find the optimal threshold: We adjust the threshold parameter by plotting out the accuracy rate of sentiment polarity prediction against different threshold values for VADER scores. We finally choose the threshold value that gives us the highest accuracy rate.

5) Retrain VADER lexicon model: We retrain the model and get the sentiment score, which helps us to get a deeper understanding of customers' attitude towards clothes in different departments.
6) Model Evaluation: We assessed the model's accuracy for sentiment score prediction using the VADER lexicon and compared it to the labeled sentiment polarity. Additionally, we utilized the confusion matrix for precision and recall rates, and VADER scores for actual negative and positive reviews.

2.2 SVM

1) Create labels: To perform the supervised model, we first need to create labels by mapping ratings to sentiment. Since the original dataset is unbalanced, with over 70% ratings above 4. Thus, we create the label of sentiment based on the 'rating' column. We labeled the 'Rating' with 1 or 2 as negative, 3 or 4 as neutral, and 5 as positive.
2) Train SVM model: Given labels in the data, we train our own supervised linear SVM classification model, and get the predicted sentiment score for all review texts in the dataset. We add the predicted sentiment score to a new column 'SentimentScore'.
3) Evaluate the model: Finally, we evaluate the model performance using metrics such as accuracy rate, precision, recall, and F1-score.

3 Synthesis of Topic Modeling and Sentiment Analysis

With our goal to predict customers' sentiments toward clothes under each department, we want to combine the results from both topic modeling and sentiment analysis.
We compute the predicted sentiment score for each topic under each department based on the SVM model we developed in the sentiment analysis part. We display the sentiment score as the probability that one topic turns out to be "positive"(it should be noted that the "positive" here is not the one that we used to define labels for SVM, but predicted by the SVM model).

Results and Discussion

1 Topic Modeling

The topics we extract from topic modeling and named by ChatGPT are as follows:

Table 1 Topics

Department name	Topics	Department name	Topics
Bottoms	1: Fit and Comfortable Jeans and Pants	Jackets	1: Comfortable and Fitting Jackets
	2: Skirts in Various Colors and Fabrics		2: Coats and Vests for Different Occasions
Dresses	1: Sizing and Fit for Petite Women		3: Fitting and Sizing Issues
	2: Flattering and Comfortable Dresses	Tops	1: Fitting and Sizing Issues for Tops
	3: Overall Look and Back Design		2: Versatile and Stylish Tops for Different Occasions
Intimate	1: Fit, wearability, and love for tops and dresses		3: Quality and Design Feedback
	2: Fitting and Sizing Issues for Suits and Leggings		4: Soft and Comfortable Sweaters in Various Colors
		Trend	1: fit and fabric
			2: color and overall appearance

(For topic selection steps, see Appendix Table 2,3)
The insights we can gather from the topics generated are the most probable topics that customers are discussing or interested in regarding each department's products. From these topics, we can see the most common topic is "fit", indicating that customers prioritize the comfort of the clothes they wear.
Also, in different departments, customers have different focuses. For example, for the bottoms department, customers are most likely concerned about the fit and size of pants and jeans, while for dresses, customers are interested in the dress size and fit, as well as the style, color, and fabric that flatters the wearer.

2 Sentiment Analysis

2.1 VADER Lexicon Model

We identified the sentiment of each review in different departments. For a sample of our result, see Appendix Table 4. The optimal threshold value is -0.009, based on which, the model accuracy rate is 0.77. The precision for this VADER lexicon-based sentiment analysis tool is 0.8. The recall(sensitivity) for this VADER lexicon-based sentiment analysis tool is 0.98.

For confusion matrix, VADER scores for actual positive reviews and actual negative reviews, see Appendix Figure 5-7. Results show that VADER model is better at identifying the positive reviews than negative ones.

2.2 SVM Model

For a sample of review text and their predicted sentiment score(column 'SentimentScore') and actual sentiment('sentiment'), see Appendix Table 5.

The Accuracy rate is 0.79, which is improved compared to the VADER model. For more detailed accuracy rates, precision, and recall rate, see Appendix Table 6.

3 Synthesis of Topic Modeling and Sentiment Analysis

Table 2 Results of combining topic modeling and sentiment analysis

Department Name	Topic	Positive Score	Negative Score	Neutral Score
Bottoms	Fit and Comfortable Jeans and Pants	0.663930	0.053914	0.282157
	Skirts in Various Colors and Fabrics	0.724138	0.041379	0.234483
Dresses	Flattering and Comfortable Dresses	0.728230	0.042105	0.229665
	Overall Look and Back Design	0.458425	0.110503	0.431072
	Sizing and Fit for Petite Women	0.596101	0.069954	0.333945
Intimate	Fit, wearability, and love for tops and dresses	0.676811	0.056604	0.266586
	Fitting and Sizing Issues for Suits and Leggings	0.619565	0.043478	0.336957
Jackets	Coats and Vests for Different Occasions	0.764706	0.058824	0.176471
	Comfortable and Fitting Jackets	0.677387	0.073367	0.249246
	Fitting and Sizing Issues	0.550000	0.050000	0.400000
Tops	Fitting and Sizing Issues for Tops	0.600812	0.077454	0.321733
	Quality and Design Feedback	0.467758	0.115660	0.416581
	Soft and Comfortable Sweaters in Various Colors	0.750522	0.026618	0.222860
	Versatile and Stylish Tops for Different Occasions	0.852632	0.005263	0.142105
Trend	Color and Overall Appearance	0.600000	0.000000	0.400000
	Color and Overall Appearance	0.500000	0.105263	0.394737

As we can see from the table, the possibilities of a topic under one department being "positive", "negative", and "neutral" sum to one. Most of the topics obtained a "positive" score higher than 0.5, but "Overall Look and Back Design" from "Dress" department and "Quality and Design Feedback" under "Tops" department. We assume this is a valuable finding, and can guide the cloth producer to increase customers' satisfaction in response to customers' reviews.

Conclusion

Using topic modeling and sentiment analysis, we have examined customer review data and combined these approaches to derive significant recommendations for clothing sellers. Our analysis has led to the following key insights:

- Fit is crucial: Customers prioritize well-fitting clothes that provide freedom of movement without being overly tight or restrictive. Accurate sizing information is vital for customers to make informed choices without physically trying on the garments.
- Tailored approach: Different clothing departments elicit varying customer preferences. Sellers should refer to our topic table and take targeted actions based on these findings.
- Addressing concerns: Customers express negative opinions regarding the "Overall Look and Back Design" of dresses and the "Quality and Design Feedback" of tops. Suggestions include collaborating with experienced designers and focusing on fabric quality and overall aesthetic appeal.

Our project has two main limitations. Firstly, self-selection bias may skew the overall satisfaction level as users who are extremely positive or negative are more likely to comment. Secondly, the dataset collected six years ago may not accurately reflect current customer preferences. Nonetheless, we consider our methods and findings to be instructive and valuable.

Overall, our project offers an efficient means for sellers to gain insights and improve their understanding of customer needs.

Appendix

Table 1 Introduction of variables

• **Clothing ID:** Integer Categorical variable that refers to the specific piece being reviewed.
• **Age:** Positive Integer variable of the reviewers age.
• **Title:** String variable for the title of the review.
• **Review Text:** String variable for the review body.
• **Rating:** Positive Ordinal Integer variable for the product score granted by the customer from 1 Worst, to 5 Best.
• **Recommended IND:** Binary variable stating where the customer recommends the product where 1 is recommended, 0 is not recommended.
• **Positive Feedback Count:** Positive Integer documenting the number of other customers who found this review positive.
• **Division Name:** Categorical name of the product high level division.
• **Department Name:** Categorical name of the product department name.
• **Class Name:** Categorical name of the product class name.

Figure 1 Descriptive analysis: distribution of customer ratings by age

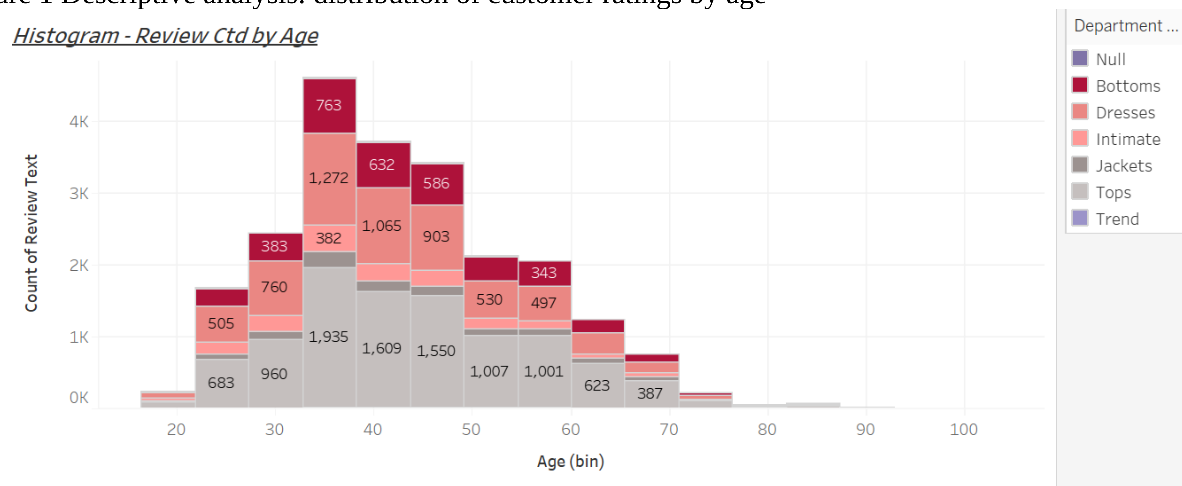

Figure 2 Descriptive analysis: line chart of customer ratings by age

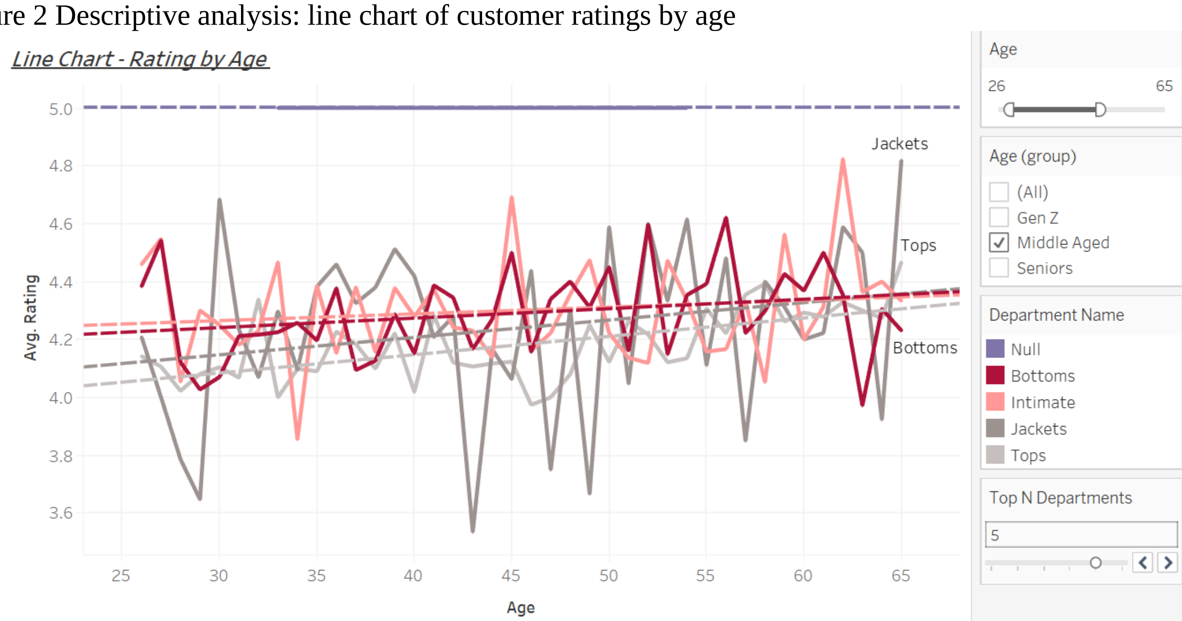

Figure 3 Descriptive analysis: Number of Ratings and average rating by department

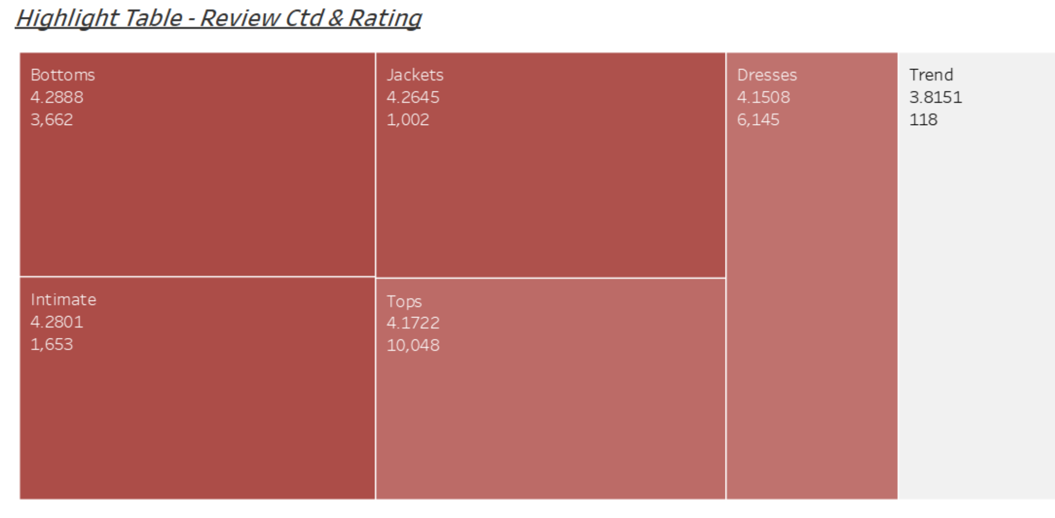

Figure 4 Descriptive analysis: summary of numeric variables

	Unnamed: 0	Clothing ID	Age	Rating
count	23486.000000	23486.000000	23486.000000	23486.000000
mean	11742.500000	918.118709	43.198544	4.196032
std	6779.968547	203.298980	12.279544	1.110031
min	0.000000	0.000000	18.000000	1.000000
25%	5871.250000	861.000000	34.000000	4.000000
50%	11742.500000	936.000000	41.000000	5.000000
75%	17613.750000	1078.000000	52.000000	5.000000
max	23485.000000	1205.000000	99.000000	5.000000

	Recommended IND	Positive Feedback Count
count	23486.000000	23486.000000
mean	0.822362	2.535936
std	0.382216	5.702202
min	0.000000	0.000000
25%	1.000000	0.000000
50%	1.000000	1.000000
75%	1.000000	3.000000
max	1.000000	122.000000

Table 2 The best number of topics for each clothing department in topic modeling

```python
#print best number of topics for each group
for name, num_topics in best_num_topics.items():
    print(f"Department Name: {name}, Best Number of Topics: {num_topics}")
```

```
Department Name: Bottoms, Best Number of Topics: 4
Department Name: Dresses, Best Number of Topics: 5
Department Name: Intimate, Best Number of Topics: 4
Department Name: Jackets, Best Number of Topics: 5
Department Name: Tops, Best Number of Topics: 5
Department Name: Trend, Best Number of Topics: 4
```

Table 3 Sample of raw results for topics identification

```
# Print LDA topics for each group
for name, lda_model in lda_models.items():
    print(f"Department Name: {name}")
    for topic_id, topic in lda_model.print_topics(num_topics=best_num_topics[name]):
        print(f"Topic {topic_id+1}: {topic}")
```

```
Department Name: Bottoms
Topic 1: 0.027*"fit" + 0.022*"size" + 0.020*"pant" + 0.018*"love" + 0.017*"jean" + 0.013*"wear" + 0.013*"look" + 0.013*"lik
Topic 2: 0.028*"skirt" + 0.016*"love" + 0.015*"size" + 0.012*"color" + 0.011*"look" + 0.011*"fit" + 0.010*"waist" + 0.010*"
Department Name: Dresses
Topic 1: 0.045*"dress" + 0.034*"size" + 0.026*"fit" + 0.025*"5" + 0.016*"small" + 0.015*"order" + 0.015*"petit" + 0.013*"lo
Topic 2: 0.063*"dress" + 0.020*"love" + 0.020*"look" + 0.014*"fit" + 0.013*"wear" + 0.012*"great" + 0.012*"color" + 0.012*"
Topic 3: 0.040*"dress" + 0.018*"like" + 0.014*"look" + 0.011*"back" + 0.010*"review" + 0.009*"one" + 0.009*"fit" + 0.008*"f
Department Name: Intimate
Topic 1: 0.019*"fit" + 0.017*"love" + 0.014*"wear" + 0.014*"size" + 0.013*"top" + 0.012*"like" + 0.011*"great" + 0.011*"loo
Topic 2: 0.018*"size" + 0.012*"love" + 0.012*"wear" + 0.010*"color" + 0.010*"look" + 0.010*"like" + 0.010*"fit" + 0.010*"su
Department Name: Jackets
Topic 1: 0.025*"jacket" + 0.019*"size" + 0.018*"fit" + 0.018*"look" + 0.017*"love" + 0.014*"like" + 0.012*"wear" + 0.011*"c
Topic 2: 0.020*"love" + 0.017*"coat" + 0.014*"look" + 0.013*"wear" + 0.009*"vest" + 0.008*"jacket" + 0.008*"great" + 0.008*
Topic 3: 0.024*"jacket" + 0.016*"fit" + 0.013*"look" + 0.012*"small" + 0.012*"like" + 0.011*"size" + 0.011*"order" + 0.011*
Department Name: Tops
Topic 1: 0.033*"size" + 0.029*"top" + 0.026*"fit" + 0.023*"small" + 0.020*"5" + 0.014*"wear" + 0.014*"order" + 0.013*"xs" +
Topic 2: 0.028*"top" + 0.026*"love" + 0.024*"jean" + 0.020*"great" + 0.020*"look" + 0.019*"wear" + 0.017*"color" + 0.010*"b
Topic 3: 0.032*"look" + 0.029*"top" + 0.028*"like" + 0.014*"back" + 0.013*"would" + 0.012*"love" + 0.012*"fabric" + 0.011*"
```

Figure 5 Confusion matrix

Predicted:	negative	positive	All
True:			
negative	123	347	470
netural	86	534	620
positive	67	3541	3608
All	276	4422	4698

Figure 6 VADER scores for actual positive reviews

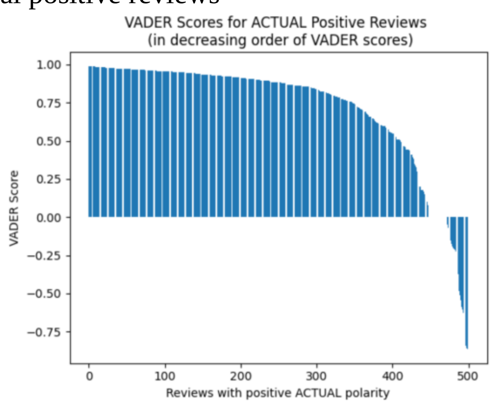

Figure 7 VADER scores for actual negative reviews

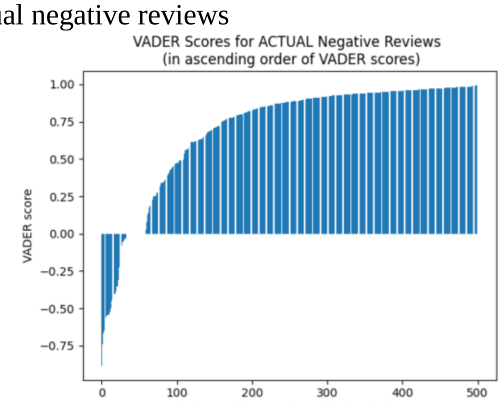

Table 4 Sample of VADER sentiment analysis result

	Department Name	Sentiment
0	Bottoms	[positive, positive, positive, positive, posit...
1	Dresses	[positive, positive, positive, positive, posit...
2	Intimate	[positive, positive, positive, positive, posit...
3	Jackets	[positive, positive, positive, positive, posit...
4	Tops	[positive, positive, positive, positive, posit...
5	Trend	[positive, positive, positive, positive, posit...

Table 5 Sample of SVM sentiment analysis

Clothing ID	Age	Title	Review Text	Rating	Recommended IND	Positive Feedback Count	Division Name	Department Name	Class Name	sentiment	SentimentScore
767	33	NaN	Absolutely wonderful - silky and sexy and comf...	4	1	0	Initmates	Intimate	Intimates	neutral	positive
1080	34	NaN	Love this dress! it's sooo pretty. i happene...	5	1	4	General	Dresses	Dresses	positive	positive
1077	60	Some major design flaws	I had such high hopes for this dress and reall...	3	0	0	General	Dresses	Dresses	neutral	neutral
1049	50	My favorite buy!	I love, love, love this jumpsuit. it's fun, fl...	5	1	0	General Petite	Bottoms	Pants	positive	positive
847	47	Flattering shirt	This shirt is very flattering to all due to th...	5	1	6	General	Tops	Blouses	positive	positive

Table 6 Accuracy rates

	precision	recall	f1-score	support
negative	0.81	0.53	0.64	2407
neutral	0.73	0.66	0.69	7948
positive	0.82	0.92	0.87	13131
accuracy			0.79	23486
macro avg	0.79	0.71	0.74	23486
weighted avg	0.79	0.79	0.79	23486

Reference

Nicapotato (2018) *Women's e-commerce clothing reviews*, *Kaggle*. Available at: https://www.kaggle.com/datasets/nicapotato/womens-ecommerce-clothing-reviews (Accessed: May 1, 2023).