Women’s E-commerce Clothing Reviews

Group Members: Yijia Chen, Yanbei Cheng, Charlie Mei, Changfei Guan, Chloe Du, Yvonne Yi-Lun Lan

Dataset: https://www.kaggle.com/nicapotato/womens-ecommerce-clothing-reviews

This is a Women’s Clothing E-Commerce dataset that includes multiple dimensions of commercial data. The data includes both customer information, product information, and product purchasing information. Each row also includes a review title and a detailed review text from a customer corresponding to a particular clothing product. The dataset has 10 feature variables and 23,486 rows. The variables include: Clothing ID, Age, Title, Review Text, Rating, Recommended IND, Positive Feedback Count, Division Name, Department Name, and Class name.

There is no overlap or conflicting coding schemes in the dataset. Overall, the dataset provides good information regarding customer reviews, ratings, and customer age for us to explore different business questions. We plan to conduct our analysis through text mining and clustering analysis.

Problem Statement

After carefully reviewing the data, we seek to explore the underlying sentiments within each customer review. We would also like to know whether the sentiments and other variables such as age, product department, and product class can be used to build predictive models to predict ratings and provide recommendations. Specifically, we will seek to answer the following questions from the dataset:

1) What is the overall sentiment for recommended clothing items?
2) What is the overall sentiment for clothing items that are not recommended?
3) Are there differences in sentiment for different departments, product divisions and/or product classes?
4) What is the relationship between specific words and sentiment on the rating for clothing items?
5) What is the relationship between specific words and sentiment on the recommendation for clothing items?

Analytical Techniques

We will be using **Text Mining** to conduct our analysis and build predictive models for the dataset.

Text Mining Introduction

Text mining is the process of obtaining and extracting high-quality information from texts to conduct analysis for a wide variety of scientific research, government, and business purposes.
The process typically includes dimensionality reduction, information retrieval, natural language processing, named entity recognition, disambiguation, recognition of pattern identified entities, document clustering, and sentiment analysis. The primary goal is to convert text into useful data for analysis. This goal can be achieved through applications such as natural language processing to effectively interpret data and extract information tailored to different needs.

Text can be used to identify keywords or groups of words and then be analyzed further to identify underlying themes and patterns. By using text as an input we can effectively enhance the accuracy and application of predictive models. The analytical process starts with information retrieval, which includes extracting useful information from text search engines, opinion mining, or summarizing documents. Lexical analysis is the process of converting a sequence of characters or words into a sequence of tokens. One approach for text preparation is the bag of words approach, which includes creating corpurs, converting characters to lower cases, removing punctuations, removing brackets, removing stopwords, removing web URLs, stripping white spaces, and replacing abbreviations. After the preparation stage, more techniques such as pattern recognition, tagging/annotation information extraction, and data mining techniques can be applied to analyze the data.

The application of text analysis is broad. The analysis can be applied to feature extraction, entity extraction, theme extraction, sentiment analysis, translation, and document search etc. Feature extraction can be achieved through identifying specific words or keywords in text. Entity extraction can be conducted through identifying entities from texts from places such as companies, countries, or organizations. Theme extraction can be achieved through clustering and topic models. As a result, text mining is a suitable technique for conducting analysis for this dataset.

Text Mining on Women’s E-commerce Dataset

The Women’s E-commerce dataset includes information on review points, review texts, and recommender ID, which are ideal for conducting sentiment analysis, building predictive models, and providing recommendations. The first step is to prepare the dataset for text mining. There are no missing values in the dataset, and we combined “title” and “review” columns into one column. The next step is to conduct basic text cleaning: converting all texts to lowercase, removing punctuations, removing stopwords, and stripping white spaces. Then, we created a document term matrix to tokenize the texts, remove sparse terms, and complete stems. The most frequently appearing words in the comments of our dataset are visualized in Figure 0. The top five words are: Love (13,738), Dress (13,726), Fit (12,028), Size (10,908), Look (9,540).

Figure 0: Wordcloud of top 100 most commonly used words

1. Sentiment Analysis on Customer and Clothing Reviews

We conducted sentiment analysis for the clothing reviews in the dataset using “Bing” Lexicon and “Afinn” Lexicon from the tidytext package. The “Bing” lexicon categorizes words into positive and negative groups. The “Afinn” lexicon assigns words with a sentiment score ranging from -5 to 5. Negative scores indicate negative sentiment and positive scores indicate positive sentiments.

We have computed two sentiment scores - one for consumer sentiment and one for average clothing sentiment. Consumer sentiment is the average score based on all words they have used in their comments. Clothing sentiment is the average consumer sentiment observed for each clothing item in the dataset. There are over 1,000 clothing items in the data, so multiple consumers have rated particular clothing items.

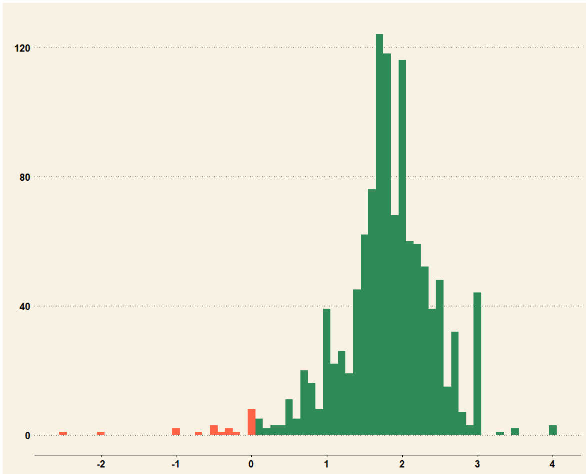


Figure 1: Distribution of average consumer sentiment
Largely positive reviews by consumers, average score of 1.8, median 2.0.

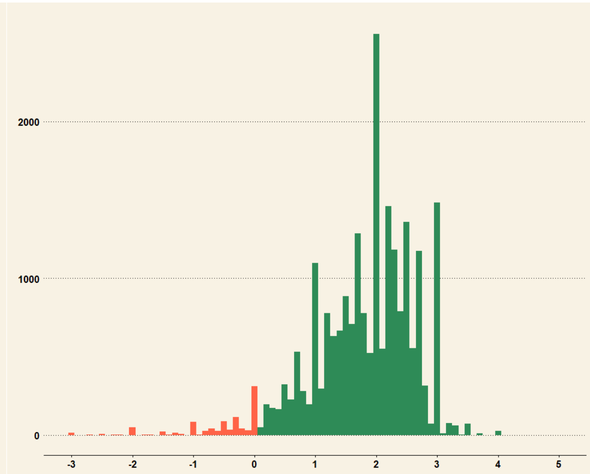


Figure 2: Distribution of average sentiment towards clothing products
Largely positive reviews for clothing items, average score of 1.75, median 1.76.

Figure 1 displays the distribution of average consumer sentiment. Customer sentiments are largely positive, with the average score being 1.8 and median score being 2.0. Figure 2 displays the distribution of the average sentiment towards the clothing product purchased. The sentiments are also largely positive, with the average score being 1.75 and median score being 1.76.

2. Sentiment Analysis on Clothing Division, Product Department, Product Class, and Age

Using these sentiments, we then compare the differences in sentiment towards different clothing items. We want to know whether customers have different sentiments towards clothing from different divisions, departments, classes, and ages. We utilized the "Bing" and "Afinn" lexicons to evaluate the differences in sentiments across these different groups.

3. Predictive Models: Random Forest

The final step of our analysis is to build predictive models for ratings and whether a product is recommended or not. We build two random forest models - one for ratings and one for recommendations. Predictor variables include: word frequency, sentiment scores, age, product division, product department and product class. The dataset was divided into train and test groups, using 80% of the dataset as the train set and 20% of the dataset as the test set. For both of the random forest models, we ran 1,000 bootstrapped trees ("num.trees = 1000"), each of which considers ten of the predictor variables ("mtry = 10").

To evaluate the models, we examined the model accuracy on the test sample. We also considered the relative importance of the predictor variables for each model to see which predictors matter most for ratings and recommendations.

Results Discussion

PART 1: Sentiment for clothing items

1）What is the overall sentiment for recommended clothing items?
2）What is the overall sentiment for clothing items that are not recommended?

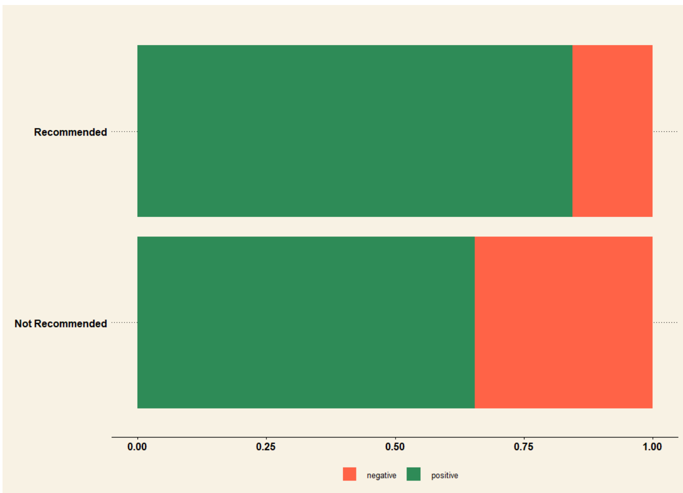


Figure 3: Overall sentiment comparison of recommended and not recommended clothing items

As shown in Figure 3, 84% of recommended items gained positive sentiment from customers, while of those which are not recommended, positive sentiment decreased to 66% and negative sentiment more than doubled, about 34% . However, overall, most people have positive feelings about products, whether they recommended them or not.

3）Are there differences in sentiment for different departments, product divisions and/or product classes?

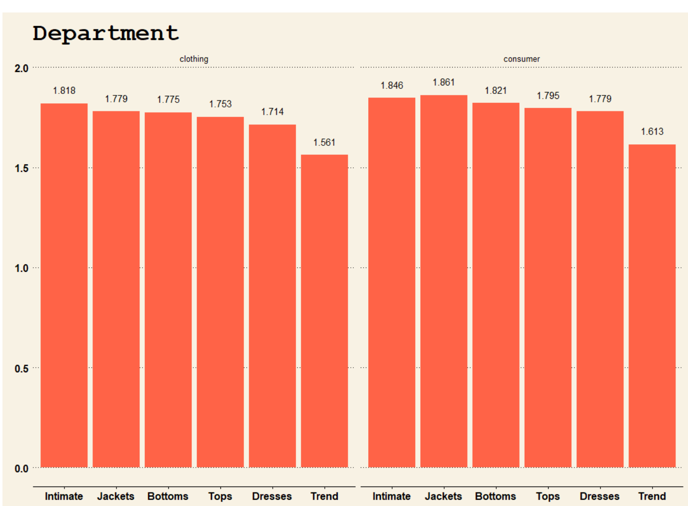


Figure 4: Average department sentiment by item and consumer

Figure 4 shows the average sentiments for clothing items in different departments and consumers' general sentiments toward clothing items in different departments. First, clothing products from 4 departments receive an above-average sentiment score (1.75) and there are 3 departments towards which the consumer sentiments are above average consumer sentiments (1.80). Thus, overall, the sentiments toward each department are largely positive. Second, the average sentiment of departments by item and consumer is consistent, which indicates that intimate, jacket, bottom and top departments are relatively popular and well received but the performance of the trend division is relatively flat. Lastly, it is worth noting that the slight difference related to score between clothing class and consumers result from the fact that the same clothing item can be reviewed by more than one customer.

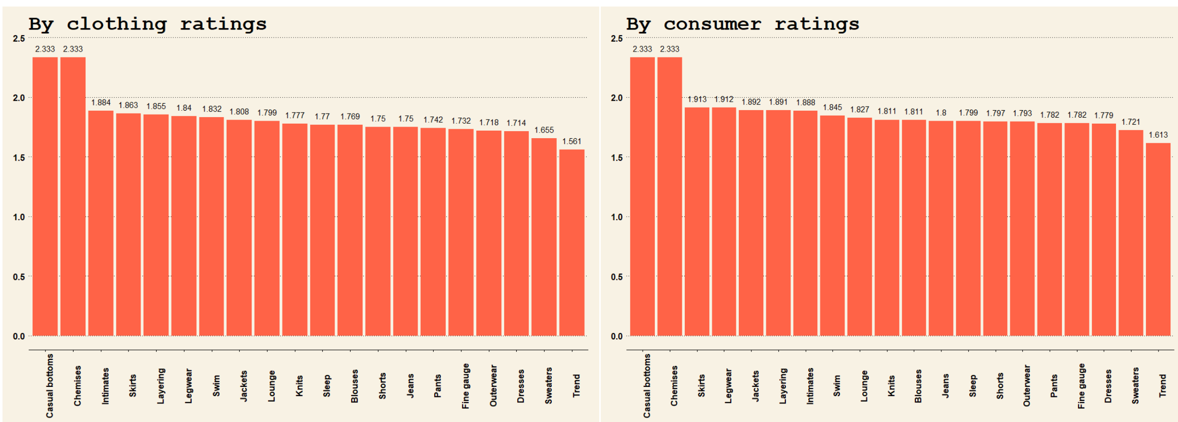


FIgure 5: Average product sentiment by item and consumer

Figure 5 displays average sentiments different product classes gained and customers held toward these product lines respectively. It can be clearly learned that casual bottoms and chemises got a highly positive sentiment, which indicates these two product classes are wildly favored by women shopping online. Moreover, trend products may be relatively personalized, so the average sentiment toward this class is relatively less positive.

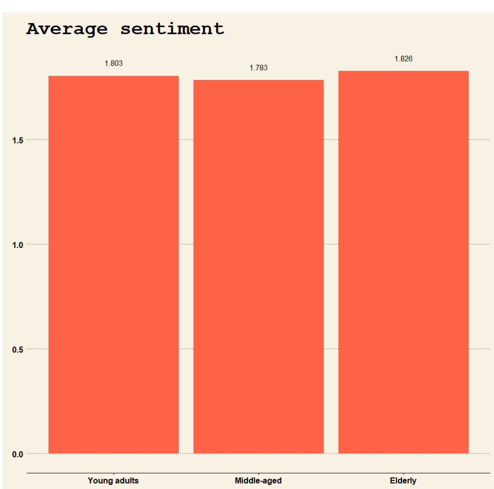


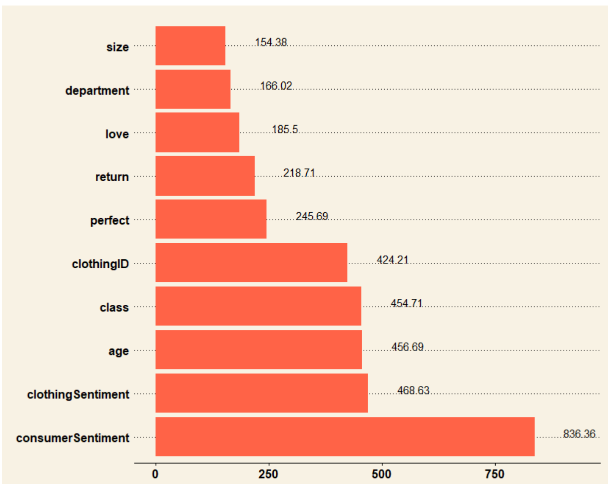
Figure 6 indicates the average sentiment of different age groups. Customers of all ages were generally in a positive mood toward E-commerce clothing and there is a slight difference among them . When looking closely, the elderly seem to be more likely to like products while the middle-aged hold a sentiment below average sentiment score.

Figure 6: Average sentiment by age group

PART 2: Predicting product ratings and recommendations with random forest

Through random forest models, we predict whether the words used in customer reviews and their overall sentiment impact product ratings and whether a clothing item is recommended.

4）What is the relationship between specific words and sentiment on the rating for clothing items?



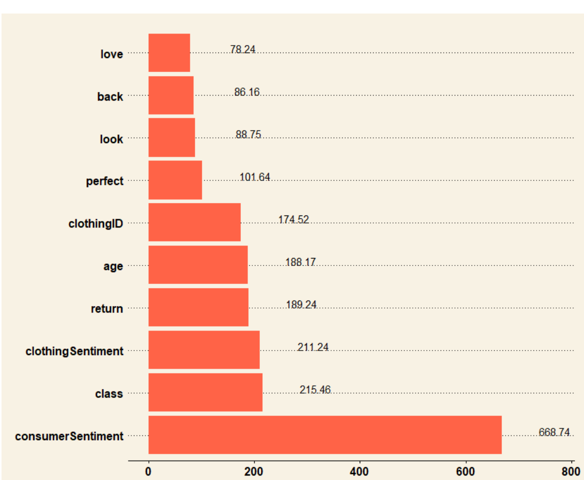

```
prop_matrix <- prop.table(conf_matrix, 1); prop_matrix
 pred_test
            1           2           3           4           5
1 0.072289157 0.102409639 0.307228916 0.048192771 0.469879518
2 0.018867925 0.066037736 0.311320755 0.106918239 0.496855346
3 0.015845070 0.019366197 0.244718310 0.151408451 0.568661972
4 0.001008065 0.006048387 0.067540323 0.102822581 0.822580645
5 0.000407498 0.002444988 0.011002445 0.018337408 0.967807661
```

Figure 7: Variable importance from random forest model for product ratings (N = 17,988) *Figure 8: Confusion matrix for product ratings*

For customer ratings, we find that consumer and clothing sentiment both largely influence their ratings. Other important variables include (in decreasing order of importance): age, class, clothingID, and then words: perfect, return, love.

This model has an overall accuracy of 58.88%. From the confusion matrix in Figure 8, we can see that our model was particularly good at predicting high ratings but has difficulty in making correct predictions on low product ratings.

5）What is the relationship between specific words and sentiment on the recommendation for clothing items?




```
> prop_matrix2 <- prop.table(conf_matrix2, 1); prop_matrix2
  pred_test2
            0         1
  0 0.3193588 0.6806412
  1 0.0233252 0.9766748
```

Figure 10: Confusion matrix for recommendations

Figure 9: Variable importance from random forest model for whether products are recommended (N = 17,988).

For customer recommendations, we find that consumer sentiment has the highest predictive power. Other important variables include (in decreasing order of importance): class, clothing sentiment, and then words: return, perfect, and look.

This model has an overall accuracy of 85.82%. From the confusion matrix in Figure 10, we can see that our model was again better at predicting clothing items that are recommended, resulting in a high model sensitivity of 97.67%. However, prediction of non-recommended products is poor again, as evidenced in the low model specificity of 31.93%.

Conclusion & Recommendations

Findings and Recommendations

Sentiment analysis is the computational treatment of attitudes, sentiments and subjectivity of text. The capacity to leverage sentimental text to gain consumer and market insights as well as improve brand reputation can make a significant impact in today's organizations.

This research project identified and predicted how likely consumers will react to clothing products. People tend to express their opinions on the Internet whether they approve or disapprove of the products. Hence, analyzing consumer and product sentiments could help provide insights on customers' ratings and their relative attitude towards E-commerce clothing.

Text from consumer reviews were extracted, tokenized and awarded sentiment values. Word frequency counts were also constructed to identify most commonly used words in consumer review. We find that sentiments from reviews were mainly positive. More than 80% of recommended items received positive reviews whereas only 34% of items not recommended showed negative reviews.

Using this text analysis, we also implemented random forest models to examine the relationship between specific words and sentiment on clothing ratings and recommendations. Sentiment values appear to be particularly good predictive variables, accounting for the largest portions of predictive power in our models. Secondary to that are word frequencies counts. Ultimately, it is evident that consumer and product sentiment is an important first step to understand consumer ratings. Thus, we would highly recommend completing this intermediary sentiment analysis stage when attempting to predict consumer ratings and recommendations.

Directions for future work

This research project serves as an effective approach to understanding how consumers rate clothing products and make their recommendations. Existing sentiment analysis models can be improved further with more feature selections with semantic and advanced text mining techniques. Possible directions for future work include exploring other contexts outside of clothing to understand whether natural language processing is effective in interpreting product ratings and recommendations.