Factors Impacts on Online Shopping Consumers' Purchasing Decisions: Evidence of Women dresses from Amazon

Diyang Lu[1, *]

[1]Department of Economics and Management, Beijing Jiaotong University, Beijing, China

*Corresponding author: 18241199@bjtu.edu.cn

Abstract. Online reviews have become an important way for consumers to understand product quality and merchant services with the prosperity of online shopping. This paper adopts the method of natural language learning to conduct sentiment analysis on Amazon women's online shopping reviews, then establishes a bag of words model to vectorizes the text. Finally, a prediction model is established based on the Naive Bayes classifier, which predicts consumers' recommendation willingness by analyzing the review text. Through training and testing, the accuracy of the prediction model reaches 0.87, and improved to 0.92 by TF-IDF algorithm. This study analyzes the key influencing factors of women's online shopping which has a high reference value for merchants to improve services in a targeted manner to obtain more consumer recommendations.

Keywords: NLP; Sentiment Analysis; Online Reviews; TF-IDF; Bag-of-words model; Naive Bayesian Classifiers

1. Introduction

Thousands of online clothing shopping comments have been generated on major e-commerce platforms as clothing online retail has become an indispensable part of people's daily life. These reviews contain the real feedback of consumers on clothing products and a large amount of commercial competitive intelligence information, such as consumers' attention to different brands and different categories of clothing, consumers' purchasing intentions, and the relationship between sales and evaluations. It has become an urgent problem for e-commerce to extract intelligence information from clothing review texts so that business operators can respond to changes in the consumer market in a timely and effective manner[1-3].

Through the literature review, it is found that the important effect of online shopping reviews has already attracted the attention of many scholars in the field of e-commerce and related research at home and abroad. In the early days, scholars focused on the external features of the commentary rather than the text itself. To be specific, Liu et al. studied the relationship between consumer trust tendency, reviewer credibility, reviewer's emotional tendency, online review quality, online review quantity, online review timeliness and consumers' purchase intention [4]. Liu and Li roughly divided the reviews into positive and negative reviews [5]. Besides, some scholars have realized the value of the review content itself. Chen and Li proposed that the quality of online reviews has a significant positive predictive effect on purchase intention, and verified it with a questionnaire survey [6]. As text data mining technology becomes more and more mature, more and more scholars use machine learning algorithms to build models to study online shopping reviews from the perspective of natural language learning. Zhang applied the Bi-LSTM+CRF model to achieve automatic extraction of product attributes and attribute evaluations in reviews [7]; Li et al. constructed a high-frequency word co-occurrence network to analyze the structural characteristics of online shopping reviews [8]. The research on online shopping reviews in recent years undoubtedly points to a general direction—big data text analysis, which is to use massive data to clarify the consumption intentions pointed to by various words in the reviews, so as to facilitate targeted optimization by merchants. The purpose of this study is to use online shopping reviews to predict consumers' recommendation intentions. The technical route of the full text is shown in Fig 1.

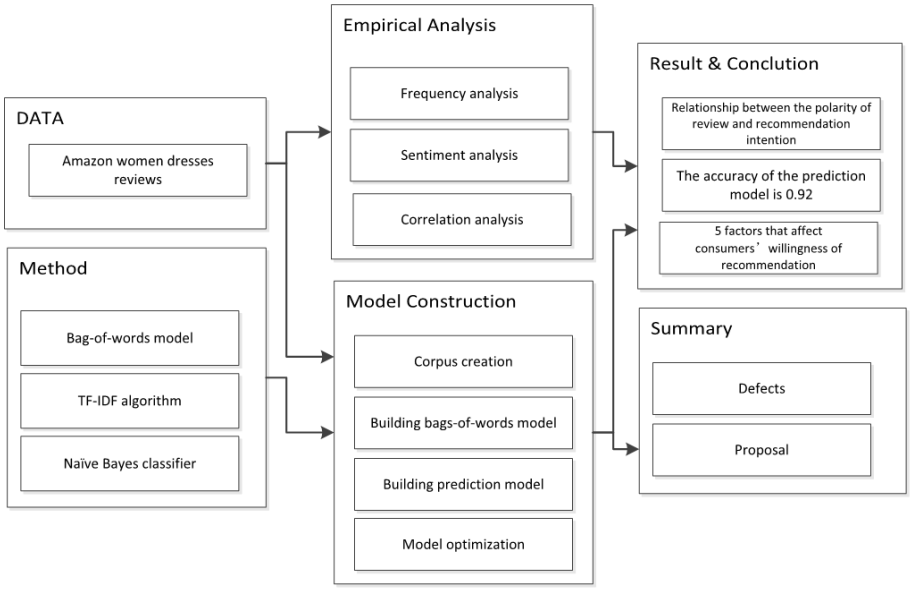


Fig 1. Technical route

2. Data & Method

2.1 Data descriptions

The dataset Women's E-Commerce Clothing Reviews used in this article comes from Kaggle. Because this is real commercial data, it has been anonymized, and references to the company in the review text and body have been replaced with "retailer". This dataset includes 23486 rows and 10 feature variables. The variables and their meanings are shown in Table 1.

Table 1. Variable table

Varaibles	Notes
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.

The research subject of this paper is consumer reviews, so records with empty reviews are deleted in advance. In addition, the irrelevant columns Clothing ID and Title in the dataset were removed. Finally, the dataset used for analysis contains 8 variables and 845 records, i.e., the dataset size is 845 rows × 8 columns.

Before conducting the correlation analysis between Review Text and Positive Feedback Count, this study first classifies and counts various types of data contained in the dataset according to fields and visualize them. The histograms of each age, product score, category of clothes are shown in Fig. 2-4. Fig. 2 shows the age distribution of the dataset. The user data collected in this dataset are mainly young and middle-aged. 80% of the users are between the ages of 28 and 61. Users of this age group are often financially independent, have strong purchasing power. They are proficient in using the Internet to consume, and have a high frequency of online shopping. Therefore, their online shopping reviews and recommendation willingness have high data value.

Seen from Fig. 3 that most records in this dataset have high ratings. The amount of data increases in steps from 1 to 5 points. Fig.4 indicates that when people give low scores, they tend to be reluctant to recommend the product to others. Conversely, when people give high scores, they have a strong

willingness to recommend it. As illustrated in Fig. 5, Dresses, suits and blouses occupied the first, second and third places respectively. Figures 6 and 7 depict the age data distribution of each department and each division when the recommended integers are 1 and 0 respectively. It can be seen from Fig 6 and 7 that whether the recommended int is 1 or 0, the data-count distribution of different ages in the data set is similar. Besides, it can be considered that age is not related to customer recommendation willingness.

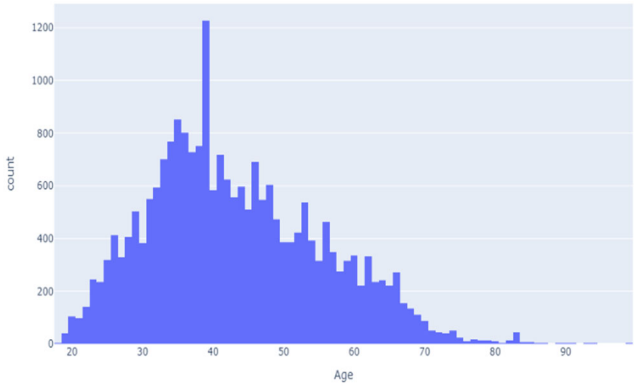


Fig. 2 The histograms of each age

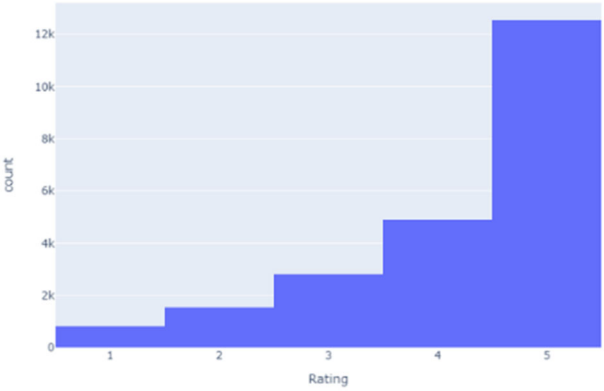


Fig. 3 The histograms of each rating

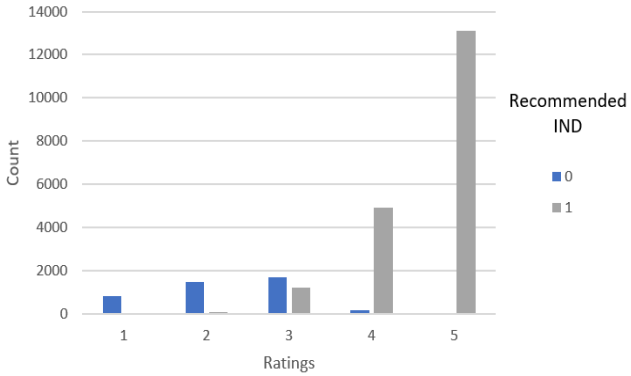


Fig. 4 Ratings under different recommendation intentions

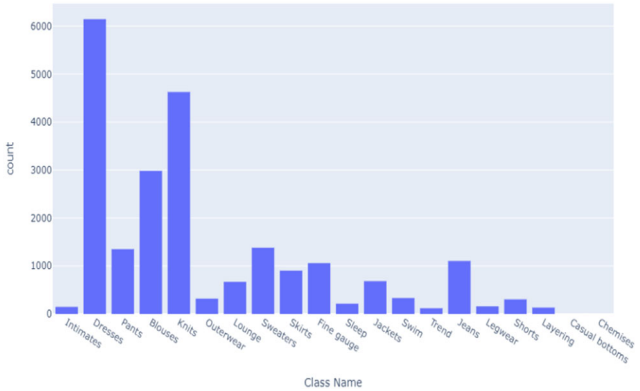


Fig. 5 The histograms of each category

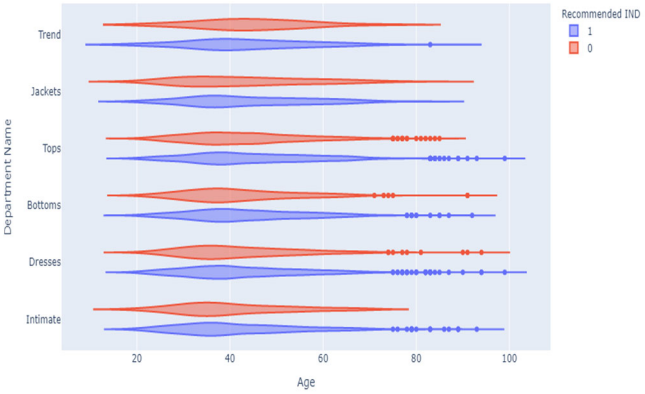


Fig. 6 The relationship between consumers' age and Recommended IND in different department

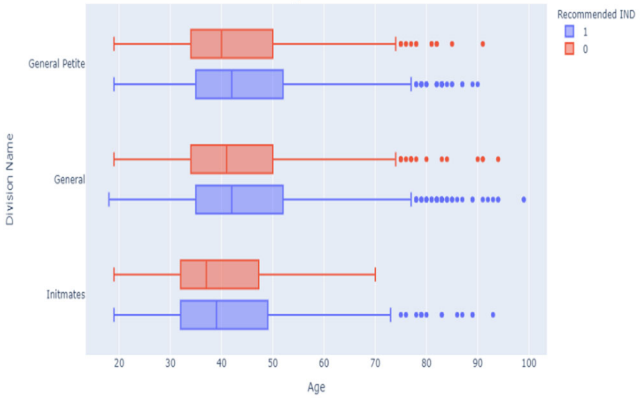


Fig. 7 The relationship between consumers' age and Recommended IND in different divisions

2.2 Description of the models

2.2.1 Bag-of-words model

The bag-of-words model transformed a document into a vector ignoring its word order, grammar, syntax and other elements. Each dimension of the vector represents a keyword, and the value of the dimension is the frequency of a certain word.

2.2.2 TF-IDF algorithm

TF-IDF is a statistical method to assess the importance of a word to a document set or one of the documents in a corpus. TF-IDF is actually TF * IDF. TF refers to Term Frequency, IDF refers to Inverse Document Frequency. The main idea of IDF is that if there are fewer documents containing a certain term, that is, the smaller the number of documents containing this term, the larger the IDF, it means that the term has a good ability to distinguish between categories.

2.2.3 Naive Bayes classifier

Naive Bayes classifiers are a series of simple probabilistic classifiers based on the application of Bayes' theorem. It assumes that each feature of the sample is uncorrelated with other features. For a given item to be classified, the classifier calculates the probability that it belongs to each category, and selects the most likely category as its final category .

3. Empirical analysis

3.1 Frequency analysis

The following formally enters the analysis of consumer and review texts. Text length is one of the important characteristics of this data. After statistics, the length and number of words of consumer comments are given in Fig. 8. The number of characters and words of the comments are irregularly distributed, and the maximum values are around 500 characters and 100 words, respectively.

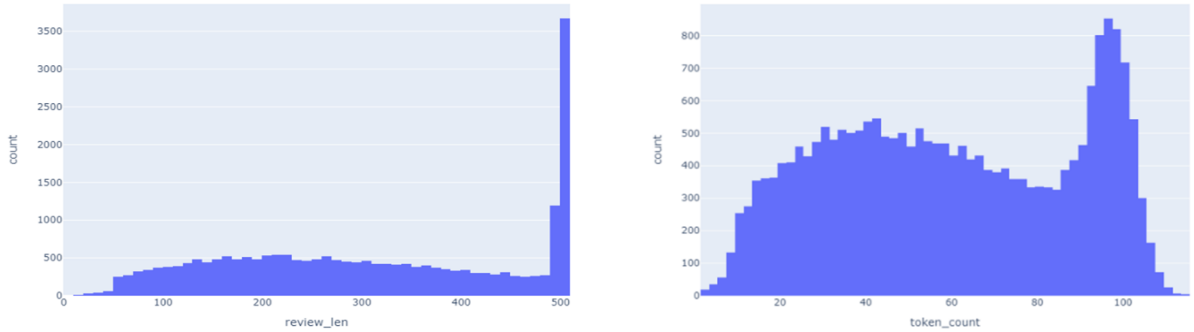


Fig. 8 The histograms of review length and reviews' words.

3.2 Sentiment analysis

Since users often do not follow strict writing standards when writing online shopping reviews, the texts are often mixed with irregular characters. Therefore, when using machine learning algorithms for sentiment analysis, text normalization should be performed first. Replaced spaces and '&' with ',' for a total of 86 replacements in the entire comment text.

TextBlob is used to perform sentiment analysis on user reviews[8], which can quantify the sentimental tendencies of review texts by polarity. The value of polarity is between [-1.0, 1.0], -1.0 is the most negative, 1.0 is the most positive. Table 2 and Fig 9 shows the polarity of a part of comments obtained.

Table 2. Review Text Polarity Table (part)

Index	Polarity
0	0.63
1	0.34
2	0.07
3	0.55
...	...
23481	0.55
23482	0.09
23483	0.41
23484	0.32

As shown in Fig. 9, the polarity of review texts is normally distributed with an average value of 0.215, and the polarity of most review texts is >0. If the comments are divided into three categories according to their polarity, namely positive[0.5,1], neutral[0,0.5], and negative [-1,0], then the various attitudes in the comment text The proportion is presented in Fig.10. Emotional tendencies in reviews

may vary when consumers are willing or unwilling to recommend a product to others. The polarity distribution of comments under different recommendation intentions is shown in the figure 5. As shown in Figure 11, under all departments, review polarity and recommendation willingness exhibit a fixed relationship—reviews willing to recommend a product to others tend to have higher polarity.

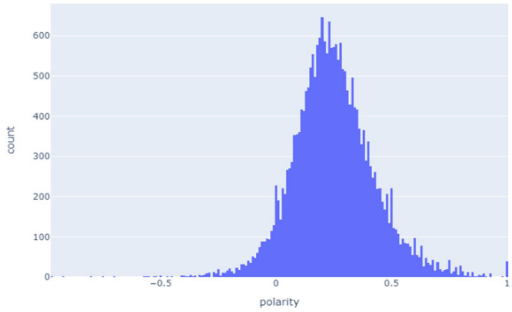


Fig. 9 Polarity of Review Text

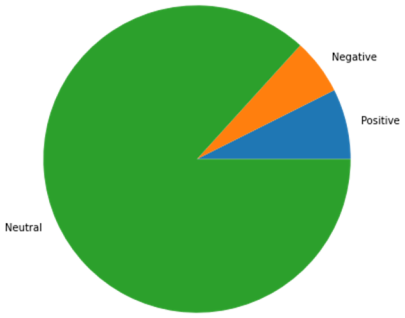


Fig. 10 Review text polarity distribution map

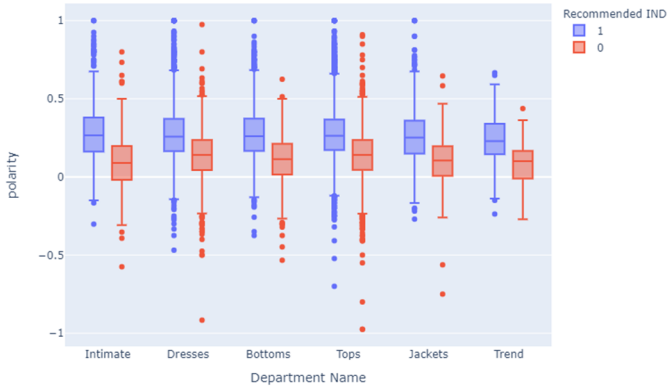


Fig. 11 The relationship between polarity and recommendation intention in different departments

3.3 Correlation analysis

Among the elements discussed above, there is a clear correlation between some of the data, as shown in Figure 12. In order to eliminate the influence of multicollinearity as much as possible, the correlation coefficient between variables should not exceed 0.8. It can be seen that the correlation coefficient between review_len and token_count is as high as 0.99, so only one of these two variables should be retained, and the text length is selected here. The final correlation heatmap is shown in Figure 13.

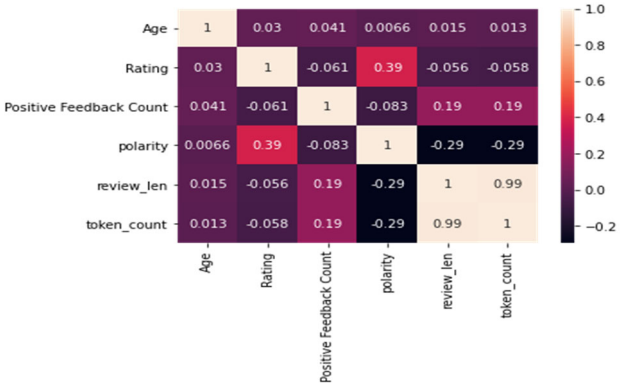


Fig. 12 Correlation HeatMap (6 viriables)

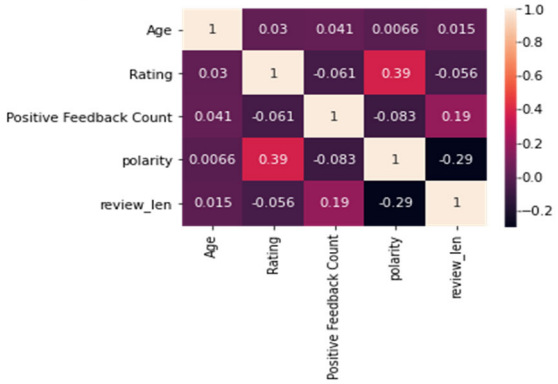


Fig. 13 Correlation HeatMap (5 variables)

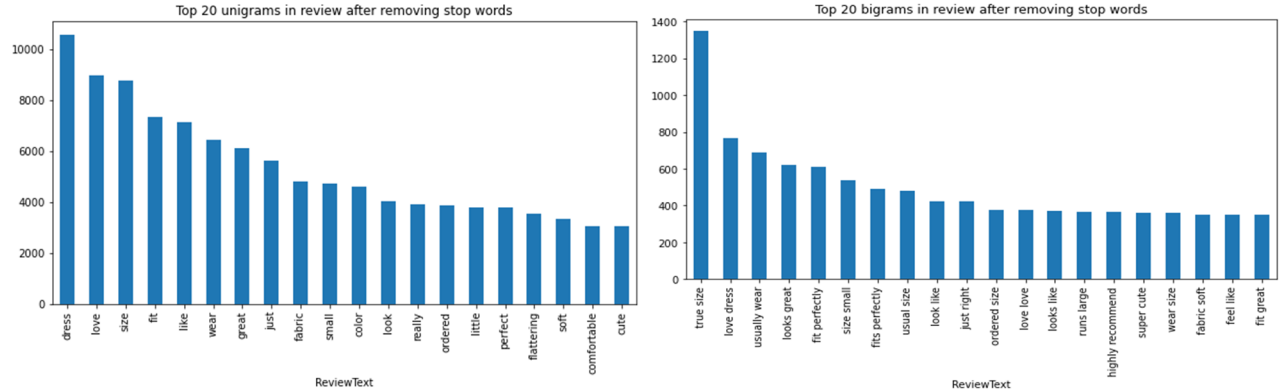
4. Model construction

4.1 Corpus Creation

Before modeling, the text data need to be convert into a corpus. To create a corpus of consumer reviews based on consumer reviews, it is necessary to perform three steps of word segmentation, stemming, and removal of stop words. The applied toolkit is NLTK (Natural Language Toolkit).[9]

- Step1 Use NLTK to segment the text, and the thesaurus is the default English thesaurus of NLTK;
- Step2 Use NLTK.stem to extract the sentence stem;
- Step3 Remove stop words, the stop word list is the default stop word list of NLTK.

Finally, a text set that can be used for vectorization is obtained. Fig 14 shows the 20 most frequently occurring words, bigrams, trigrams in the obtained text set.



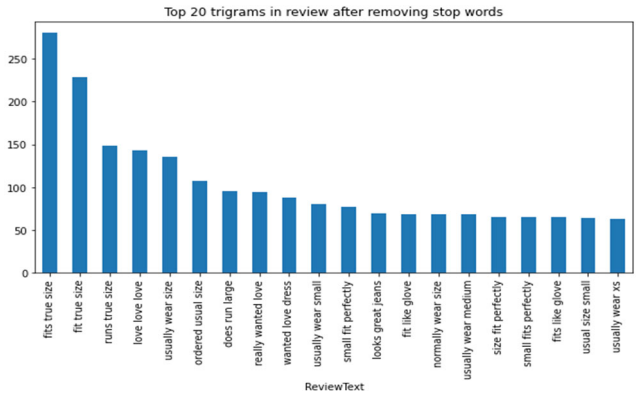


Fig. 14 Top 20 unigrams, bigrams, trigrams in reviews

4.2 Building bags-of-words medol

Having creating the corpus, the model could be build mathematically. The advantage of the bag-of-words model is that it is simple and clear, each document has a corresponding vector, regardless of word order, and the calculation speed is fast. Sklearn's CountVectorizer is used for text-to-vector conversion. Finally I get matrix X.

$$X = [n_1, n_2, n_3, \dots, n_k] \tag{1}$$

where n_i denotes no.i document .

The purpose of the model is to use user comments to predict the user's recommendation willingness, so the two columns of data are taken out and divided into training set and test set according to the ratio of 75% and 25%. Use CountVectorizer to train the training set to get the model. The model was tested with the test set data and the accuracy was 0.87.

4.3 Medol Optimization

However, using word frequency as the weight of text vectorization will lead to some words with low discriminative power but high word frequency occupy high weight. Therefore, this paper attempts to use the TF-IDF algorithm to optimize the transformation results of the bag-of-words model. Matrix X is weighted using Sklearn's TfidfVectorizer. Retraining with the Naive Bayes classifier yields a model accuracy of 0.92. As the accuracy of the model increased by 0.05 after using the TF-IDF algorithm for optimization, this optimization idea is reasonable and valuable. At the same time, the high-frequency elements of the text set are classified and summarized, and the factors that affect consumers' willingness to recommend can be divided into 5 categories, namely size, color, texture, style and others. The importance of its impact on consumers is shown in Table. 3 and Fig 15.

Table. 3 Factors of consumer recommendation willingness

factors	importance(%)
size	67.12
color	11.88
texture	10.1
price	8.23
others	2.67

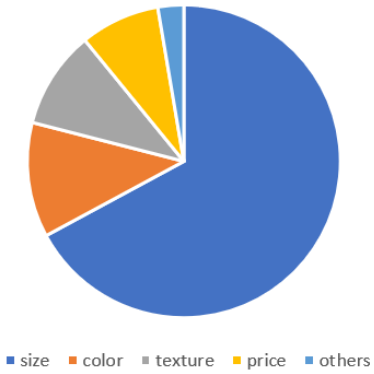


Fig.15 Proportion plot of factors of consumer recommendation willingness

5. Summary

In summary, this paper investigates the relationship between consumers' online shopping reviews and their willingness to recommend based on Amazon women clothing reviews. A total of 23,486 pieces of data were collected in this study, and sentiment analysis was conducted on consumer online shopping reviews. The results of the analysis showed that reviews are willing to recommend a product to others tend to have higher polarity. Then this paper uses the bag of words model and Naive Bayes classifier. A predictive model is built to predict consumer recommendation willingness based on reviews. The accuracy of the final optimized model reached 0.92. According to the analysis, the impact factors in the comments can be divided into five categories: size accounts for the highest proportion, color and texture second, and price and others the smallest.

Moreover, based on the research results, two recommendations are made to women's clothing merchants on the Amazon platform. First, increase the dimension of clothing size data. Clothing that fits perfectly can often improve consumers' evaluation of the product. Adding clothing size data such as sleeve length, shoulder width, bust circumference, and hip circumference allows consumers to better choose the size without trying it on. Second, use real lighting and filters to take photos of clothing. Consumers usually hope that there is no color difference between the pictures they see online and the actual product or the color difference is as small as possible. The use of beautifying filters may increase the probability of consumers buying clothing, but it will reduce consumers' evaluation of products after purchase.

Nevertheless, the analysis in this paper has some shortcomings and defects. First, the impact of the word order of comments on semantics is not considered. It creates a fallacy in the semantics of some comments while simplifying the model. Second, the amount of data is relatively small and the source is single. Consumers on different e-commerce platforms have different consumption habits and review habits. Just using the data from the Amazon platform ignores the possible impact of this. Third, the relationship between the number of product recommendations and sales has not been studied. Consumers' willingness to recommend is not equal to product sales. Only by clearly quantifying the relationship between the two can reviews and merchants' revenue be directly linked.

In the future, the research direction of online shopping reviews will still be dominated by big data and text mining. NLP algorithms such as LDA and RNN will become the mainstream research methods in this field. Overall, these results offer a guideline for women's clothing e-commerce, which is of great significance for the efficient classification and utilization of online shopping reviews.

References

[1] Hu Jueliang, Xu Yaoyao, Dong Jianming. Consumer focus intelligence analysis based on clothing online shopping comments. Journal of Zhejiang Sci-Tech University(Natural Science Edition), 2021, 45(01): 21-30.

[2] Li Ang, Zhao Zhijie. A review of online review usefulness research. Business Economics Research, 2019(6): 77-80.

[3] Tarraf P, Molz R. Competitive Intelligence at Small Enterprises. SAM Advanced Management Journal, 2006, 71.

[4] Liu Hongwen, Li Xiaohong, Nurul Hanim Romainoor. Meta-analysis of the influence of online reviews on clothing purchase intention. Journal of Silk, 2021, 58(01): 59-71.

[5] Liu Wenxin, Li Zhenfei. An Empirical Study on the Impact of Online Commodity Reviews on Impulsive Consumption of Online Shopping. Shopping Mall Modernization, 2020(08): 27-29.

[6] Chen Youqing, Li Xiufei. The relationship between online reviews and purchase intention:Analysis of a moderated mediation model. Hubei Agricultural Sciences, 2021, 60(01): 187-191.

[7] Zhang Shilin. Product attribute extraction from Chinese online shopping reviews based on Bi-LSTM and CRF. Computer and Modernization, 2019(02):93-97.

[8] Li Taoying, Li Feng, Lv Xiaoning. Analysis of structure characteristics of high frequency word co-Occurrence network of online shopping reviews. Computer Application Research, 2019, 36(01): 53-57.

[9] TextBlob library. Retrieved from: https://textblob.readthedocs.io/en/dev/

[10] NLTK library. Retrieved from: https://www.nltk.org/