Problem Chosen	2020	Team Control Number
C	MCM/ICM Summary Sheet	**2001335**

Evaluation and Prediction

Summary

Sunshine company plans to market and sell three new products in the online marketplace:a microwave oven, a baby pacifier, and a hair dryer. We establish a model to both provide marketing strategies and identify potentially important design features.

Our model takes into account the following factors:star_rating, help_votes, totle_votes, review and verified_purchase. Before establishing the model, we used verified_purchase for data cleaning and deleted reviews that reviewers didn't purchase.Then we define helpful_votes and total_votes as a useful evaluation coefficient to reduce the number of influencing factors. Finally, we use Natural Language Processing(NLP) to quantify the emotional score of reviews. We develop necessary assumptions and essential symbols to establish the model.

Firstly, information entropy is a value that measures the amount of information carried by a set of data. We use the information entropy of star_rating, help_factors and review_score as a reference to evaluate the amount of information contained in each factor. In order to define the score that describes product evaluation, we establish an AHP model (The Analytic Process). By comparing the importance to the score among the three factors, we assign certain weight to each factor, so as to get the evaluation system.

Besides, we found that the fluctuation of product sales was related to seasons, so we established a time series model to predict the future sales of this product. At the same time, we also predicted the reputation of the product. Comparing the two curves, we found that the change of the reputation and sales about the product is related to the season, so we can find the peak of the reputation and sales prediction curves, and formulate the marketing strategy.

After that, we defined a screening model to screen out evaluations with a higher evaluation score than the preset value, that is, successful products. And then select the top ten products from the sales, and extract the characteristics that consumers love from these successful products, that is, mining success potentially important design features of the product and the results are given to provide a reference for Sunshine Company.

Finally, we used Pearson correlation coefficient to analyze the correlation between review and star_rating, and found that they have weak correlation.

After these works, we analyzed the sensitivitystrengths and weaknesses of the model and write a letter to CMO of Sunshine Company, reporting our analysis and results based on our model, and propose some marketing strategies for marketing products and design features.

Keywords: Natural Language Processing (NLP);Naive Bayes;The Analytic Process (AHP) Model;Time series model;Correlation analysis

Contents

1 **Introduction** 1

1.1 Background 1

1.2 Restatement of Problem 1

2 **Analysis of the Problem** 1

3 **Symbols** 2

4 **Simplifying Assumptions** 3

5 **Date Processing** 3

6 **Model and Solution** 3

6.1 Natural Language Processing for sentiment analysis 3

6.1.1 Naive Bayes Method for Emotional Analysis 3

6.1.2 Review Emotion Quantification 5

6.2 AHP Model based on information entropy 6

6.2.1 Model Building 6

6.2.2 Model Preparation 6

6.2.3 Model Solution 7

6.3 Reputation prediction model 8

6.3.1 Model Preparation 8

6.3.2 Time series prediction model 8

6.3.3 Model Solution 8

6.3.4 Results Analysis and Sales strategy 9

6.4 Screening model 10

6.4.1 Screening criteria 10

6.4.2 Analysis of results 11

6.5 Correlation analysis model 12

6.5.1 Correlation calculation 12

6.5.2 Result analysis 12

6.5.3 Relationship between Specific Quality Description and Rating Leveling 12

7 Sensitivity Analysis **13**

8 Strengths and Weakness **14**

8.1 Strengths 14

8.2 Weakness 14

9 References **17**

Appendices **18**

1 Introduction

1.1 Background

With the development of e-commerce and the improvement of the level of logistics services, more and more people choose to shop online this way. In order to give customers a better shopping experience, e-commerce websites have introduced shopping rating and review systems.Amazon established an online review system as early as 1995, and has developed a relatively complete review system to date. In this system, buyers can not only express their satisfaction with the products by using ratings, but also write reviews to further express their opinions on the products. These ratings and reviews allow sellers to understand market demand and development rules in a timely manner, help them improve product quality and formulate online sales strategies, and then make them successful in the online market.

Therefore, considering that Sunshine expects to launch three products in the online market, a microwave oven, a baby pacifier, and a hair dryer, we think it is necessary to determine the correct marketing strategy and the characteristics of successful products. .

1.2 Restatement of Problem

Consider the background information,the requirements identified in the problem statement,and the information provided in the problem attchment to address the following issues:

1. It is necessary to establish an evaluation metric model to help sunshine company succeed in the three products listed in the future. The model will be able to evaluate the products in the current market through star_rating, helpful_votes, review_score, and use the final score last_score to indicate the product's score in this shopping.
2. Study the relationship between prestige and time, and establish a model that can predict the prestige trend of the product in the future time, so as to understand the future prestige of the product in the market and provide reference for deciding the future sales strategy.
3. Use a combination of text-based and rating-based metrics as a condition to indicate potential successful and failed products, and determine the characteristics of successful and failed products
4. Discuss the relationship between star_rating and review
    - whether star_rating will cause a specific type of review
    - whether the specific review is related to star_rating.

2 Analysis of the Problem

Determining the correct marketing strategy and extracting successful product characteristics for Sunshine Company can be divided into four sub-problems:

problem 1:Set up indicators to measure merchandise scores. According to the given data, we first quantify the three data that are of reference value to the product experience: text comments, stars and voting usefulness coefficient. Our main purpose is to compare the importance of these three factors to the score through the evaluation model and obtain the weight corresponding to the three factors.

problem 2: According to the data in problem one, the average evaluation score of the product in a period of time is calculated to express the reputation score of the product. then we establish a Time series forecasting model to fit and determine the relationship between reputation and time, predict the reputation of the product in the future according to the previous product reputation data, and finally analyze the reputation changes of the product

problem 3:The key to the problem is to set up a score measurement standard to evaluate the success or failure of commodities, so as to screen out successful products and failed products, and to extract the characteristics of products from the title of products and user comments in combination with product sales.

problem 4:After the text reviews are quantified, correlation analysis can be carried out on the text reviews and the star marking rate, and whether high-star data will cause good comments and low-star data will cause bad comments can be analyzed. Secondly, emotional words in comments are extracted to analyze whether comments with high scores will lead to high stars and comments with low scores will lead to low stars.

3 Symbols

Definitions	Descriptions
H	Useful evaluation coefficient
$R_{i,j}$	Indicates the reputation value of the product
P	Comprehensive satisfaction score
$\alpha_{i,j}$	A month's total comprehensive evaluation score
$\beta_{i,j}$	Sales per month
s_t	Periodic term
Y_t	Random noise term
d	cycle
$P_{i,j}$	Represents the posterior probability of the j^{th} word in the i^{th} review
X	The word set of review
Y	Contains two categories Positive and Negative, represent the emontional color
m_t	Trend item
n	Year

4 Simplifying Assumptions

1. Because the sales volume in the previous few years in the data is small, based on the fact that online shopping was not yet popular at the time, the sample was considered less, and the effect of the years with less sales in the previous years in the data was ignored.

2. Ignore the inconsistent meaning between the review headline and the review body.Text-based metrics depend on the review body

3. Ignore the differences in the popularity of online shopping in different periods

4. Accept the errors caused by text analysis due to ambiguity of some words in the reviews.

5. When analyzing the review text, ignore the impact of different word order on the prediction of the emotional tendency of the review. That is, we assume that "I love this hair dryer" and "this hair dryer love me" have the same emotional color.

5 Date Processing

For some data whose verified_purchase is N and not vine, it should be cleared first. In fact, reviews of unpurchased products are not informative and may affect the evaluation results.

According to helpful_votes and totle_votes, useful evaluation coefficients H can be defined.

$$H = 1 + \frac{helpful\_votes}{totle\_votes + 1} \cdot \lg(totle\_votes + 1) \tag{1}$$

6 Model and Solution

6.1 Natural Language Processing for sentiment analysis

In order to extract the information of commodity evaluation from the comments of customers, we need NLP (natural language processing) to extract the emotion-related words in the reviews, and judge the emotional tendency of buyers from these words. Progress quantifies the content of the reviews. We will use this quantified value as a measure of customer satisfaction.We mainly use naive bayes to achieve the above goals

6.1.1 Naive Bayes Method for Emotional Analysis

First of all, we want to get the goal that we can predict the positive or negative probability of the emotion of a review through a word in the review, so as to determine whether the review is positive or negative. In the field of Natural Language Processing (NLP), Naive Bayes is usually used to Text-Categorization, and it is direct and efficient when dealing with problems. Therefore, we use Naive Bayes to classify user reviews.

When the Naive Bayes method classifies words into emotions, the learned model is used to calculate the posterior probability distribution of the input word x, and the class with the greatest posterior probability is used as the class output of the input word x.

Suppose $x \in X$, X is a collection of review words,$c_k \in Y$, $Y = \{Pos, Neg\}$ represents both positive and negative comments.$Y = 1$ represents this review as a positive comment, while $Y = 0$ represents this review as a negative review.Then we need to get its posterior probability $P(Y = c_k|X = x)$,According to Bayes formula,we have:

$$P(Y = c_k|X = x) = \frac{P(X = x|Y = c_k)P(Y = c_k)}{\sum_k P(X = x|Y = c_k)P(Y = c_k)} \tag{2}$$

Because $\sum_k P(X = x|Y = c_k)P(Y = c_k) = P(X = x|Y = Pos)P(Y = Pos) + P(X = x|Y = Neg)P(Y = Neg)$ is a fixed value,the problem we solved becomes the following optimization problem:

$$y = argmax_{c_k} P(Y = c_k)P(Y = c_k)P(X = x|Y = c_k) \tag{3}$$

Among them,$y \in Y$ is the category with the greatest posterior probability of x,We output this category as the category of the word. We find a labeled corpus of Amazon reviews on the kaggle website[3] as a training set to train our Bayesian classifier, and then use it to analyze the sentiment tendency of reviews.

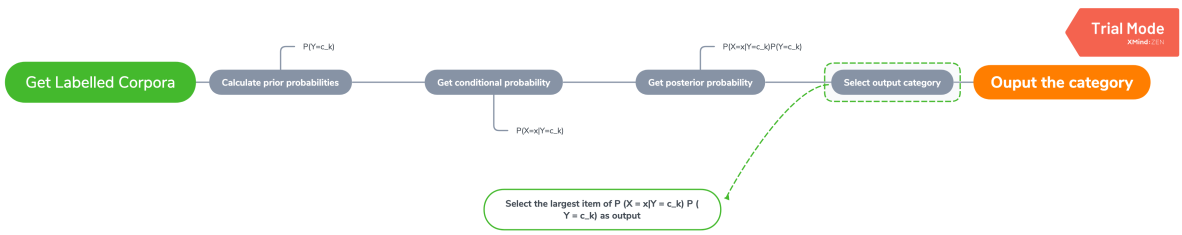

Figure 1: Bayesian classifier training process

After training, our Bayesian classifier can obtain a word cloud with positive evaluation and a word cloud with negative evaluation respectively according to classification accuracy.

Figure 2: Positive

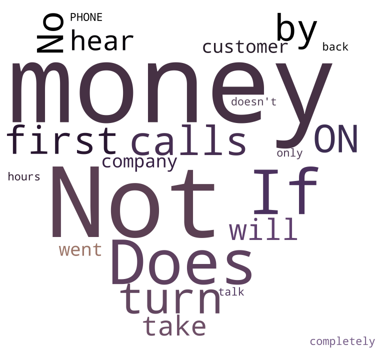

Figure 3: Negative

The larger the size of the word, the more effective the word is in classification. According to the above words with good classification effect, we can visualize their corresponding posterior probabilities:

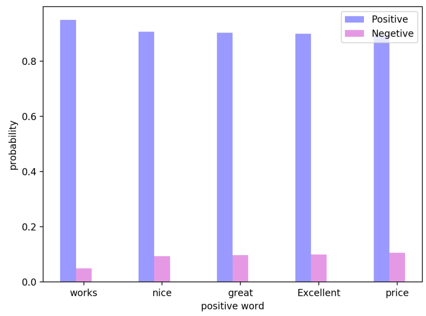

Figure 4: Positive

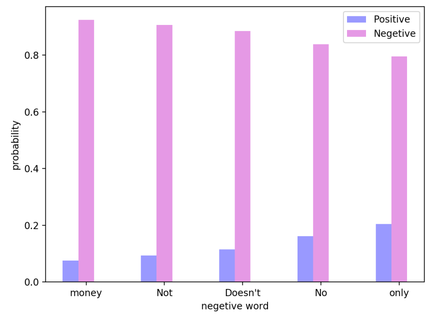

Figure 5: Negative

Based on these words with significant classification effect, we can quantify the emotional tone of the reviews.

6.1.2 Review Emotion Quantification

When we quantify the review, for each word in the review, we will calculate the probability that the review is positive and the review is negative if it exists in the review. Then we will give the review a "customer satisfaction" score according to the following formula.

Suppose y_i represents the score of the review, $P_{i,j}$ represents the posterior probability of the j^{th} word in the i^{th} review, and C is a constant, then the quantization formula is

$$y_i = \log \sum_{j}^{n} P_{ij} + C - \log(C) \tag{4}$$

$$P_{ij} = \begin{cases} P(Y = Pos|X = x_j) & \text{if } P(Y = Pos|X = x_j) \geq 0.5, \\ -P(Y = Neg|X = x_j) & \text{if } P(Y = Pos|X = x_j) < 0.5 \end{cases} \tag{5}$$

$$C = \min_i \sum_j P_{ij} \tag{6}$$

We propose this quantitative formula mainly based on the following considerations and assumptions:

- Although not every word in the commentary is of great help for judging the emotion of the commentary, at the same time, the posterior probability calculated from these words is generally close to 0.5, which is smaller than the posterior probability calculated from words with significant effect (generally close to 0.9 or 0.1). On the whole, these irrelevant words will not have great influence on the overall quantified value.
- For words whose emotion is positive, we will impose a "reward"; We will impose a "punishment" on words whose emotions are negative. This is why we add this probability to the overall score when the positive posterior probability is greater than 0.5 and subtract the corresponding negative posterior probability when the probability is less than 0.5.
- In order to make the score more gentle on the whole, we use logarithmic function to

reduce the influence of extreme values on the whole, and also reduce the variance of reviews on the whole, because it is very difficult to accurately quantify the text emotion. In order to realize this operation, we add a constant c after summing the posterior probabilities of words in the review, so that all sums are greater than 0, and then subtract the logarithm of this constant. in order to make this score fluctuate up and down, if the score is greater than 0, it means that the review tends to be positive, and if the score is less than 0, it means that the review tends to be negative.

6.2 AHP Model based on information entropy

6.2.1 Model Building

Through the above model, we have completed the quantification of the review text and obtained a satisfaction score from the review. In order to obtain the standard to measure the product situation and objectively evaluate the reputation of the product, we have established a AHP model, which focuses on the following three aspects:

- Star_rating: A means used by consumers to score products in an evaluation system.The website will show the average score on the product page, and each comment will show the score. When a consumer purchases a product, the score will affect the consumer's purchase intention and thus affect the sales volume of the product, so it is taken into account.
- Useful evaluation coefficient H: In the evaluation system, the index that reflects the quality of the review.Consumers will browse product reviews when purchasing products, and vote on reviews when they think the reviews are useful. The more votes a comment has, the more reference it will have.The formula (1) obtains the coefficient by combining helpful_vote and total_vote.
- Review: [1]In the evaluation system, the way consumers describe the experience of purchasing products.When consumers have a desire to buy this product, they will know more about the product information in detail, including comments. The quality of the comments will increase or decrease consumers' desire to buy, so they will be taken into consideration. Here, review has been quantified as a review_score.

6.2.2 Model Preparation

In order to reduce the influence brought by the subjective effect of AHP model, we use information entropy to help determine the weight of three factors. Information entropy is a measure that represents the uncertainty of random variables. Let X be a discrete random variable with a finite number of values, and its probability distribution is:

$$P(X = x_i) = p_i \quad i = 1, 2, ..., n \tag{7}$$

The information entropy of random variable X is defined as:

$$H(x) = -\sum_{i=1}^{n} p_i \log p_i \tag{8}$$

The greater the value of information entropy, the greater the uncertainty of random variables. However, the amount of information contained in a random variable is directly related to its uncertainty. For example, we need to know a lot of information if we want to find out what we don't know or what we don't know. So from this point of view, it can be considered that the amount of information is equal to the amount of uncertainty. Because we can use information entropy to measure the information contained in a random variable.

According to the above principle, we calculated the information entropy of the three factors of the three products.The results are as follows:

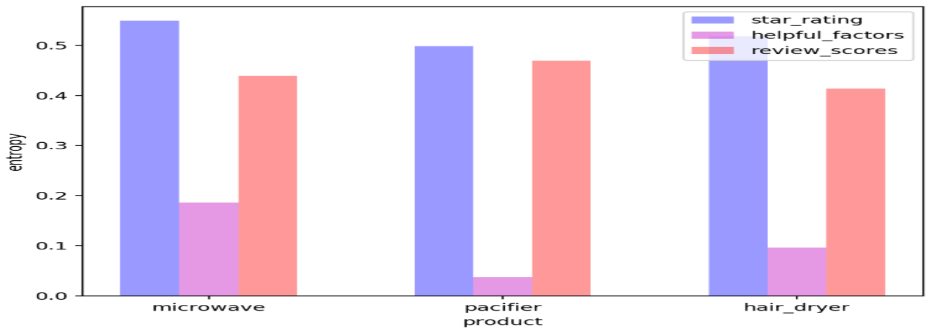

Figure 6: information entropy

6.2.3 Model Solution

According to the information entropy of the three factors, star_rating has the largest information entropy and the most hidden information, so we regard it as the most important.However, helpful_factor(Useful evaluation coefficient) has the smallest information entropy, that is, gives helpful the smallest proportion. According to the difference of information entropy of different factors, different proportional relations are given.The comparison matrix is as follows:

Aspect	star_rating	H	review_score	Weight
star_rating	1	3	2	0.5390
H	1/3	1	1/2	0.1638
review_score	1/2	2	1	0.2973

By calculating the weight of the three factors, the maximum eigenvalue $\lambda = 3.0092$ is obtained. Consistency ratio $CR = 0.0079$,$CR < 0.1$,therefore, consistency is correct.Through the weight of the three factors, we can get the comprehensive satisfaction score of the product,Its symbol is P.

6.3 Reputation prediction model

6.3.1 Model Preparation

Based on the respective weights of star_rating, H, and review_scores obtained through the AHP model, we first normalized the respective scores of star_rating, H, and review respectively, and multiplied by the corresponding weights, we can obtain the comprehensive evaluation score for each purchase.According to the actual situation, we define the average evaluation score of the product within one month as reputation value.

Let the $\alpha_{i,j}$(i for month, j for year) is a month's total comprehensive evaluation score,$\beta_{i,j}$ is one month's sales.The reputation value for the month is $R_{i,j} = \dfrac{\alpha_{i,j}}{\beta_{i,j}}$.

6.3.2 Time series prediction model

In order to discover the inherent regularity of reputation data changes with time, and then make judgment and prediction, we have constructed a time series prediction model.We first calculate the monthly reputation values $R_{i,j}$ in the three data sets.Taking time as dependent variable and reputation value as dependent variable, the change of reputation value $R_{i,j}$ of future products can be predicted through model fitting curve.The typical decomposition formula of a time series is:

$$X_t = m_t + s_t + Y_t \tag{9}$$

Where m_t is the trend term and s_t is the periodic term (seasonal term) with known period d; Y_t is a random noise term First of all, we found that there is no obvious trend in the time series of various products' reputation-time relationship diagram, so we can remove the trend item m_t, while for the seasonal item Y_t, we observed that seasonal changes have a significant effect on the sales volume and score of products, thus the seasonal item can be retained. The above formula is simplified as follows:

$$X_t = s_t + Y_t \tag{10}$$

6.3.3 Model Solution

$$R_{i,j}(i = 1,2,...12; j = 1,2,...n)$$

There are n years of data, 12 data per year.

1. Select seasonal items.

   - Calculate the average value of the j-th year.

   $$\overline{R_j} = \frac{\sum_{j=1}^{12}}{12} \tag{11}$$

   - Zero-averaging monthly data

   $$st_{i,j} = R_{i,j} - \overline{R_j} \tag{12}$$

- The seasonal items are

$$s_i = \frac{\sum_{j=1}^{n}}{n}(i = 1,2...12) \tag{13}$$

2. Acquire data after season item are removed.

$$Y_{i,j} = R_{i,j} - s_i(i = 1,2,...12; j = 1,2,...n) \tag{14}$$

$$Z = (Y_{1,1}, Y_{2,1}, ...Y_{12,1}, Y_{1,2}, Y_{2,2}, ..., Y_{12,2}, ..., Y_{1,n}, Y_{2,n}, ..., Y_{12,n}) = z(z_1, z_2, ..., z_{12n}) \tag{15}$$

3. Regression fitting Polynomial regression was used to fit the data $z_1, z_2, ..., z_{12n}$

4. Forecast

   The predicted value for the next 12 months after eliminating seasonal items is

$$\widehat{z}_{12n+1}, \widehat{z}_{12n+2}, ..., \widehat{z}_{12n+12}$$

That is $\widehat{Y}_{1,n+1}, \widehat{Y}_{2,n+1}, ..., \widehat{Y}_{12,n+1}$ The predict value in that original data for the next 12 months is $\widehat{R}_{i,n+1} = \widehat{Y}_{i,n+1} + \widehat{s}_i(i = 1,2,...12)$

6.3.4 Results Analysis and Sales strategy

The following are the results of prestige predicted by the time series model of microwave, pacifier, and hair_dryer.

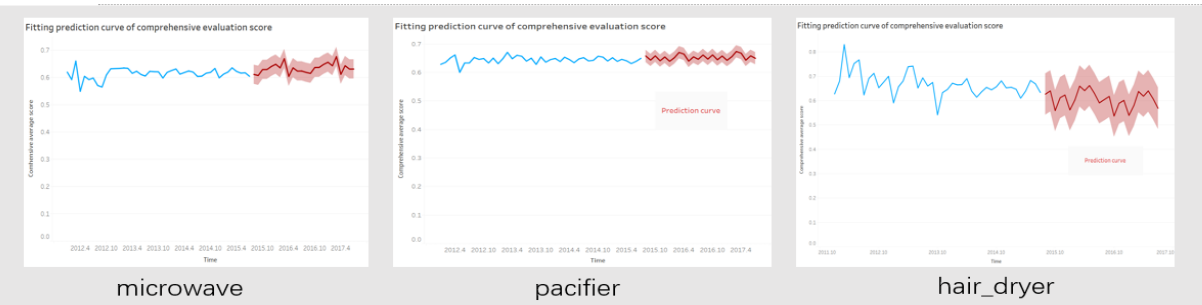

Figure 7: Prestige forecast curve

The following are the sales forecast results of microwave, pacifier and hair_dryer

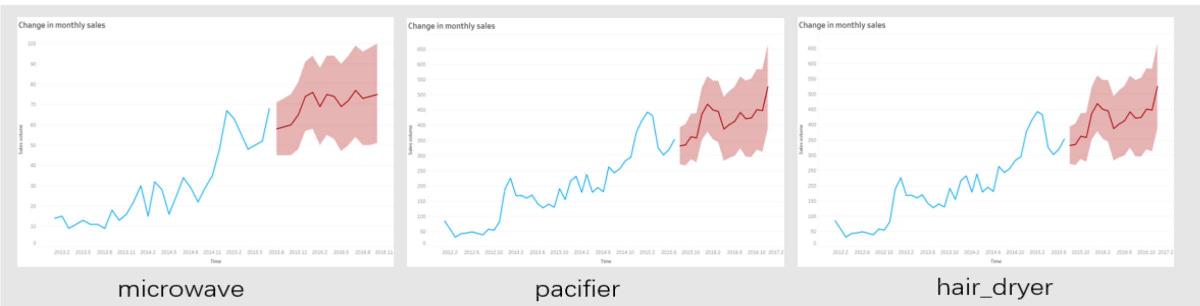

Figure 8: Sales volume forecast curve

Considering that the e-commerce platform was not mature enough in previous years and the number of users who bought online was relatively small, the reference value of product sales in previous years was relatively small, and only the product sales in recent years were analyzed.

- **Microwave:** According to the results, the popularity and sales volume of microwave have seasonal changes, and the sales volume generally increases in autumn and winter. That is, due to the cold weather, consumers have greater demand for microwave in autumn and winter, and the popularity score of products generally increases, i.e. consumers have higher reviews on products and better shopping experience. Therefore, Sunshine Company can increase its production in the coming autumn and put it on the e-commerce platform in the autumn, which will lead to better sales volume and reputation. The higher the prestige, the more consumers will be attracted to buy, realizing a virtuous circle.
- **Pacifier:** According to the data research of *Journal of Applied Econometricsz*, American mothers generally tend to give birth in spring. The results show that pacifier's sales volume is relatively stable, but it will reach a small peak in spring, and pacifier's reputation score will also reach a peak in spring. It can be inferred that pacifier is more likely to be loved by Baoma and has better evaluation in spring. Therefore, sunshine Company can produce pacifier near spring and put it on the e-commerce platform in spring, which will get better sales volume and reputation.
- **hair_dryer:** From the analysis of the results, it can be seen that the sales volume of hair_dryer fluctuates in the season, but it will reach the peak within one year in winter, and its reputation will peak in winter. Because the temperature is lower in winter, consumers need hair_dryer to blow their hair. In contrast, the summer temperature is higher and the consumer's demand for hair_dryer is lower. Therefore, sunshine Company can produce pacifier near winter and put it on the e-commerce platform in winter, which will get better sales volume and reputation.

  According to the above analysis, the final marketing strategy results are as follows:

	Spring	Summer	Fall	Winter
Microwave	NO	NO	YES	YES
Pacifier	YES	NO	NO	NO
Hair_dryer	NO	NO	NO	YES

Table 2: Time schedule for product launch

6.4 Screening model

6.4.1 Screening criteria

We use comprehensive satisfaction score P to measure successful products and failed products:

Successful product	P>0.65
Failed product	P<0.3

We select the data defined as successful products, and make statistics on the same product_parent, extract the top ten product_title of sales volume, and analyze the characteristics of successful products.

6.4.2 Analysis of results

Microwave product_title	Sales volume
danby 0.7 cu.ft. countertop microwave	180
microwave cavity paint 98qbp0302	55
whirlpool wmc20005yw countertop microwave, 0.5 cu. ft., white	52
whirlpool stainless look countertop microwave, 0.5 cu. feet, wmc20005yd	45
whirlpool wmc20005yb 0.5 cu. ft. black countertop microwave	41
sharp microwave drawer oven	32
sharp 950-watt 1-2/5-cubic-foot over-the-range microwaves	27
samsung counter top microwave	26
ge microwave oven magnetron and diode kit om75p (10) part # wb27x10017	24
samsung mc11h6033ct countertop convection microwave with 1.1 cu. ft. capacity, slim fry technology, grilling element, ceramic enamel interior, drop down door, and eco mode in stainless steel	16

Products	Features
Microwave	Volume:0.5-0.7 cu.ft. Type:countertop Material:stainless steel
Pacifier	Preformance:free soothie Exterior:animal shaped
Hair_dryer	Type:ionic hair dryer Functionfold-n-g Color:deep purple,silver,black,pink

Two tables of statistics on hair dryers and pacifiers are attached to the appendix.

From the above statistical information of successful products, the characteristics that a successful product should possess can be extracted:

6.5 Correlation analysis model

6.5.1 Correlation calculation

In order to analyze the relationship between the star_rating and review, we made a correlation analysis between the star_rating and review_score Since the overall data distribution of the star marking rate and the quantitative value of the comments is close to the normal distribution, we use Pearson correlation coefficient to measure the correlation between the two.

$$\rho_{X,Y} = \frac{\sum(X-\overline{X})(Y-\overline{Y})}{\sqrt{\sum(X-\overline{X})^2\sum(Y-\overline{Y}^2)}} \tag{16}$$

$\rho_{X,Y}$	Range
0.8-1.0	Very strong correlation
0.6-0.8	Strong correlation
0.4-0.6	Moderate correlation
0.2-0.4	Weak correlation
0.0-0.2	Extremely weak correlation

By calculation,$p_{X,Y} = 0.291$

6.5.2 Result analysis

Compared with the table, it is weakly correlated. That is, when the star_rating is low, it will be accompanied by bad reviews, and when the star_rating is high, it will be accompanied by good reviews. Consumers are easy to be influenced by the first impression of reviews before purchasing goods. If there is any deliberate bad review by black powder, the reputation of the product will decline. Therefore, Sunshine Company can invite vine to experience the product and write reviews with reference value for the reference of consumers, leaving a good first impression on consumers. Moreover, vine's reviews have more reference value and consumers are more willing to refer to them.

6.5.3 Relationship between Specific Quality Description and Rating Leveling

For some words with emotional colors (such as great, like, well, good, awesome), we counted the star ratings of excellent (4-5 stars) and poor (1-3 stars) when these words appeared in the reviews. Probability, we found that for some words with strong emotions, the rating rating is often closely related to these words.

For several words with positive emotions and negative emotions, we counted the probability that these words appear in the reviews with high and low quality star rating:

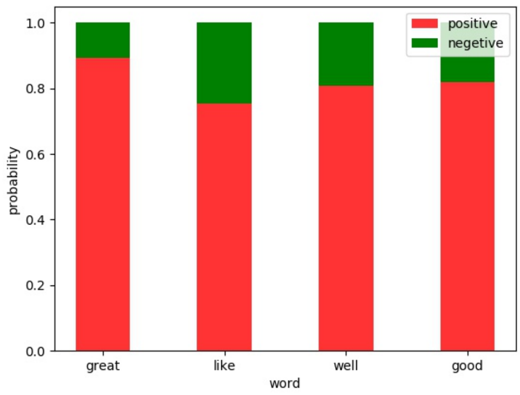

Figure 9: Positive word

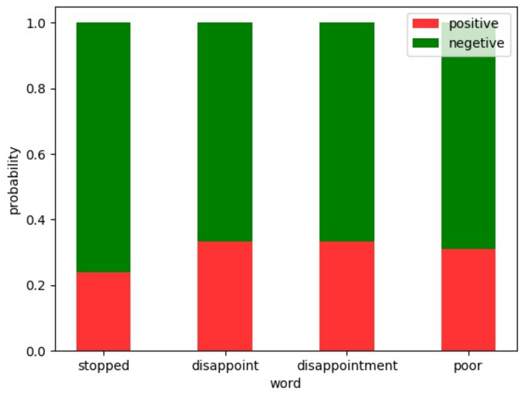

Figure 10: Negative word

From the figure, we can see that these words have a relatively large degree of discrimination in scoring. If these words appear in comments, the corresponding scoring will often be linked to their positivity or negativity. However, not all words have similar discrimination, such as some unrelated words (price) or some neutral words (much). These words have no discrimination for scoring, so we think these words have no direct relationship with the rating level.

7 Sensitivity Analysis

We extract the potential design functions of successful products according to the top ten successful products in sales volume, and define successful products as determined by star_rating, review_score. these three factors will lead to differences in the selected successful products.

Firstly, we increase the weight of one factor and equally remove the difference to the other two factors. We make the same progress for the other factors.Secondly, we decrease the weight of one factor and equally remove the difference to the other two factors. We make the same progress for the other factors.

We show the final sales volume of the top four successful products in the figure below.

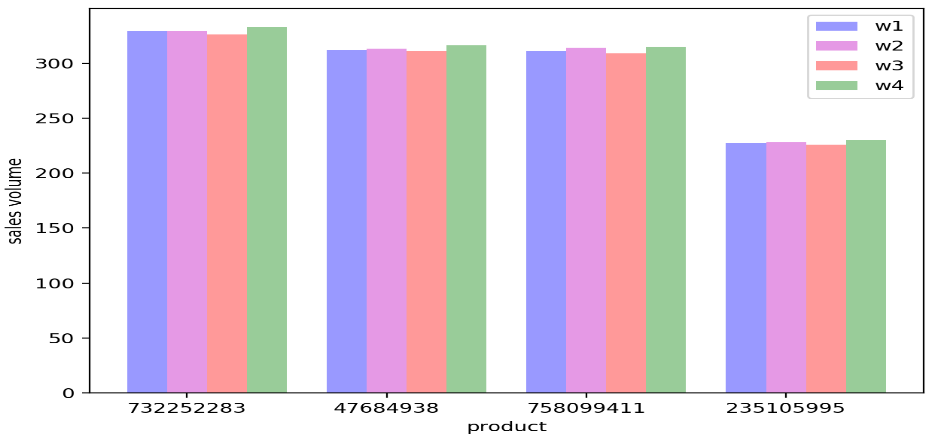

Figure 11: Final sales chart

We found that after changing the weight, the sales volume of the four selected successful products showed little change trend, indicating that the public still had the desire to purchase these products. In fact, the extracted potential design functions are basically unchanged. Therefore, our model is reliable.

8 Strengths and Weakness

8.1 Strengths

- **The objectivity of quantitative results is good**
  We quantified the text reviews objectively, and the data obtained are relatively stable and reliable, which enables the text reviews to participate in the measurement of commodities in the form of scores and objectively comment on the quality of commodities.

- **The prediction results are relatively accurate**
  We use the given data to predict the time series and consider the seasonal influence. The model fits the curve of product reputation with time better and the prediction is more accurate.

- **High degree of visualization**
  Our data processing and results are displayed with a large number of charts, which are highly visible, readable and easy to understand.

- **High flexibility**
  On the basis of AHP, we can also change the weights of star_rating, helpful_votes and review according to our own needs to meet different evaluation needs.

8.2 Weakness

- **Ignoring vine's role**
  In the process of establishing AHP model, we considered vine's comments to be consistent with those of ordinary consumers, ignoring vine's possible guiding role. However, vine's comments generally have a large number of votes, so vine's influence has been introduced into the model through useful evaluation coefficient H, so ignoring vine's influence on the model will not be very big.

- **Ignore the characteristics of commodities outside the top 10 sales volume**
  The total number of products in the top ten sales is close to about half of the total. They represent the popular direction of the public, and they have characteristics that are valued by the public. Therefore, ignoring the characteristics of products other than the top ten sales will not have a great impact.

A letter to Sunshine Company
March 9,2020

Dear CMO of Sunshine Company:

Hearing that Sunshine Company is planning to introduce and sell three new products in the online marketplace. We are more than glad to share our mathematical model to give some recommendations.

We strongly advise you to market different products in different seasons. The results of which products market in different seassons have been shown in the following table. The reason for this is that, based on our time series prediction model, we find that product sales and reputation vary seasonally and peak in a given season. In order to get a good reputation for your company's products when they are marketed, we suggest you to regulate the time of their marketing.

After considering the top 10 successful products in sales, we extracted the potential design features of successful products from their titles and reviews. The results of the potential design features of each product are given in the following table. We strongly recommend that you take these characteristics into account in your marketing products, which will help you achieve higher sales.

Finally, we found that reviews with more votes had more impact. Therefore, we suggest that your company invite vine to experience and evaluate the product after its marketing, which will give consumers more reference information and promote their purchase.

We hope that our model can help Sunshine Company to sell the products well. And we also hope that your product will get better.

Products	Features
Microwave	Volume:0.5-0.7 cu.ft.
	Type:countertop
	Material:stainless steel
Pacifier	Preformance:free soothie
	Exterior:animal shaped
Hair_dryer	Type:ionic hair dryer
	Functionfold-n-g
	Color:deep purple,silver,black,pink

Table 3: Characteristics of Successful Commodities

	Spring	Summer	Fall	Winter
Microwave	NO	NO	YES	YES
Pacifier	YES	NO	NO	NO
Hair_dryer	NO	NO	NO	YES

Table 4: Time schedule for product launch

Yours sincerely,

Team # 2001335
MCM2020

9 References

References

[1] wang junkui. Research on the usefulness of online comments on e-commerce websites [D]. Xi 'an: xidian university of science and technology, 2014.

[2] Gao Baojun, king haemorrheological nature, yellow Tian, Hou Yangyang. Impact analysis online evaluation system for the sale of goods, based on the data of jingdong and Tmall [J]. Journal of price theory and practice, 2015 (8) : 103-105.

[3] `https://www.kaggle.com/marklvl/sentiment-labelled-sentences-data-set`

Appendices

Pacifier product_titile	Sales volume
philips avent bpa free soothie pacifier, 0-3 months, 2 pack, packaging may vary	563
wubbanub infant pacifier - giraffe	371
wubbanub brown monkey pacifier	354
mary meyer wubbanub plush pacifier, lamb	353
wubbanub brown puppy pacifier	206
fctry - baby pacifiers	165
philips avent bpa free contemporary freeflow pacifier	157
wubbanub lamb infant pacifier	155
wubbanub elephant pacifier	142
wubbanub tabby kitten pacifier	111

Hair_dryer product_titile	Sales volume
remington ac2015 tlstudio salon collection pearl ceramic hair dryer, deep purple	329
andis 1875-watt fold-n-go ionic hair dryer , silver/black (80020)	312
conair 1875 watt tourmaline ceramic hair dryer	311
revlon essentials 1875w fast dry hair dryer, rv408	227
conair corp pers care 146np conair ionic conditioning 1875-watt hair dryer	197
conair 1875 watt cord keeper 2-in-1 hair dryer, black	180
andis 1600w quiet hangup hair dryer with night light	160
andis micro turbo hair dryer	135
conair minipro tourmaline ceramic styler / hair dryer; pink	134
oster tourmaline wall mount hair dryer 76932-710	133