Problem Chosen	2020 MCM/ICM Summary	Control Team Number
C		2014597

Evaluation system and Time series model based on reviews

In the online marketplace, Amazon provides customers with an opportunity to rate and review purchases. In order to help Sunshine Company succeed in their three new online marketplace product offerings, we have built evaluation models and time series models based on ratings and reviews.

We firstly construct an evaluation model, make a quantitative analysis of the whole indicators, and choose six indicators for the principal component analysis. We find that the "star rating","helpful votes" are most likely to show the success of one product.

After that, in order to better quantify the customer's reviews into our system, we disassemble the words of each review sentence, and compare them with the commendatory and derogatory words in the semantic database to calculate the score of each review. And the final score of each product's review is determined according to the weight of the corresponding helpful votes of each review. We select the top ten products with reviews, normalize the two indicators of star rating and review's score, and then determine the weight of them using entropy weight method in order to build a complete evaluation model.

In order to find the relationship between star rating and the number of reviews and research the consumer's group psychology, we take the most popular product as the research object, first of all, get the correlation coefficients of the two variables. The time series analysis explains the reasons for fluctuations, according to the sequence diagram.

We use the previous evaluation model to calculate the score of the product in each month, and then use the expert model to automatically match the appropriate time series model based on the window sliding method.The overall trend of product change which is obtained through seasonal analysis is used to evaluate the success.

In order to predict a product's reputation effectively, we use K-nearest Neighbor combined with Bag of Words, Term Frequency - inverse document frequency, Word2Vec and tf-idf weighted Word2Vec to judge the most accurate model. Due to the limitation of data volume, we here only predict the reputation of the whole category of products. And we found that all the reputation will be increased, and the change of pacifier is biggest.

What's more, to figure out association between specific descriptors and rating levels, we extract the words from each comment sentence and get the decision coefficient through naive Bayes. The closer the absolute value of the decision coefficient is to 0, the word is more likely associated with a higher star rating.

Key Words: Natural Language Processing; Principal Component Analysis; Correlation Analysis; Time series; K-nearest Neighbor

Letter

Dear Marketing Director of Sunshine Company:

It's our pleasure to give some recommendations to you to help craft successful products. Next I will give our advice from 3 aspects: the quality of your products, the time to lauch your products, and the marketing strategy after your lauching.

First of all, we should always give priority to the quality of our products. We try to figure out the association between specific descriptors and rating levels, and get these words that more usually related to higher star rating:

Hair dryer: easy, powerful, small, light, quiet

Microwave: easy, large, big, powerful

Pacifier: easy, soft, cut, clean

From the key words above, we can draw a conclusion that customers are inclined to buy products that are easy to handle with good quality and effect. And some exclusive characters to corresponding products should also be considered: such as a quiet, small hair dryer, a high-capacity microwave and a soft and safe pacifier.

What's more, we deem that the time to lauch your product is also of great significance. We have counted the reviews of three kinds of products in each month, and found that January, August and December are the peak of the reviews, which to some extent indicates a high level of sales, so we suggest you choose to launch in these three months.

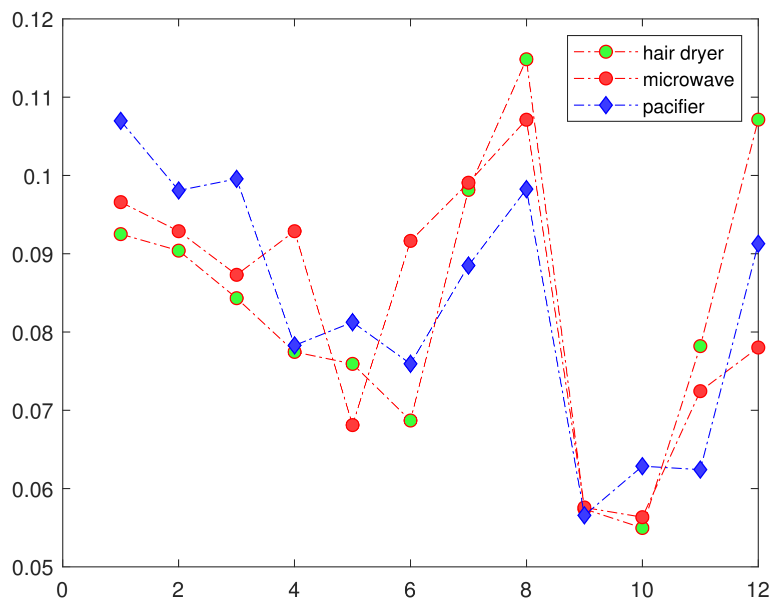


Figure 1: Proportion of monthly comments

Last but not least, to figure out which indicators that best show the success of one product, we build two evaluation systems, and find out the most informataive indicators below: star rating, helpful votes and the content of reviews especially those that have more helpful votes. People who have seen a low star rating are possible to express their dissatisfaction towards the product, because people are easily influenced by others' reviews, So our suggestion is to ensure the after-sales service of the products, solve the problems in time when they occur, and try not to make the comments of low stars last too long.

I'd appreciate it if you can take my advice into consideration, wish your products a success.

Sincerely yours

Team #2014597

Contents

1 Restatement and Clarification of the Problem

1.1 Problem Background

Living in an era when online shopping is quite generic, you are presumably purchasing your products online more or less. However, different from the traditional means of shopping, you can not see the goods face to face, thus it's of significance to judge from the reviews from other people who have bought already, three of which are star ratings, reviews, and helpfulness ratings.

Sunshine Company will introduce and sell three new products in the online marketplace: a microwave oven, a baby pacifier, and a hair dryer. Now we have got the the customer-supplied ratings and reviews associated with other competing products. Our goal is to find out the key patterns, relationships, measures, and parameters in the data to help formulate their online sales strategy and identify potentially important design features that would enhance product desirability.

1.2 Restatement of the Problem

We need to solve two main problems, the first is to figure out the quantitative or qualitative patterns and relationships between star ratings, reviews, and helpfulness ratings. The second is to a series of specific problems:

- Build a measuring system based on the ratings and reviews.
- Predict a product?s reputation is increasing or decreasing based on time patterns.
- Judge a successful or failing product by text-based measure and ratings-based measure.
- Figure out the relationship between star ratings and the number of reviews.
- Find out the association between specific descriptors and rating levels.

2 Explain Assumptions

2.1 Assumptions

- When using time series, we think the data is stable as a whole.
- The inconsistency between comments and stars is too small to influence the subsequent analysis.
- To some extent, the number of comments can reflect the sales volume.
- The score of reviews is only related to the meaning and quantity of commendatory and derogatory words.

3 Evaluation models

3.1 The evaluation model based on Principal Component Analysis

As we know, several indicators are related with a product's success. Therefore, the whole products' data are used to find out the indicators or parameters.

The indicators are dimensionally reduced based on the principal component analysis, and the indicators are left one category with certain correlation. We then extract the principal component to replace the original variable, and examine each principal component for the success of a product.

We select from the existing indicators and make a new indicator X4(the number of words in each review sentence). When dealing with the "vine" and "verified purchase" indicators, we deem that the value of "Y" is 1 while the value of "N" is 0 to simplify the following process.

Table 1: Six indicators

X1	X2	X3	X4	X5	X6
star rating	helpful votes	total votes	number	vine	verified purchase

In order to investigate whether the data can be used for PCA, we first test the original data, and the test results are shown in the following labels. Kmo test values of the three types of products are 0.539, 0.518 and 0.567, which meet the feasibility standard of PCA, The corresponding sig value are all 0.000, indicating that the data collected in this paper is suitable for PCA.

Table 2: KMO and Bartlett test for hair dryer

KMO measure of sampling adequacy		.539
Bartlett test of sphericity	Approximate chi square	57232.004
	Freedom dgree	15
	Saliency	.000

Table 3: KMO and Bartlett test for microwave

KMO measure of sampling adequacy		.5689
Bartlett test of sphericity	Approximate chi square	9528.228
	Freedom dgree	15
	Saliency	.000

We carry out the PCA, and get the total variance of interpretation.

As can be seen from the above three tables, the eigenvalues of the first three principal components are greater than or close to 1, and the cumulative variance contribution rate are 77.3%, 81.9% and 73.7%, which can show that the first three principal components can reflect

Table 4: KMO and Bartlett test for pacifier

KMO measure of sampling adequacy		.517
Bartlett test of sphericity	Approximate chi square	70633.255
	Freedom dgree	15
	Saliency	.000

Whole analysis						
Component	Sum	Variance percentage of initial eigenvalue	Accumulative %	Sum	The Variance % of the sum of load squares and	Accumulative %
1	2.206	36.769	36.769	2.206	36.769	36.769
2	1.415	23.582	60.351	1.415	23.582	60.351
3	1.018	16.964	77.315	1.018	16.964	77.315
4	0.702	11.699	89.014			
5	0.655	10.908	99.922			
6	0.005	0.078	100			

(a) hair dryer

Whole analysis						
Component	Sum	Variance percentage of initial eigenvalue	Accumulative %	Sum	The Variance % of the sum of load squares and	Accumulative %
1	2.435	40.587	40.587	2.435	40.587	40.587
2	1.542	25.703	66.29	1.542	25.703	66.29
3	0.936	15.608	81.898			
4	0.653	10.885	92.783			
5	0.43	7.171	99.954			
6	0.003	0.046	100			

(b) microwave

Whole analysis						
Component	Sum	Variance percentage of initial eigenvalue	Accumulative %	Sum	The Variance % of the sum of load squares and	Accumulative %
1	2.148	35.792	35.792	2.148	35.792	35.792
2	1.282	21.362	57.154	1.282	21.362	57.154
3	0.996	16.592	73.746			
4	0.824	13.728	87.474			
5	0.737	12.29	99.764			
6	0.014	0.236	100			

(c) pacifier

Figure 2: The total variance

the original data roughly. Thus we think that the star rating, helpful votes and total votal are most related to a product's success.

What's more, in order to figure out the hidden relationship between each indicator, we did Pearson correlation analysis and find that at 0.01 level (double tail), the correlation between star rating and number, vine and verified purchase are significant.

3.2 The evaluation model based on entropy weight method

In order to eliminate the effects from the different value range, dimension and meaning of each index, the data was normalized firstly.

Here we select 10 products that have the most reviews in total and 2 indexes(star rating and review score) of each product. The observed data was represented by: $X_{ij}(i = 1,2,...,10; j = 1,2)$. The normalization method can adopt the following formula:

$$y_{ij} = \frac{x_{ij} - x^*_{jm}}{x^*_{Jm} - x^*_{jm}} \tag{1}$$

where $x^*_{Jm} = maxx_{ij}, x^*_{jm} = minx_{ij}$

After the above transformation, all the data are in the range from 0 to 1.

Table 5: Date after normalization

Product ID	Star rating	Review score
'B003V264WW'	0.829120858	0
'B0009XH6TG'	0.629871355	0.787384438
'B00132ZG3U'	1	0.474499448
'B00005O0MZ'	0.520712985	1
'B000R80ZTQ'	0.732081446	0.805424702
'B000A3I2X4'	0.830814891	0.252305889
'B001UE7D2I'	0.088810813	0.948828142
'B001QTW2FK'	0.624671976	0.704889728
'B0009XH6WI'	0	0.134025413
'B0009XH6V4'	0.959827151	0.404495664

Entropy weight method is used to express the discrete degree for the data. Therefore, it is selected to determine the objective weight of indicators. The steps are as follows:

a. Obtain sample data, which includes p indicators, m samples, and then we establish the data matrix.

$$R = (r_{ij})_{mp} \tag{2}$$

b. Calculate the proportion matrix of the ith sample under the j-th index

$$p_{ij} = \frac{r_{ij}}{\sum_{i-1}^{m}} r_{ij} \tag{3}$$

c. Calculate the entropy of the j-th index

$$e_j = -k \sum_{i-1}^{m} p_{ij} ln p_{ij} \tag{4}$$

d. Calculate the objective weight of the j-th index by entropy weight method

$$w_j'' = \frac{(1-e_j)}{\sum_{j-1}^{p}(1-e_j)} \tag{5}$$

We then get the weights of these two indexes, which are 0.4603 and 0.5397. And the final score are as follows(3):

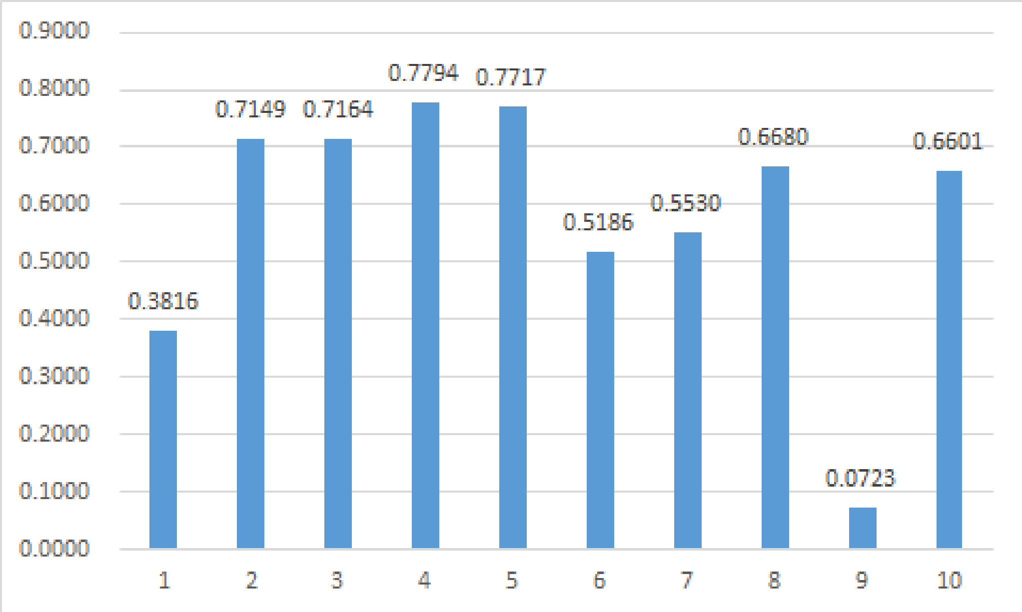


Figure 3: The final score of 10 products

4 Time series model

4.1 Time series model based on stars and reviews

In order to find the relationship between the star rating and the number of reviews, i.e. whether the customers who buy goods on Amazon website will be transferred along with others' star rating, for example, if they see that the product score is low and they have distrust of the product, they will give poor rating or give improvement suggestions to the business, but they may also see that the star rating of the product is high and they like the product To give sincere praise, let's take the model with the highest sales volume among microwave ovens, hair dryers and baby pacifiers as an example to analyze the relationship between the two variables and whether people have herd consumption psychology.

Before our analysis, we first carry out the data processing: first, select the product model with the highest sales volume, and arrange the product sales in chronological order. Take each month as the time unit, calculate the average star rating value of this month, count the evaluation quantity of this month, and get the data set we use for the next analysis.

By analyzing the correlation between the two variables, the Pearson correlation coefficient in the whole time period is obtained, and it is found that there is a certain correlation. The correlation coefficient between the star rating and the number of comments of microwave reaches 0.730, which has a great positive correlation. However, such analysis is too rough, because the research time is too long, there may be a significant correlation relation during a certain time period, so we do the following analysis.

Establish time prediction model. First of all, according to the existing data, the sequence diagrams of three products are made respectively, including the relationship between stars and time, the relationship between reviews and time, and the relationship between stars and reviews.

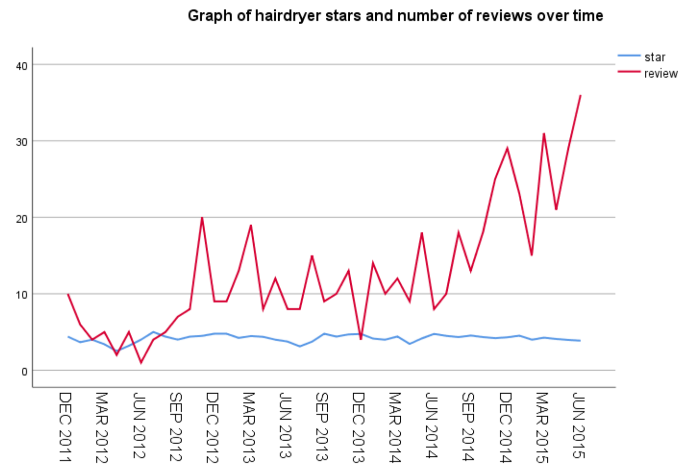


Figure 4: Sequence diagram of hair dryer

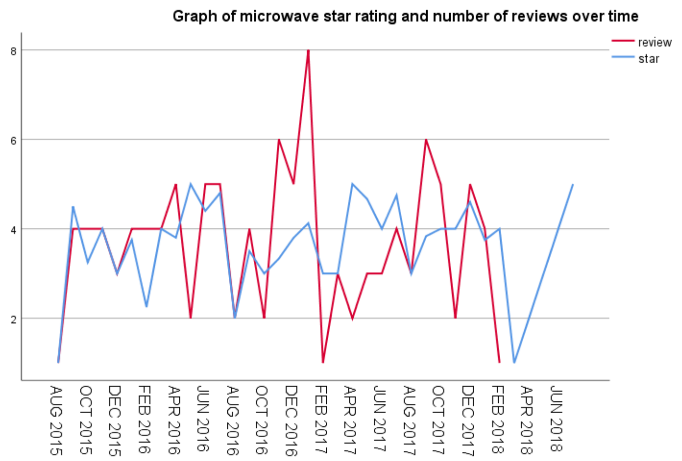


Figure 5: Sequence diagram of microwave

Through the sequence diagram figure(**??**), we can get very obvious results. Take the sequence diagram of star rating and comment quantity of microwave as an example, we can see that the peaks of most star rating change curves correspond to the peaks of comment quantity, and the troughs correspond to the troughs, showing a significant positive correlation, indicating that high star rating indicates that people recognize products and attract more consumers, However, in many products, there are also peaks of interstellar and troughs of comments, especially the sales of hair dryer products.

We can see that, with the decrease of stars, the amount of comments shows an obvious upward trend, which indicates that poor comments have a negative impact on the goods, leading more consumers to gradually find a series of defects such as product quality problems, and

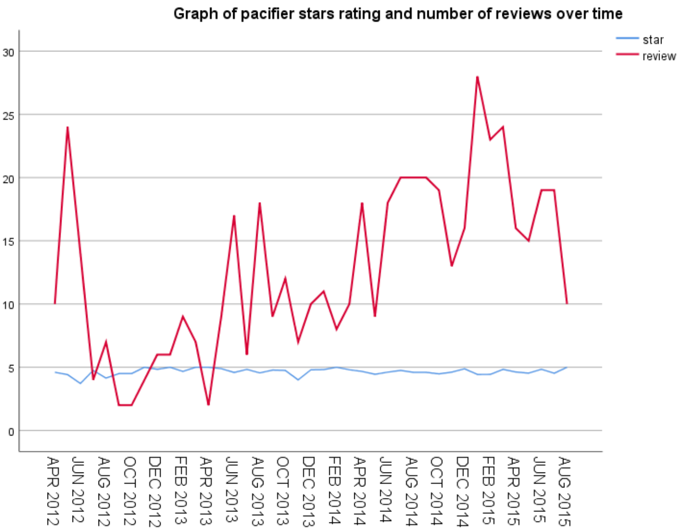


Figure 6: Sequence diagram of pacifier

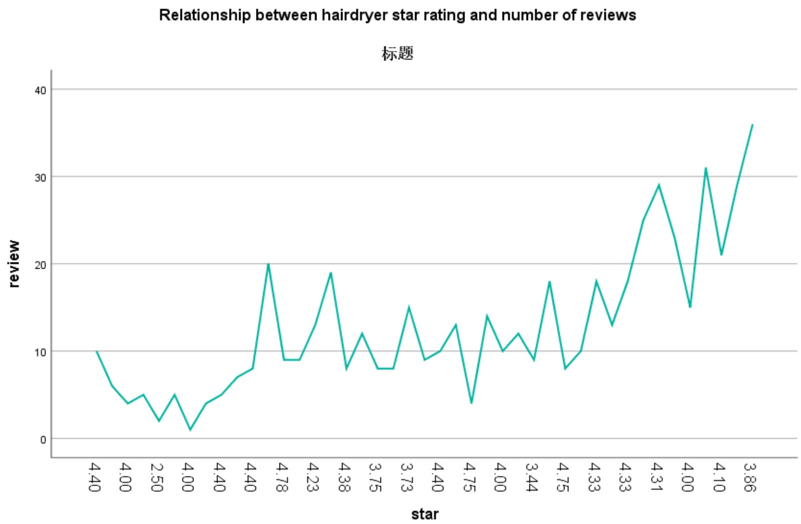


Figure 7: Star rating and the number of reviews of microwave

the users who buy the products then put forward more opinions, which makes the amount of comments increase.

Build time series analysis of ARIMA model. Create a traditional model with star rating as the independent variable and review as the dependent variable, conduct time series analysis, take month as the time unit, and predict the corresponding evaluation quantity according to the known star rating. The results are basically consistent with the actual situation and meet our expectations. The following table is the analysis of the fitting degree of each product model.

4.2 Time series model based on sliding window prediction

On the basis of the evaluation model based on entropy weight method mentioned above, we get the combinations of text-based measures and ratings-based measures, with the ignored influence of helpful votes. We have chosen one product for each category as a sample to show how we determine whether a product potentially successful or failing. The historical composite score of microwave is shown below:

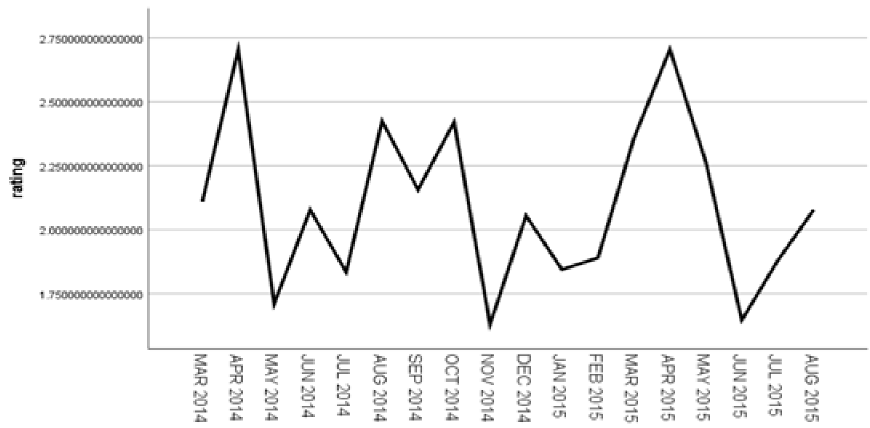


Figure 8: Historical composite score of microwave

We use the expert model for prediction, which will provide a model with minimum error based on the known sequence data. The most suitable model in the model description table is the simple seasonal model.

The stable R-square obtained by model fitting test is 0.908, which shows that the model is highly correlated. Both ACF and PACF are within their respective ranges, with no significant difference from 0. The test passed.

To solve data quality problems for time series analysis and decision-making, a prediction based sliding window prediction algorithm was proposed. The method first split given product's time series into subsequences so as to build a forecasting model to predict future values, and then the predicted value is also assumed to be known value, to continue to predict the future data.

First, we used nearly two years of data to predict the trend of products in the next four months. Then, the previous monthly data were used to predict where the product would go in the next three months. The predicted results are as follows:

It can be seen from the figure that the time sequence diagrams of real data and fitted data coincide well, which indicates that the simple seasonal model has a better fitting effect on the

Table 6: Test results

Statistical matching	AVG
stable R2	0.908
R2	0.813
RMSE	0.147
MAPE	4.888
MaxAPE	14.943
MAE	0.099
MaxAE	0.297
BIC	-3.512

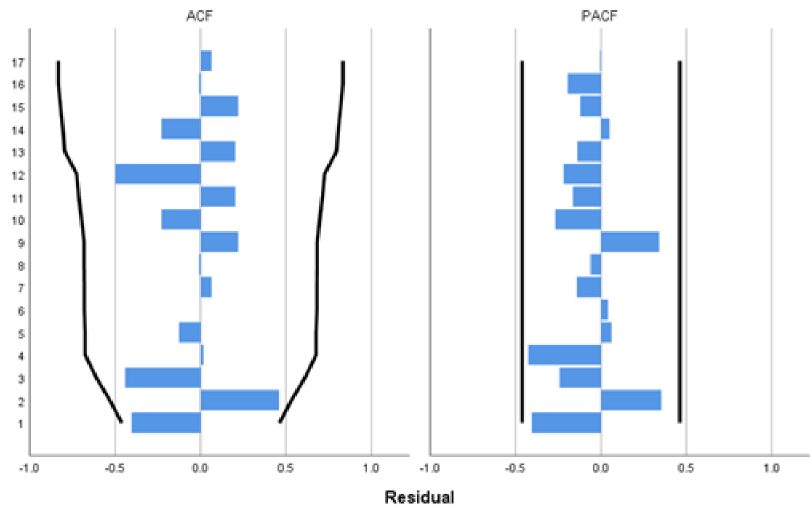


Figure 9: Residual ACF and residual PACF

original data. We do a seasonal breakdown of the data.

The seasonal factor in the first and the three quarters was positive, while the seasonal factor in the second and the fourth quarters was negative, which indicating that the average score of the product in the first and the third quarters was higher than that in the second and the fourth quarters. Seasonal decomposition is to separate the seasonal curve and observe the overall trend of the product. As you can see from the chart above, the product's rating has not been on the rise for a long time, so we don't consider it a success.

In the same way, we evaluate the hair dryer and the pacifier. The Hair-dryer is no more on the rise than the microwave, but the rising trend of pacifier is obvious. So we can say that this pacifier is successful.

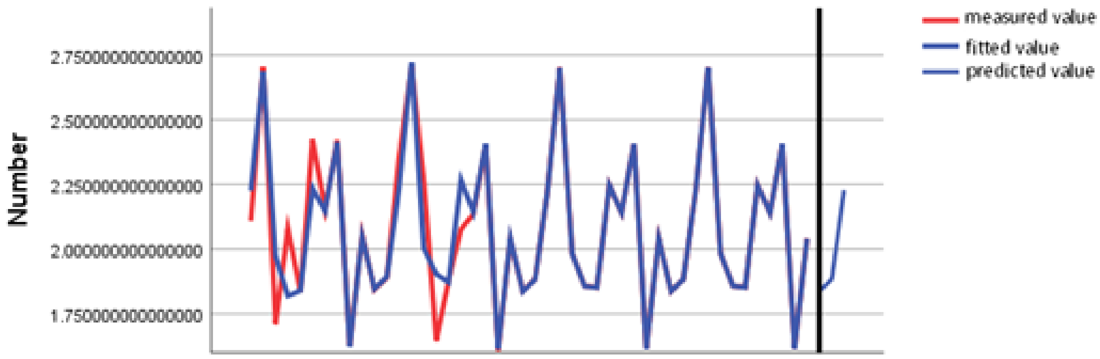


Figure 10: Predicted results

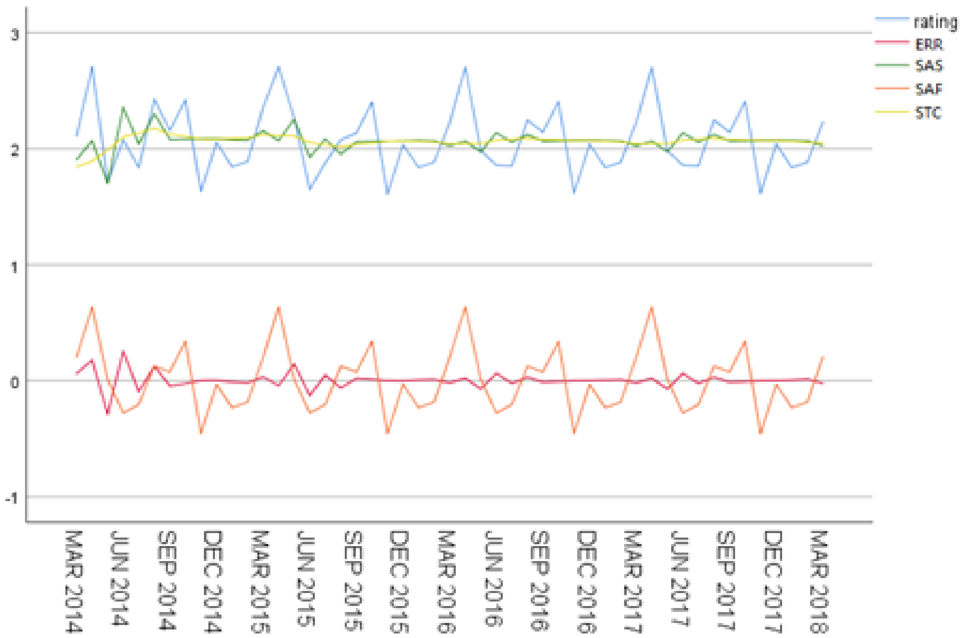
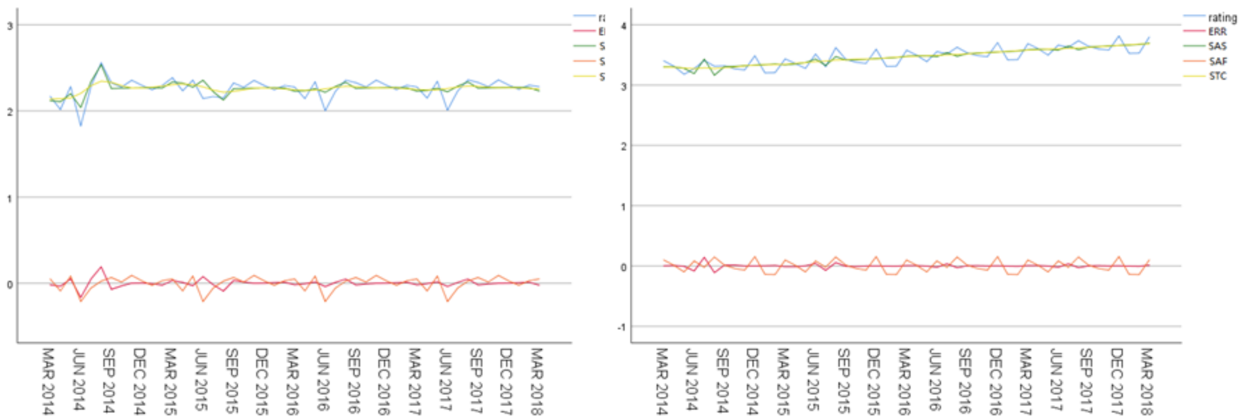


Figure 11: Seasonal decomposition curve of hair dryer

Table 7: Seasonal factor

Cycle	Seasonal factor
1	0.202886503616550
2	0.635392891679713
3	0.010326788519055
4	-0.278963673215924
5	-0.204705249341441
6	0.124851080845626
7	0.075438649012786
8	0.338933294328234
9	-0.456152812465383
10	-0.031461649503738
11	-0.232407556089949
12	-0.184138267385530



(a) microwave (b) pacifier

Figure 12: Seasonal decomposition curve

5 Models based om Natural Language Processing

5.1 Text Classification using SpaCy

We classify the positive and negative emotions of the evaluation through natural language processing, and process the evaluation information in data. We will explore texual data using amazing spacy library and build an emotional classification model. Below we take the data information of the hair dryer as an example to introduce the operation steps and the establishment of the model in detail.

1.**First, we read the data of star rating.** In this indicator, we determined that the rating 4 to 5 is a positive review, 1 to 2 is a negative review, and the data of rating 3 can be considered bad or not bad. The positive data is recorded as 1, the negative evaluation is recorded as 0, and the data is unified to make a histogram. Then we selected randomly 500 reviews in the sample to join the training set.

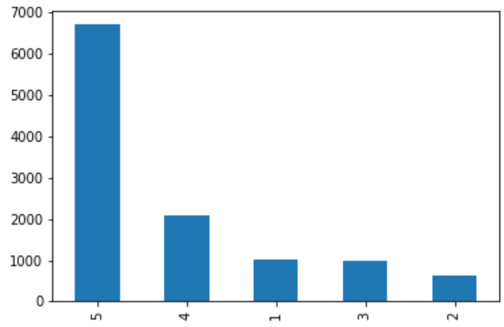


Figure 13: Score

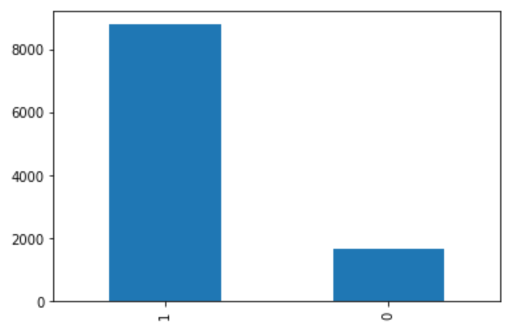


Figure 14: Score boolean

Figure 15: Classification of score

2.**Visualize Data include explacy that explains how parsing is done and displaCy that visualizes named entities.** We set up Text classification model–SpaCyText Categorizer. We will extract linguistic features like tokenization, part-of-speech tagging, dependency parsing, dependency parsing, named entities recognition, Sentence Boundary Detection for building language models later.

Frist for tokenization, first step in any NLP pipeline is tokenizing text i.e breaking down paragraphs into sentences and then sentences into words, punctuations and so on. We will load English language model to tokenize our English text. Every language is different and have different rules. Spacy offers 8 different language models. We visualize the sentence structure of the sentence and check if the previous work is in place. There is not much difference between parsed review and original one. But we will see ahead what has actually happened. We can see how parsing has been done visually through explacy.

Second for Part-of-speech tagging, after tokenization we can parse and tag variety of parts of speech to paragraph text. SpaCy uses statistical models in background to predict which tag will go for each word based on the context. Then analyze the composition of the sentence, you can decompose the specific part of speech of each word, restore the part of speech, and filter out punctuation, numbers and stop words. Lemmatization is the process of extracting uninflected/base form of the word. Lemma can be like For eg. Adjectives: best, better $\rightarrow$ good

Adverbs: worse, worst → badly Nouns: ducks, children → duck, child Verbs: standing,stood → stand

Third for Named Entity Recognition (NER), Named entity is real world object like Person, Organization etc. Spacy figures out entities automatically.

Fourth for Dependency parsing, Syntactic Parsing or Dependency Parsing is process of identifying sentences and assigning a syntactic structure to it. As in Subject combined with object makes a sentence. We need Sentense Boundry Detection to figure out where sentence starts and ends is very important part of NLP.

The last for Processing Noun chunks , Visualize using Scattertext and Word vectors and similarity.

3.**Set up training model and focus on scoring our hair dryer's customer reviews**

We will train a multi-label convolutional neural network text classifier on our procduct reviews. The document tensor is then summarized by concatenating max and mean pooling, and a multilayer perceptron is used to predict an output vector of length nr class, before a logistic activation is applied elementwise. The value of each output neuron is the probability that some class is present.

Finally, using the model, we tested the remaining evaluation texts. After the test, we found that Positive review is indeed close to 1 and Negative review is close to 0.

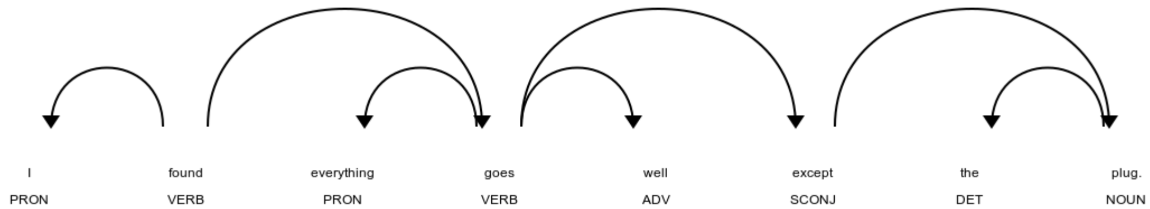


Figure 16: The structure of sentence

	text	lemma	pos	tag	dep	shape	is_alpha	is_stop	is_punctuation
0	I	-PRON-	PRON	PRP	nsubj	X	TRUE	TRUE	FALSE
1	found	(find)	VERB	VBD	ROOT	xxxx	TRUE	FALSE	FALSE
2	everything	(everything)	PRON	NN	nsubj	xxxx	TRUE	TRUE	FALSE
3	goes	(go)	VERB	VBZ	ccomp	xxxx	TRUE	FALSE	FALSE
4	well	(well)	ADV	RB	advmod	xxxx	TRUE	TRUE	FALSE
5	except	(except)	SCONJ	IN	prep	xxxx	TRUE	TRUE	FALSE
6	the	(the)	DET	DT	det	xxx	TRUE	TRUE	FALSE
7	plug	(plug)	NOUN	NN	pobj	xxxx	TRUE	FALSE	FALSE
8	.	(.)	PUNCT	.	punct	.	FALSE	FALSE	TRUE
9	Why	(why)	ADV	WRB	advmod	Xxx	TRUE	TRUE	FALSE
10	the	(the)	DET	DT	det	xxx	TRUE	TRUE	FALSE
11	left	(left)	NOUN	NN	nsubj	xxxx	TRUE	FALSE	FALSE
12	and	(and,)	CCONJ	CC	cc	xxx	TRUE	TRUE	FALSE
13	right	(right)	ADJ	JJ	conj	xxxx	TRUE	FALSE	FALSE
14	is	(be)	AUX	VBZ	ROOT	xx	TRUE	TRUE	FALSE
15	opposite	(opposite)	ADJ	JJ	acomp	xxxx	TRUE	FALSE	FALSE

Figure 17: The analysis of words

text	root	root.text	root.dep_	root.head.text	
0	I	I	I	nsubj	found
1	everything	(everything)	(everything)	nsubj	goes
2	the plug	(plug)	(plug)	pobj	except
3	the left	(left)	(left)	nsubj	is
4	I	(I)	(I)	nsubj	have
5	the plug	(plug)	(plug)	dobj	put
6	charging	(charging)	(charging)	pcomp	for
7	Another flaw	(flaw)	(flaw)	nsubj	is
8	the big noise	(noise)	(noise)	attr	is

Figure 18: The match of words

5.2 The value of the indexes

From the Text Classification we can get the score of each score, to calculate the average review scores and star rating, we think that the helpful votes are closely related to the review scores. We determine the review score based on the propotion of helpful votes.

To validate our opinion, we make the ANOVA analysis, in which the independent variable is review score and the dependent variable is helpful notes.

ANOVA					
	Square sum	Free degree	Mean square	F	Saliency
Regression	1050.603	1	1050.603	2.81	0.94
Residual	1374083.165	3675	373.9		
Sum	1375133.769	3676			

Figure 19: The outcome of ANOVA

Then we get the star rating(calculated by average) and the review scores(calculated by the helpful votes'weigh) of each product.

5.3 Four models based on KNN

This python program is based on the given review data by applying K Nearest Neighbors (KNN) algorithm. To build generalized prediction model first step should be necessary cleaning of data as a part of data preprocessing.

We will perform following data preprocessing:

- Remove Stop-words
- Remove any punctuations or limited set of special characters like ',' or '.' etc.

- Snowball Stem the word
- Convert the word to lowercase

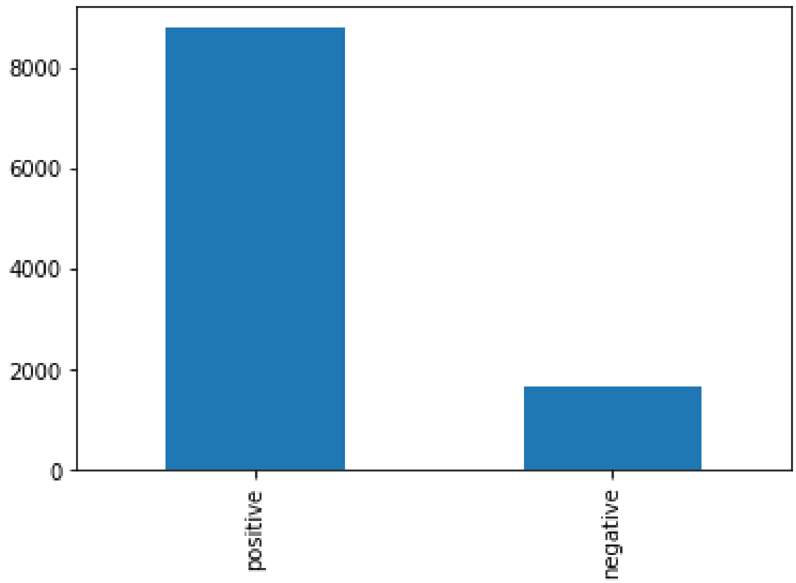
Once we the data is cleaned to be processed we'll use below Feature generation techniques to convert text to numeric vector.

- Bag Of Words (BoW)
- Term Frequency - inverse document frequency (tf-idf)
- Word2Vec
- tf-idf weighted Word2Vec

Given a review determine whether a review is positive or negative, by appling KNN algorithm and deciding the best Feature generation technique for given problem.

5.3.1 Appling KNN with 4 models based on hair dryer

In the data Preprocessing, we segregate data as positive and negative and sort data for time based splitting for model train and test dataset.

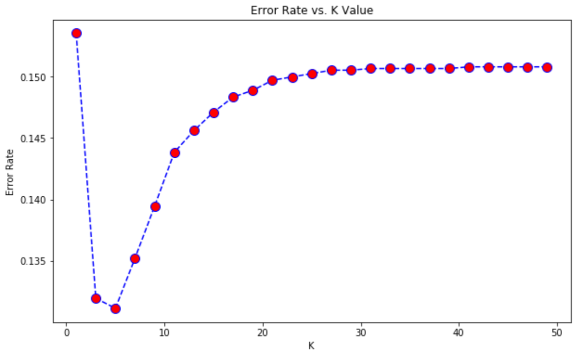


We first generate Bag of Words Vector matrix for Reviews, then split Data into Train and Test based on the timestamp of review, find Optimal K by 10 fold Cross validation.

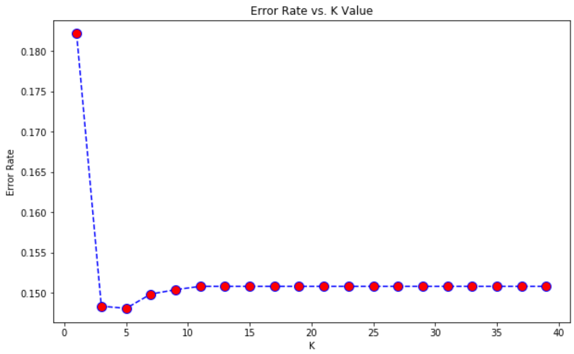
Then we use the optimal k, to get a confusion matrix

The result of feature generation techniques and machine learning algorithms vary by application. But by comparing the accuracy of all 4 developed models, KNN model with Avg. tf-idf weighted Word2vec feature generation technique gives accuracy more than 88% which is the best to predict the polarity of reviews among all models.

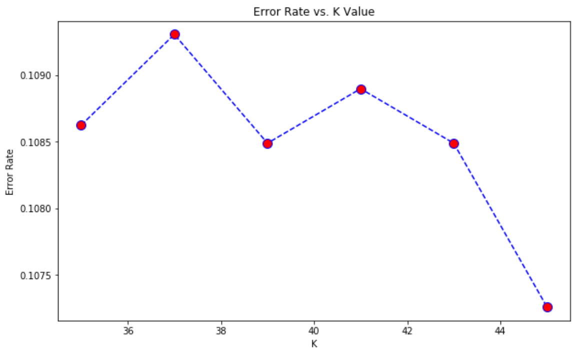
Samely, we can find out the most suitable model for each product: model for microwave is tf-idf weighted Word2vec, the model for pacifier is Bag of words. And the we can get the predicted number of positive and negative reviews from the corresponding confusion matrix.



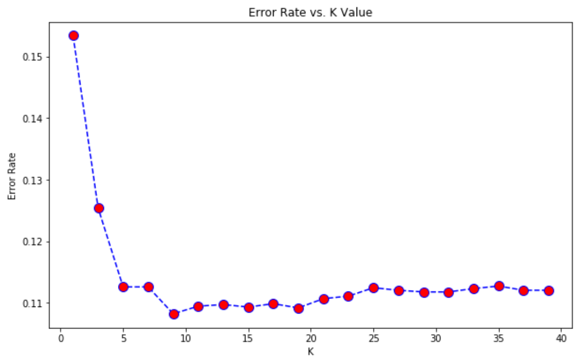
(a) BoW



(b) tf-idf

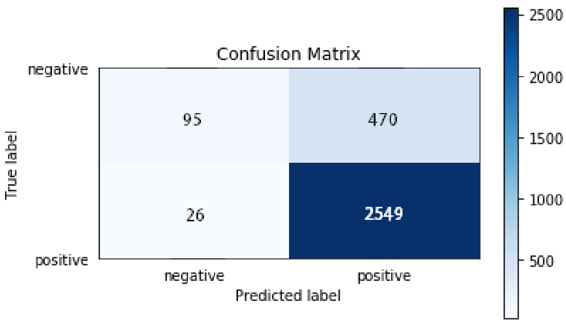


(c) Word2Vec

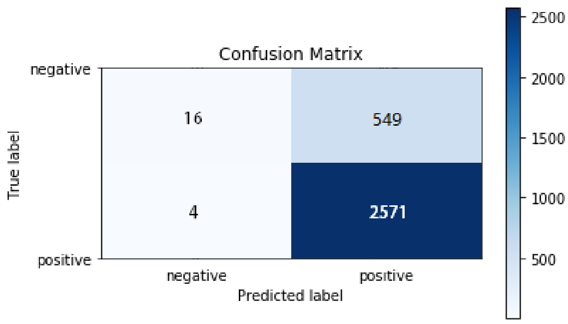


(d) tf-idf weighted Word2Vec

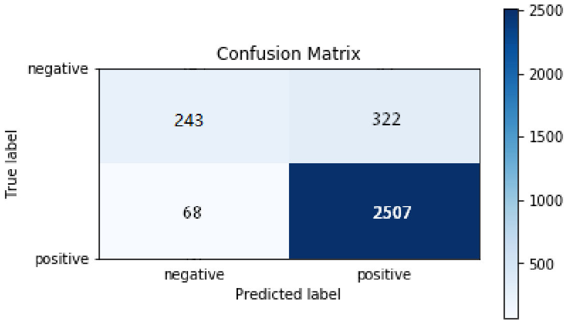
Figure 20: K value



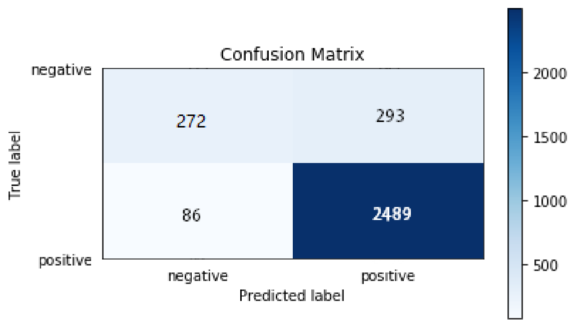
(a) BoW



(b) tf-idf



(c) Word2Vec



(d) tf-idf weighted Word2Vec

Figure 21: Confusion matrix

Table 8: Accuracy for KNN model with Bag of words is 0.842

	precision	recall	f1-score	support
negative	0.79	0.17	0.28	565
positive	0.84	0.99	0.91	2575
accuracy			0.84	3140
macro avg	0.81	0.58	0.59	3140
weighted avg	0.83	0.84	0.80	3140

Table 9: Accuracy for KNN model with tf-idf is 0.824

	precision	recall	f1-score	support
negative	0.80	0.03	0.05	565
positive	0.82	1.00	0.90	2575
accuracy			0.82	3140
macro avg	0.81	0.51	0.48	3140
weighted avg	0.82	0.82	0.75	3140

5.3.2 The trend of the three products' reputation

To simplify our model, we focus exclusively on the propotion of the positive reviews, and here we make a comparison between the training set and prediction set.

From the three figures (22 23 24), we can see that all the reputation of these three products shows an upward trend, with the proportion of positive comments rising by 4%, 7%, 11% respectively.

Table 10: Accuracy for KNN model with Word2Vec is 0.876

	precision	recall	f1-score	support
negative	0.78	0.43	0.55	565
positive	0.89	0.97	0.93	2575
accuracy			0.88	3140
macro avg	0.83	0.70	0.74	3140
weighted avg	0.87	0.88	0.86	3140

Table 11: Accuracy for KNN model with tf-idf weighted Word2vec 0.88

	precision	recall	f1-score	support
negative	0.76	0.48	0.59	565
positive	0.89	0.97	0.93	2575
accuracy			0.88	3140
macro avg	0.83	0.72	0.76	3140
weighted avg	0.87	0.88	0.87	3140

Figure 22: Hair dryer

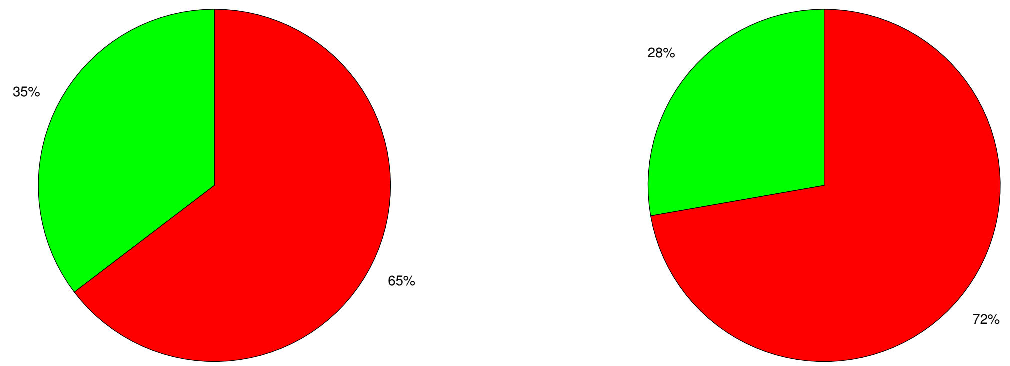


Figure 23: Microwave

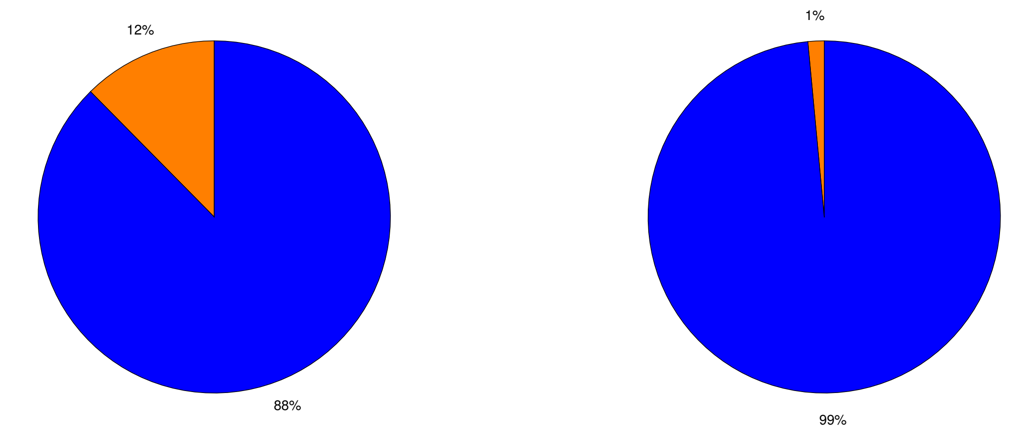


Figure 24: Pacifier

5.4 Assosiation between quality descriptors and rating levels

In this question we use NLTK(Natural Language Toolkit) to deal with the text-based reviews. Steps are as follows:

- Load and prepare the Dataset: Load columns (Star ating and Review body); Here we define reviews with stars more than 4 as positive reviews and reviews with stars less than 4 as negative reviews.
- Data exploration: Display general informations about the data.
- Filter adjectives with NLTK:

  ① Create a list for each review and split all sentences in a review; Split all words from each sentence and add tags all words.

  Review : Sentence 1: [(word 1, tag 1), ... , (word n, tag n)]; Sentence n: [(word 1, tag 1), ... , (word n, tag n)]

  ② Add all adjectives to a list (without tag, only word)

  Review: ['Adjective 1', Adjective 2, ... , Adjective n]

  ③ Transform the list of adjectives of each review to one string each review, which is needed for the model later on.

  Review: [Adjective1 Adjective2 Adjective3 Adjective 4]
- Multinomial Naive Bayes: Fit model with train data and calculate R2-score with test data.
- Confusion-Matrix: Because the already classified data we trained our model with is not very balanced (much more positives than negatives), we don't know whether the R2 score is reliable. The Model could be just labeling almost everything as positive, even if it should be negative. With the Confusion matrix, we can evaluate the accuracy of the classification more clearer and see exactly where our multinomial naive Bayes has its errors.
- Classify adjectives as positive or negative and get the most used adjectives.

In order to better show which descriptors of the product are most closely related to the rating levels, we removed adverbs such as "little" and "much" and nouns of the product itself, and then arranged the remaining words in ascending order of absolute value from small to large, and got the following table.

Adjectives with higher coefficients (good, great, easy, nice, perfect, powerful) are correlated to positive reviews and adjectives with lower coefficients (bad, low, expensive, hard) reduce the likelihood of an adjective having a positive meaning/being in a positive review.

Table 12: The specific descriptions and coefficients in microwave

Descriptors	Coefficient	Descriptors	Coefficient	Descriptors	Coefficient
great	-2.31	new	-3.22	happy	-3.84
good	-2.61	nice	-3.41	turntable	-3.97
small	-2.75	large	-3.71	powerful	-4.15
easy	-2.90	big	-3.78	fit	-4.20
old	-3.00	perfect	-3.84	clean	-4.30

Table 13: The specific descriptions and coefficients in hair dryer

Descriptors	Coefficient	Descriptors	Coefficient	Descriptors	Coefficient
great	-2.26	powerful	-3.45	new	-3.96
good	-2.60	nice	-3.67	low	-3.99
old	-3.35	small	-3.78	heavy	-3.99
easy	-3.35	happy	-3.85	quiet	-4.02
hot	-3.39	light	-3.90	retractable	-4.30

Table 14: The specific descriptions and coefficients in pacifier

Descriptors	Coefficient	Descriptors	Coefficient	Descriptors	Coefficient
great	-2.13	nice	-3.65	small	-3.86
easy	-2.62	new	-3.72	cute	-3.93
old	-2.73	perfect	-3.73	big	-3.97
good	-2.86	different	-3.81	clean	-4.09
soft	-3.65	happy	-3.84	hard	-4.15

6 Strengths and Weaknesses

6.1 Strengths

- When we articulate our own metrics for assessment, we try our best to include all the important elements to make the rating more accurate. Time factor, review, star-rating, helpful votes are all discussed in the model.
- We states a distinct quantification system which is expected to live up to the common sense.
- Our model can be practically implemented in the ?eld. By varying the initial parameters based on the speci?c case, optimal results can be determined.
- We come up with various criteria to compare different situations. Hence an overall comparison can be made based on these criteria.
- Our models are fairly robust to the changes in parameters based on sensitivity analysis. It means a slight change in parameters will not cause a signi?cant change in the result.

6.2 Weaknesses

- When using time series analysis, we only choose one of the three types of products, which is not extensive.
- In principal component analysis, there is some sloppiness in processing 0 and 1 for vine and verified purchase.
- Too few indexes in principal component analysis.
- Only one factor of high praise is considered in the prediction of product reputation change.

References

[1] Shasha Li Study on mining method optimization based on Improved AHP and entropy weight method,2020

[2] Jianghua Sun Research on the influence of traditional newspaper microblog based on principal component analysis and cluster analysis,2015

[3] YiChen Dong,Siyi Li,Xueqin Gong. Time Series Analysis: An application of ARIMA model in stock price forecasting[P]. Proceedings of the 2017 International Conference on Innovations in Economic Management and Social Science (IEMSS 2017),2017.

[4] YiChen Dong,Siyi Li,Xueqin Gong. Time Series Analysis: An application of ARIMA model in stock price forecasting[P]. Proceedings of the 2017 International Conference on Innovations in Economic Management and Social Science (IEMSS 2017),2017.

[5] Yuxing Qian,Huayang Zhou,Hao Li,Meiling Ren,Wenxuan Gui,Liqin Zhou. Analysis of Users' Health Knowledge Requirement and Health Perception in Senior Online Community Based on Web Text Mining[P]. Proceedings of the 2nd International Conference on Social Science, Public Health and Education (SSPHE 2018),2019.

[6] Weimin Ouyang,Qinhua Huang. Mining Weighted Association Rules in Data Streams with a Sliding Window[P]. Proceedings of the 2015 International Conference on Computer Science and Intelligent Communication,2015.

[7] Johannes K. Chiang,Chun-Cheng Chen. Sentimental Analysis on Big Data ? On case of Financial Document Text Mining to Predict Sub-Index Trend[P]. Proceedings of the 2015 5th International Conference on Computer Sciences and Automation Engineering,2016.

[8] Huali Cai,Fang Wu,Qi Duan,Jian Kang,Yawei Jiang. Simple Mining Method of Rich Text Data from User Feedback and Its Application[P]. Proceedings of the 2016 International Conference on Engineering Science and Management,2016.

[9] Huilong Fan,Yongbin Qin. Research on Text Classification Based on Improved TF-IDF Algorithm[P]. Proceedings of the 2018 International Conference on Network, Communication, Computer Engineering (NCCE 2018),2018.

[10] Liwen Liang. A Method of Antarctic Temperature Forecasting Based on Time Series Model[P]. Proceedings of the 2017 5th International Conference on Frontiers of Manufacturing Science and Measuring Technology (FMSMT 2017),2017.

[11] Wengang Chen,WenWei Chen,Bing Fan,Runze Wu,Yanru Wang,Junli Fan. A Data Preprocessing Optimization Method based on Sliding Time Window in Communication Alarm Events[P]. Proceedings of the 2016 3rd International Conference on Materials Engineering, Manufacturing Technology and Control,2016.

[12] Weisi Dai. Improvement and Implementation of Feature Weighting Algorithm TF-IDF in Text Classification[P]. Proceedings of the 2018 International Conference on Network, Communication, Computer Engineering (NCCE 2018),2018.

[13] Lamia Al-Horaibi,Muhammad Badruddin Khan. Sentiment analysis of Arabic tweets using text mining techniques[P]. International Workshop on Pattern Recognition,2016.

[14] Zhifang Dong,Jiasong Wu,Jiyong Gui. A novel sliding window algorithm for 2D discrete Fourier transform[P]. International Symposium on Multispectral Image Processing and Pattern Recognition,2015.

[15] Yadav, C.S.,Sharan, A.,Joshi, M.L.. Semantic graph based approach for text mining[P]. Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014 International Conference on,2014.

[16] Isah, H.,Trundle, P.,Neagu, D.. Social media analysis for product safety using text mining and sentiment analysis[P]. Computational Intelligence (UKCI), 2014 14th UK Workshop on,2014.

[17] Jian-hong YU,Xiao-juan LE. Sales Forecast for Amazon Sales Based on Different Statistics Methodologies[P]. Joint International Conference on Economics and Management Engineering (ICEME 2016) Economics and Business Management (EBM 2016),2016.

[18] Somprasertsri, G., Lalitrojwong, P.. Extracting product features and opinions from product reviews using dependency analysis[P]. Fuzzy Systems and Knowledge Discovery (FSKD), 2010 Seventh International Conference on,2010.

[19] Haiping Zhang, Zhengang Yu, Ming Xu, Yueling Shi. Feature-level sentiment analysis for Chinese product reviews[P]. Computer Research and Development (ICCRD), 2011 3rd International Conference on,2011.

[20] Bei LIU. Oceanographic Big Data Text Categorization Algorithm Based on Improved Mutual Information[C]. Science and Engineering Research Center.Proceedings of 2017 2nd International Conference on Artificial Intelligence: Techniques and Applications (AITA 2017).Science and Engineering Research Center:Science and Engineering Research Center,2017:96-100.

[21] Jie Chen. Optimized TF-IDF Algorithm with the Adaptive Weight of Position of Word[C]. Science and Engineering Research Center.Proceedings of 2016 2nd International Conference on Artificial Intelligence and Industrial Engineering (AIIE2016).Science and Engineering Research Center:Science and Engineering Research Center,2016:125-128.

[22] Sukanya, M.,Biruntha, S.. Techniques on text mining[P]. Advanced Communication Control and Computing Technologies (ICACCCT), 2012 IEEE International Conference on,2012.

[23] Yunfei Yi,Lijun Liu,Cheng Hua Li,Wei Song,Shuai Liu. Machine Learning Algorithms with Co-occurrence Based Term Association for Text Mining[P]. Computational Intelligence and Communication Networks (CICN), 2012 Fourth International Conference on,2012.

[24] Reategui, E.,Klemann, M.,Finco, M.D.. Using a Text Mining Tool to Support Text Summarization[P]. Advanced Learning Technologies (ICALT), 2012 IEEE 12th International Conference on,2012.

[25] Wu Maowen,Zhang Cai Dong,Lan Weiyao,Wu Qing Qiang. Text topic mining based on LDA and co-occurrence theory[P]. Computer Science & Education (ICCSE), 2012 7th International Conference on,2012.

[26] Albathan, M.,Yuefeng Li,Algarni, A.. Using Patterns Co-occurrence Matrix for Cleaning Closed Sequential Patterns for Text Mining[P]. ,2012.

[27] Trappey, C.,Hsin-Ying Wu,Kuan-Liang Liu,Feng-Teng Lin. Knowledge Discovery of Service Satisfaction Based on Text Analysis of Critical Incident Dialogues and Clustering Methods[P]. ,2013.

[28] Kamatchi, K.,Siva Balan, A.. Multiphase text mining predictor for market analysis[P]. ,2013.

[29] Guelpeli, M.V.C., Garcia, A.C.B., Branco, A.H.. The process of summarization in the preprocessing stage in order to improve measurement of texts when clustering[P]. Internet Technology and Secured Transactions (ICITST), 2011 International Conference for,2011.

[30] Qingqiang Wu, Caidong Zhang, Xiang Deng, Changlong Jiang. LDA-based model for topic evolution mining on text[P]. Computer Science & Education (ICCSE), 2011 6th International Conference on,2011.

[31] Chaudhuri, C., Chaudhuri, A.. Detection of Verbatim or Partial Duplication from Multiple Source Documents Using Data Mining Techniques and Case-Based Reasoning Methodologies[P]. Emerging Applications of Information Technology (EAIT), 2011 Second International Conference on,2011.

[32] Bhujade, V., Janwe, N.J.. Knowledge Discovery in Text Mining Technique Using Association Rules Extraction[P]. Computational Intelligence and Communication Networks (CICN), 2011 International Conference on,2011.

[33] Simanovsky, A., Ulanov, A.. Mining Text Patterns for Synonyms Extraction[P]. Database and Expert Systems Applications (DEXA), 2011 22nd International Workshop on,2011.

[34] Zelong Liu, Maozhen Li, Yang Liu, Ponraj, M.. Performance evaluation of Latent Dirichlet Allocation in text mining[P]. Fuzzy Systems and Knowledge Discovery (FSKD), 2011 Eighth International Conference on,2011.

[35] Hirokawa, S., Baba, T., Nakatoh, T.. Text mining of bankruptcy information using formal concept analysis[P]. Awareness Science and Technology (iCAST), 2011 3rd International Conference on,2011.

[36] Saeed, Z., Sadaf, A., Muhammad, S.. Activity-based correlation of personal documents and their visualization using association rule mining[P]. Emerging Technologies (ICET), 2011 7th International Conference on,2011.

[37] Susheng Wang, Yan Liu, Zhichao Li, Yun Hua. The Relationship between Media Information and Stock Returns Based on Text Semantic Mining Algorithms[P]. Information Management, Innovation Management and Industrial Engineering (ICIII), 2011 International Conference on,2011.

[38] Joshi, S., Nigam, B.. Categorizing the Document Using Multi Class Classification in Data Mining[P]. Computational Intelligence and Communication Networks (CICN), 2011 International Conference on,2011.

[39] Yanxu Zhu, Gang Yin, Huaimin Wang, Dianxi Shi, Xiang Rao, Lin Yuan. Efficient Approach for Repeated Patterns Mining Based on Indent Shape of HTML Documents[P]. Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), 2011 International Conference on,2011.

[40] Goujon, B.. Text Mining for Opinion Target Detection[P]. Intelligence and Security Informatics Conference (EISIC), 2011 European,2011.

[41] Gharehchopogh, F.S., Khalifelu, Z.A.. Analysis and evaluation of unstructured data: text mining versus natural language processing[P]. Application of Information and Communication Technologies (AICT), 2011 5th International Conference on,2011.

[42] Nithya, K., Kalaivaani, P.C.D., Thangarajan, R.. An enhanced data mining model for text classification[P]. Computing, Communication and Applications (ICCCA), 2012 International Conference on,2012.

[43] Sun Yueheng, Liu Xing, Cui Xiaoyuan. The Mining of Term Semantic Relationships and its Application in Text Classification[P]. Intelligent Computation Technology and Automation (ICICTA), 2012 Fifth International Conference on,2012.

[44] Yan Chen, Yang Yang, Huisan Zhang, Haiping Zhu, Feng Tian. A topic detection method based on Semantic Dependency Distance and PLSA[P]. Computer Supported Cooperative Work in Design (CSCWD), 2012 IEEE 16th International Conference on,2012.

[45] Puri, S., Kaushik, S.. An enhanced fuzzy similarity based concept mining model for text classification using feature clustering[P]. Engineering and Systems (SCES), 2012 Students Conference on,2012.

[46] Sam, K.M., Chatwin, C.R.. Ontology-based text-mining model for social network analysis[P]. Management of Innovation and Technology (ICMIT), 2012 IEEE International Conference on,2012.

[47] Siering, M.. "Boom" or "Ruin"–Does It Make a Difference? Using Text Mining and Sentiment Analysis to Support Intraday Investment Decisions[P]. System Science (HICSS), 2012 45th Hawaii International Conference on,2012.

[48] Dongbin Hu,Lixia Luo,Lihua Xu. The study on Web product reviews mining based on an improved text categorization algorithm[P]. icebi-10,2010.

[49] https://www.kaggle.com/iwasdata/nltktoclassifyadjectivesaspositivenegative#Usingnltk-toclassifiyadjectivesinamazonfoodreviewsaspositiveornegative

[50] https://www.kaggle.com/poonaml/text-classification-using-spacy

[51] https://www.kaggle.com/jinalsalvi/amazonfoodreviewpredictionwith-knn/comments#385683

Appendices

Appendix A: Programmes Codes

Here are simulation programmes we used in our model as follow.

Input matlab source:

```
load a.txt;
 x=a;
[n,m]=size(x);%Standardization
xmin=min(x);
xmax=max(x);
dis=xmax-xmin;
for i=1:m;
     x(:,i)=(x(:,i)-xmin(i))/dis(i);
end;
p=[];
E=[];
for k=1:m
      s=sum(x(:,k));
      p=x(:,k)/s;
        s=0.0;
for i=1:n
       if(p(i)>0) s=s+p(i)*log(p(i)); end
      end
  E(k)=-s/log(n); %Get entropy
  end;
  s=sum(E);
  W1=(1-E)/(m-s); %Get weigh
```

Input Python source:

```
import numpy as np
import pandas as pd
import nltk
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import confusion_matrix
import re

df = pd.read_csv('microwave.tsv', sep='\t', usecols = ["star_rating", "review_body"])
df["star_rating"] = np.where(df["star_rating"] >= 4, "positive", "negative")

df.info()
print(df.head())

sns.countplot(data = df, x= df["star_rating"]).set_title("Score distribution", fontweigh
# plt.show()

texts = df["review_body"]
texts.apply(lambda x: re.sub('<br />','.',x))
texts.apply(lambda x: re.sub('&+#[0-9]+;','',x))

import nltk

from nltk.corpus import stopwords
texts_transformed = []
```

```
for review in texts:
    sentences = nltk.sent_tokenize(str(review))
    adjectives = []

    for sentence in sentences:
        words = nltk.word_tokenize(sentence)
        words = [word for word in words if word not in stopwords.words('english')]
        words_tagged = nltk.pos_tag(words)

        adj_add = [adjectives.append(word_tagged[0]) for word_tagged in words_tagged if

    texts_transformed.append(" ".join(adjectives))

X = texts_transformed
y = df["star_rating"]

X_train, X_test, y_train, y_test = train_test_split(X, y, random_state = 0)

cv = CountVectorizer(max_features = 50)
cv.fit(X_train)

X_train = cv.transform(X_train)
X_test = cv.transform(X_test)

arr = X_train.toarray()

print(arr.shape)

model = MultinomialNB()
model.fit(X_train, y_train)

print(model.score(X_test, y_test))

y_test_pred = model.predict(X_test)
print(confusion_matrix(y_test, y_test_pred))

def classifier(adjective):
    return model.predict(cv.transform([adjective]))
print(classifier('great'))
print(classifier('bad'))

adj = list(zip(model.coef_[0], cv.get_feature_names()))
adj = sorted(adj, reverse = True)
for a in adj:
    print(a)
```