



# International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 6, Issue 5, May 2018

## Enhancing Cosmetic Products Quality Using MRF and K-Means

Prachi Kunjir<sup>1</sup>, Madhura Argade<sup>2</sup>, Durgeshwari Rane<sup>3</sup>

U.G. Student, Department of Computer Engineering, PES Modern Engineering College, Shivajinagar, Pune, India<sup>1</sup>

U.G. Student, Department of Computer Engineering, PES Modern Engineering College, Shivajinagar, Pune, India<sup>2</sup>

U.G. Student, Department of Computer Engineering, PES Modern Engineering College, Shivajinagar, Pune, India<sup>3</sup>

**ABSTRACT:** Opinion mining on thousands of cosmetic products reviews of a company in an unsupervised manner is a task to survey the feedbacks of customers on their products and services. This is extremely helpful for owners to improve their business. In our project, we are using an opinion mining technique on reviews of cosmetic products using various data processing methods like text pre-processing and text transformation and subsequently applying K-Means Clustering and MRF feature selection algorithms. MRF feature selection is used for selecting relevant features from large number of features extracted. Then K-Means Clustering is employed for clustering into positive and negative reviews. Generated results from this, would help owner to recognize whether to specific product needs to be improved for increasing trade volume and revenue growth. Also system would generate a statistic analysis based on previous sales after the modification is applied. Here MRF can significantly reduce the number of features in the data set, so, CT is significantly decreased. Also when compared with Fuzzy-C-Means and K-Medoids. K-Means Clustering can achieve best clustering performance.

**KEYWORDS:** Data Mining, Opinion Mining, MRF feature selection, K-Means Clustering, Revenue Growth

### I. INTRODUCTION

Mostly recent systems which are under sentiment analysis have been developed for customers' use. Sentiment analysis is basically analysis of emotions or feelings or even an attitude expressed by a person towards any object. Opinion mining is also an alternative domain to sentiment analysis and both comes under data mining. Opinion mining is also used for achieving related outcomes except the opinions describe either positive or negative certainty. Opinion Mining is actually a mining of opinions expressed by numerous users to categorize an opinion according to defined aspects. These both techniques are mainly used for taking customer sentiments or opinions and processing those for analysing businesses in vast majority. Applications of these domains can be online shopping, healthcare, marketing, social media, consumer insights, corporate network etc. But also most of the applications that are implemented for customers and not for administrations, so we thought of developing something that would be helpful for business owners. Also we had interest in opinion mining thus finalized to implement a system according to business owner perspective in this domain only. In this we take user-generated online product reviews. We are gathering those reviews on cosmetic company's web platform. On those reviews we will be performing data pre-processing. From this we will be extracting the features for cosmetic products by using MRF feature selection method. According to these features K-means Clustering segregates positive and negative reviews. Positive reviews will be taken as compliments and negative reviews for improving the area of weakness.

### II. RELATED WORK

In 'Random Walk on the Red Carpet and Rating Movies', which is IEEE paper we studied Page Rank algorithm, used for movie ranking. Then in 'Opinion Mining for Thai Restaurant Reviews' paper we studied K-means clustering and Markov random field algorithm. Here various reviews are taken from customers to analyse restaurant's sales growth. In 'Grafting-Light' paper we studied MRF algorithm. In MRF there are many methods amongst which we chose chi-square

# International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 6, Issue 5, May 2018

method because it can find out the independence of two criteria of classification of qualitative variables and also relationships between categorical variables (contingency tables).

In 'Length of the Speech to Measure Opinion' paper we saw that psycholinguistic feature of the free individual expression that can be used to compute the opinion of texts. In 'Image Segmentation and Object Recognition' paper we studied MRF for both these methods. And in 'Mining Online Reviews for Predicting Sales Performance' we studied sentiment analysis, review mining and different prediction algorithms.

### III. PROPOSED SYSTEM

So, we are basically building a system for business owners in the cosmetics field, to increase their revenue growth and gain profit by modifying product's quality and enhancing product's manufacturing process or any other thing according to suggestions and compliments given in the reviews collected from the customers who have actually bought the product. Hence, customers' suggestions would always be appreciated, and thus company will yield more advantage. This system will work as follows:

Figure 1: Architecture Diagram

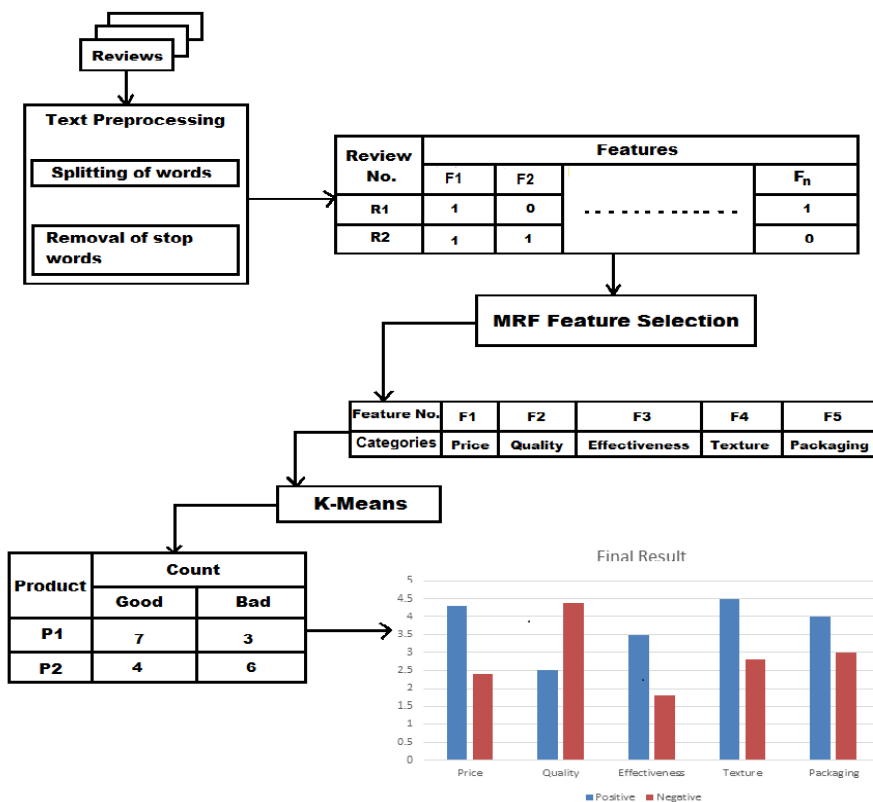


Figure 1: Architecture Diagram



# International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 6, Issue 5, May 2018

## A. Brief Methodology:

1. We are developing a web application where only authorised users can give their opinions on cosmetic products they have used.
2. By collecting those reviews we are going to perform data pre-processing which includes classification of the defined features under specified categories.
3. MRF is applied to extract those defined features from the set of features obtained from collected reviews.
4. K-means is used to segregate positive and negative reviews.
5. This overall process gives the result which is used for improving the product quality, helpful to generate economical statistical analysis.

## B. Description of the Proposed Algorithm:

1. MRF feature selection algorithm is a method applied to extract the features using Markov Random Field Optimization techniques. For determining the number of chosen features global threshold values are taken. The output is the estimated feature selector. There are many methods for feature selection which are classified under the wrapper and filter methods. We are calculating MI, Chi-square and IG. MI is Mutual Information which comes under filter method, which is actually a measure between two (possibly multi-dimensional) random variables  $X$  and  $Y$  that quantifies the amount of information obtained about one random variable, through other random variable. Chi-square test is used to test the dependency between two variables by using observed and expected values. Markov coefficient ( $\theta$ ) is calculated by Information Gain. It is an entropy based feature evaluation method which is defined as the amount of information provided by the feature items.
2. K-means Clustering Algorithm- Here output of MRF will be given to the K-Means Clustering Algorithm in the form of input vectors. The steps are given as follows :-
  - Input: 'K' : the number of clusters  
'D': data set containing n objects
  - Output: A set of K cluster
  - **Method:** Arbitrarily choose K objects from D as the initial cluster centers repeat Reassign each object to the cluster to which the object is the most similar, based on the mean value of the cluster. Update the cluster means, i.e. calculate the mean value of the objects for each cluster; until no changes.

## C. Formulae:

1. Calculate Mutual Information

$$\begin{aligned} mi &= (tp / n) * \log_2((n * tp) / (gPos * fPos)) \\ &+ (fp / n) * \log_2((n * fp) / (gNeg * fPos)) \\ &+ (fn / n) * \log_2((n * fn) / (gPos * fNeg)) \\ &+ (tn / n) * \log_2((n * tn) / (gNeg * fNeg)) \end{aligned}$$

2. Calculate Chi Square Method

$$\begin{aligned} ChiSquare &= (n * \text{Math.pow}((tp * tn - fn * fp), 2)) \\ &/ ((tp + fp) * (tp + fn) * (fn + tn) * (fp + tn)) \end{aligned}$$

3. Calculation of Markov Coefficient( $\theta$ ):

Information Gain

$$\begin{aligned} ig &= (gPos / n) * \log_2 (gPos / n) - (gNeg / n) * \log_2 (gNeg / n) \\ &+ (tp / n) * \log_2 (tp / fPos) + (fp / n) * \log_2 (fp / fPos) \end{aligned}$$



# International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijirccce.com](http://www.ijirccce.com)

Vol. 6, Issue 5, May 2018

$$+ (fn / n) * \log_2 (fn / fNeg) + (tn / n) * \log_2 (tn / fNeg)$$

#### 4. K-Means Formulation:

- Place 'K' points in the space that represents initial cluster centroids.
- Assign each data object to the cluster that has the closest centroid based on the Euclidean Distance.

$$d(x,y) = \sqrt{((x_1 - y_1)*(x_1 - y_1)) + ((x_2 - y_2)*(x_2 - y_2)) + ((x_n - y_n)*(x_n - y_n))}$$

- After the all data objects have been assigned to the respective cluster positions of the 'K' centroids are recalculated.
- Steps 2 and 3 are repeated until the centroids no longer move.

## IV. SIMULATION RESULTS

Admin can login using username and password. Admin provides login essentials to the customer at the time of purchasing the product as shown in the Figure 1. Customer can login using that username and password. Only authenticated customers can login to the system to give reviews. Customer can either give suggestions or compliment about the product as shown in the Figure 2.

After logged in to the system admin can see the result for individual product in which admin can see the graphical output for defined features under specified categories and also the count of positive and negative reviews with the customers' details as shown in the Figure 3.

This will help admin/ business owner to understand overall standing of a product. This would help owner to decide whether to make any changes in design, development, and manufacturing of product otherwise results would be taken as a compliment.

The screenshot shows a web browser window with the address bar displaying 'localhost:8080/ProductReview/register.jsp'. The page content is as follows:

**PERSONAL INFORMATION**

First Name\*

Last Name\*

Address\*

Gender\*  
 Male  Female

**BILLING INFORMATION**

Bill ID\*

Password\*

Figure 2: Customer Registration

# International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijirccce.com](http://www.ijirccce.com)

Vol. 6, Issue 5, May 2018

Registration process shown in the Figure 2 is for customers but only admin can operate it after the purchase. Here automatically username and password is newly generated for every new user. When user wants to review the product he/she can login to the website and then post the review.

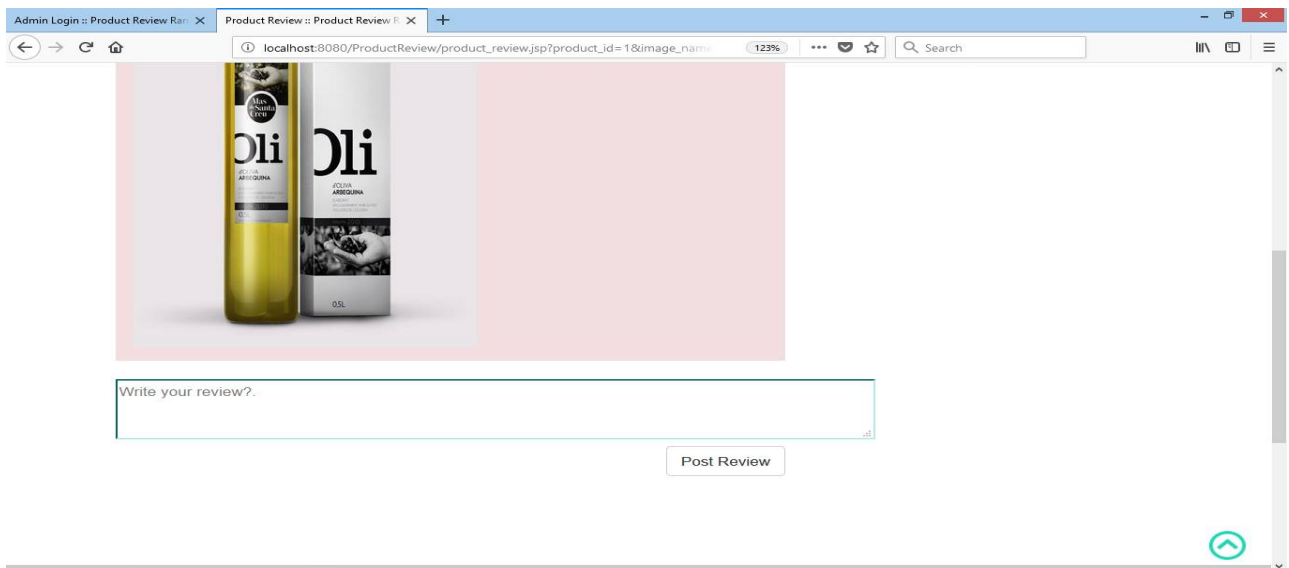


Figure 3: Post Review

Using the provided username and password user can log in to the company's website. After logging in he/she can choose a section (Hair, Face, Eyes, Body) and inside a section can select a product to post a review

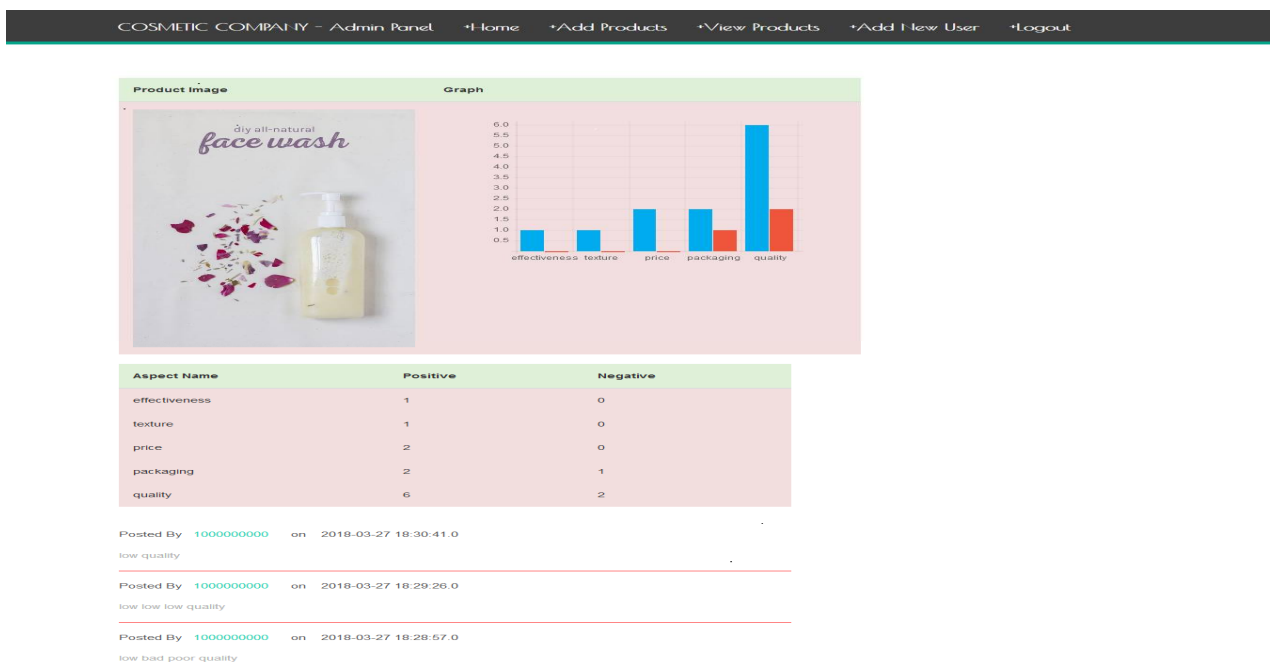


Figure 4: Main Output At Admin Side



# International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 6, Issue 5, May 2018

In Figure 4 main output for a particular product is shown. Admin will see the outcome in this format. Just like customer admin has to select a section (Hair, Face, Eyes, Body) and then the product to see the result. Also this same can be achieved through the 'View Result' tab. In this result admin can see the graphical output of features he provided in positive and negative bars. Also he gets the positive and negative count of reviews for every features. And at last admin can also see which user posted what review at what time. This will help admin to understand the reason behind the negative results.

## V. CONCLUSION AND FUTURE WORK

Therefore this proposed method is useful to give a strategy to the cosmetic companies using unsupervised K-means Clustering algorithm which is compatible with MRF feature selection since it can achieve the best performance in the clustering. This process can be applied to determine business strategy and to improve products and services according to customer requirements using opinion mining.

## REFERENCES

1. Q. Cheng, H. Zhou, and J. Cheng, "The Fisher-Markov Selector: Fast Selecting Maximally Separable Feature Subset for Multiclass Classification with Applications to High-Dimensional Data," IEEE transaction on Pattern Analysis and Machine Intelligence, Vol. 33, pp. 1217–1233, June .2011.
2. Naniek Widyastuti, AmirHamzah, "Opinion Classification Using Maximum Entropy and K-Means Clustering," IEEE transaction on International Conference on Information, Communication Technology and System, April .2017
3. Xiaohui Yu, Jimmy Xiangji Huang, "Mining Online Reviews for Predicting Sales Performance: A Case Study in the Movie Domain," APRIL 2012. ARTICLE from Published Conference Proceedings. IEEE transaction on Knowledge and Data Engineering, VOL. 24, NO. 4, APRIL 2012
4. Niphath Claypo and Saichonjaiyen, "Opinion Mining for Thai Restaurant," IEEE transaction on Knowledge and Smart Technology, pp. 105-108 MARCH 2015.
5. Nipaporn Thessrimuang and Orawan Chaowalit, "Opinion Representative of Cosmetic Products", IEEE pp. 1-6, 2016.