

# Mining Competitor for Electronic products along with Users Sentiment Analysis

<sup>1</sup>Swaraj Kothari,<sup>2</sup>Rohan Muthekar,<sup>3</sup>Nikhil Philip,<sup>4</sup>Shraddha Shelar,<sup>5</sup>P.PShevatekar

<sup>1</sup>BE Comp. Student,<sup>2</sup>BE Comp. Student,<sup>3</sup>BE Comp. Student,<sup>4</sup>BE Comp. Student,<sup>5</sup>HOD

<sup>1</sup>Computer Engineering,

<sup>1</sup>DYPIEMR, Pune,India

**Abstract:**In this competitive world of business, success is on the produce's appearance and its features to the customer at the same time its cost with its quality is seen. Large customer population buy online produces for saving time and money but it is important to know about the product which will be efficient to buy in less price. In this paper opinion mining is suggested for Analysis of Sentiments behind the product which have been purchased by customer. Our evaluation of competitiveness utilizes customer reviews, an abundant source of information that is available in a wide range of domains. We present efficient method for evaluating competitiveness and mining opinion from large datasets. Cosine similarity algorithm is implemented in this paper to find the similarity between different items.

**Index Terms - Data Mining, Web mining, positive, negative, Information retrieval, cosine similarity**

## I. INTRODUCTION

Now days everyone wants the product with the best features and the best suitable price. But is it possible to find the product that has the best feature when we have a thousand's of product with a lot of similarity. Mining Competitor is a solution that we can use in order to solve this situation. Using this algorithm, we can compare huge amount of data and find all the competitor of a particular product based on their features. With the help of this algorithm we also can find the best products available with their price constrain. We are making use of Watson's dictionary in this algorithm.

One additional feature that has been added to this project is Sentiment analysis. The user using this product can also go through various public reviews and these reviews are being differentiated based on their value for example, "Moto E4Plus is a great phone" would be a positive review. And "Moto E4Plus is a heavy phone so not easy to use" this is a negative review. So the user using this application can also differentiate the product based on various reviews put by the user.

## II. PROPOSED SYSTEM

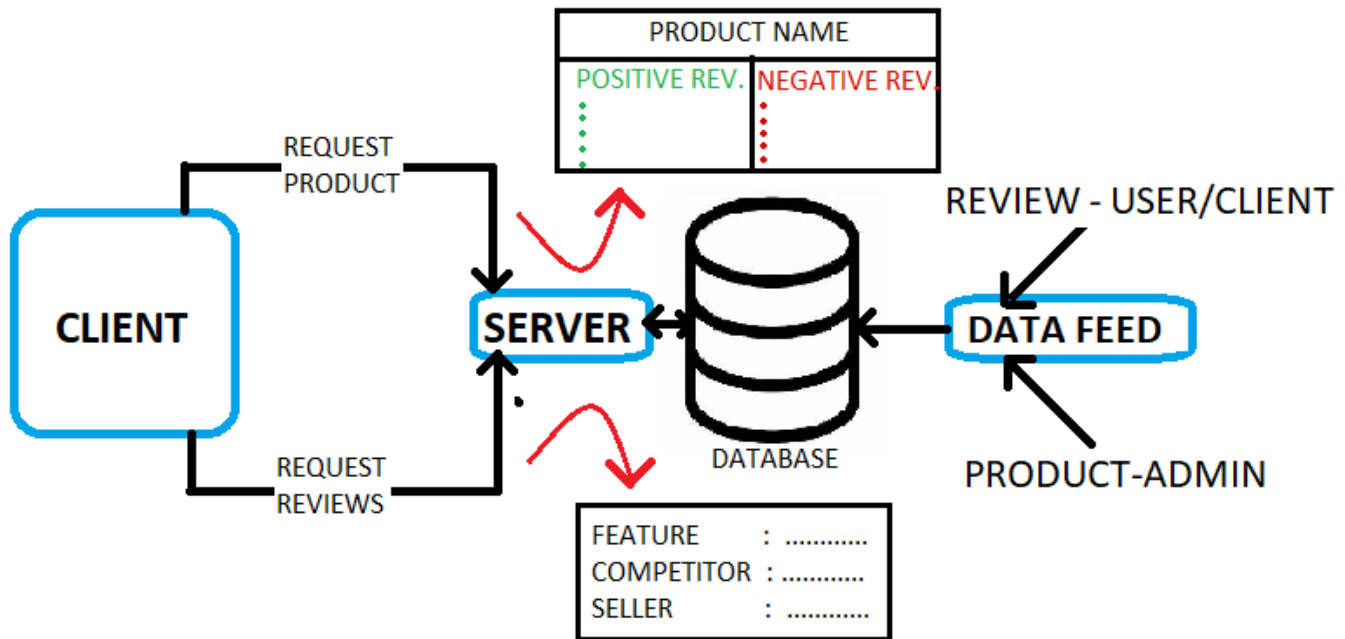
In this paper we are using the competitiveness between two product and we are trying to give out the best products that are similar to what the user has searched for. In order to increase the efficiency of our results we are using a large data set with variety of products so that the user will have the best possible result. Along with it we are also providing the user the online sellers of a particular product where our user can be redirected to the seller site when the seller option is selected.

One unique thing that we have added in our project is that we have combined the scope of sentiment analysis with mining competitor where the user can see the sentiments public have given for a particular product which is calculated using various public reviews with the help of a formal dictionary which is used in sentiment analysis.

## III. RELATED WORK

Initially public used to buy only products of reputed brands as they used to consider these reputed brands to be more reliable and with many more advance functionalities. Due to which smaller brands providing better functionalities were ignored in the market. With the help of mining competitor, the global public can now compare various products of different brands and by without being loyal to a particular brand due to its reputation. Along with it the global users can also go through the various reviews that the user has submitted based on their using experience which could be further analyzed into positive review or negative review.

### 3.1 Architectural Diagram



### 3.2 Data and Sources of Data

Data is very important to compare the devices or products and to analyse sentiments or opinion of customers who already have the product. Here both client and administrator have to involve for producing the data as new gadgets come in market administrator has to update that to the database. while client/customer will write review of their product.

### 3.3 Functional Framework

Client can perform many functions throughout the application like buy their desired product, compare the product with other competitors, reviews about their product, check different prices of the same product from different online shopping websites. But there are two main features which are the primary sources of this project they are -:

I. Request product

Customer request for the product and a stack of products are displayed according to the similarity which is calculated using **cosine similarity**. This algorithm performs computation on the data fetched from the database and the results are displayed accordingly. This provide the customer a better view of high quality product and comparison with the desired product. These values are dynamically updated based on their cosine values. The cosine values can also be calculated for string values.

**Cosine Similarity:**

Cosine similarity in layman terms can be defined as the similarity between two files or documents or entites we take into consideration.

Formula:  $\text{cosine}(\text{entity1}, \text{entity2}) = (\text{entity 1} \cdot \text{entity 2}) / \|A\| \|A\|$

Following example creates a crystal-clear view of the specified phenomenon:

entity 1 = (0, 2, 0, 0, 1, 0, 0, 2, 0, 3)

entity 2 = (1, 1, 0, 0, 1, 2, 0, 1, 0, 2)

entity 1 · entity 2 = 0\*1 + 2\*1 + 0\*0 + 0\*0 + 1\*1 + 0\*2 + 0\*0 + 2\*1 + 0\*0 + 3\*2 = 11

$\|A\| = \sqrt{(0^2 + 2^2 + 0^2 + 0^2 + 1^2 + 0^2 + 0^2 + 2^2 + 0^2 + 3^2)} = \sqrt{18} = 4.24$

$\|B\| = \sqrt{(1^2 + 1^2 + 0^2 + 0^2 + 1^2 + 2^2 + 0^2 + 1^2 + 0^2 + 2^2)} = \sqrt{12} = 3.46$

Thus, after inserting the above values in the given formula -:

$$\text{Similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{x=1}^n A_x B_x}{\sqrt{\sum_{x=1}^n A_x^2} \sqrt{\sum_{x=1}^n B_x^2}}$$

Cosine (entity1, entity 2) = 0.749

The cosine of the angle between them is about 0.749. Here by 8 dimensional vectors are taken into consideration. A primary virtue of using cosine similarity is that it converts the defined question that is beyond our imagination to a possible dynamic outcome in this case we can conclude that the result is at a very large distance from zero and is almost at a perfect agreement or in sense close to one. Thus, cosine similarity is of great advantage as it focuses mainly on orientation and gives optimistic results.

II. Request reviews

Reviews which are provided by the user are then accumulated in database and then sentiment analysis is done for a particular product with respect to its reviews. The more promising review more is the product impact over the customer. Sentiment analysis plays a keen role in providing that functionality.

### Sentiment Analysis

A sentence can be divided into 3 categories namely: -

- sentiment feature
- product feature
- stopwords

This is a simple dictionary comparison algorithm where sentences which are fetched from the database go through the following steps.

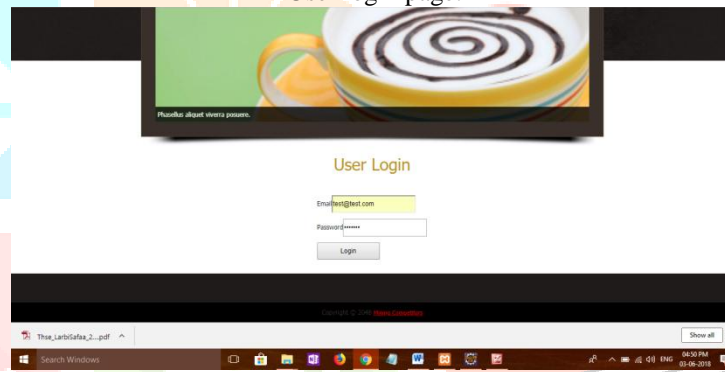
- I. Data Cleaning – removing punctuation and stop words from sentences.
- II. Tokenization – sentences are divided into words called as tokens.
- III. Data Comparison – now each word is compared with positive and negative repositories and the assigned as positive, negative or neutral

## IV.RESULTS AND DISCUSSION

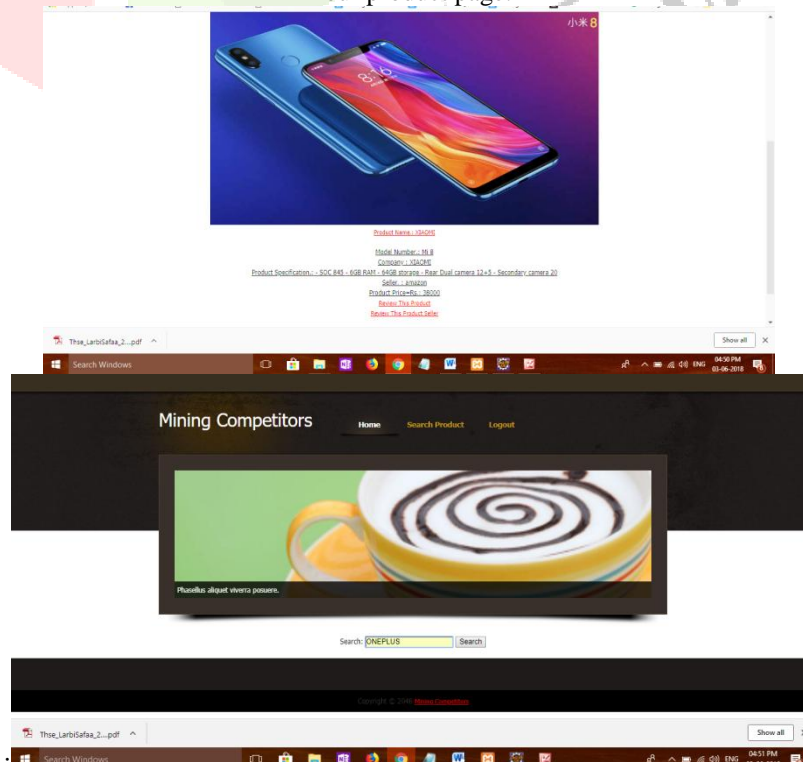
### 4.1 Result:-

The clubbing of cosine similarity along with sentiment analysis is one of the competent phenomenon in the world of business . This thereby is a great factor and brings about a dynamic revolution in social and economic perspepts. In our project we get a more of a crystal clear view of the product which might be beneficiary both in quality as well as in featurisite aspects . Thus mining competitor creates imaginary guidelines which will surely help us select the best products according to the predefined requirements.

Flow of project :  
User login page:



Your product page:





Best Seller Flip card:

Result analysis is the last phase of the project development where the cost, quality, efficiency, through is taken into consideration. Accurate, timely recognition of project profitability, for each project, for every period end, is very important in any company.

## V. FUTURE ENHANCEMENT AND CONCLUSION:

The present level of empirical research done in the electronic world is minuscule's research and studies conducted hereby are largely based on perceptions and evaluations of electronic goods which help us attain the goal of priorities in the best possible way. In the future, we can also seek other relevant research problems from the fields admits ours. In the terms of future enhancements, a variety of improvisations can be made in the techniques of cosine similarity and dictionary to simplify customer perceptions and attitudes. Everyday terabytes of data get generated from millions of shoppers and their opinions can be used to enhance the business and product related enhancement strategies. These algorithms can together help in modelling and predicting the user's mentality and in future can help the retailers to know their customers' needs better. Thus, a combination of these algorithms can be applied in other sectors like banking, medicines, education and so on. This thereby helps us gain knowledge from an enormous amount of relative data.

Clementine, SAS can further be improvised to adopt these techniques together to bring a revolution in the world of customer relative enhancements. Thereby we can also develop certain closures and niche to segregate the students, elderly as well as male and female users. These techniques can further be implied in online retailing, emerging brands and multilevel marketing propaganda for proper decision making criteria in future.

Cosine similarity clubbed together with sentiment analysis is one of the most competitive phenomenon is the platform of business and speculations. This system nullifies the protocol of forecasting and decisions which has brought great social and economic benefits. Comprehensive integration of products and reviews have resulted in explosion of information. This has greatly helped to effectively analysis and act upon the information that the users contain. A huge combinational leap has helped solve the end user's problems. These reviews and comparators have proved to be quite a beneficiary factor and have helped expand the horizon on quite a wider audience.

## VI. ACKNOWLEDGMENT

This research paper stands as a collective effort of every sentinel member in our group including teachers, online proactive guidance as well as the essence of our entire project team. This is thereby our acknowledgement to dedicate our sincere gratitude towards Mrs. P.P. Shevatekar ma'am for her valuable guidance and support throughout our project. Likewise, we would also like to express our sincere gratitude towards MR. Ishwar sir and the coherent staff whose unconditional support acted as a driving force throughout the project. The motivation of this project and the relative paper would not have been possible without them.

## REFERENCES

- [1] S. R. Barahate, V. M. Shelake, "A Survey and Future Vision of Data mining in Educational Field", Proc. 2nd Int. Conf. on Advanced Computing & Communication Technology, pp. 96-100, 2012.
- [2] Weil Kevin, "Measuring Tweets", Twitter Official Blog, February 2010, [online] Available: <http://www.internetlivestats.com/twitter-statistics>.
- [3] Graham L. Giller (2012). "The Statistical Properties of Random Bitstreams and the Sampling Distribution of Cosine Similarity". Giller Investments Research Notes (20121024/1).
- [4] Sidorov, Grigori; Gelbukh, Alexander; Gómez-Adorno, Helena; Pinto, David. "Soft Similarity and Soft Cosine Measure: Similarity of Features in Vector Space Model". *Computación y Sistemas*. 18 (3): 491–504.