

# Comment Based Product Recommendation System using Association Rule Approach and Naïve Bayes Algorithm

Patel Jaimini<sup>1</sup>, Mesariya Manisha<sup>2</sup>

<sup>1</sup>Assistant Professor, <sup>2</sup> Student

<sup>1,2</sup>Department of Computer Engineering

<sup>1,2</sup>Kalol Institute of Technology and Research Center, Kalol, Gandhinagar, Gujarat, India.

**Abstract**—Recommendation System have emerged rapidly in the past decade. Recommendation systems on E-commerce websites helps consumer to find products. A recommendation systems learns consumer behavior in order to suggest products to those consumers. Recommendation system allow consumer to have new experience discovering new products rather than needing to search for them. When making purchase decisions consumer often use the comments left by previous buyers to help them. We are going to research on Recommendation system that can help the use to find the better product from large data set.

**Index Terms:** Recommendation system, Association Rule Mining, K-Means Algorithm, Naïve Bayes Algorithm.

## I. INTRODUCTION

Data mining is defined as a process used to extract usable data from a large data set of any raw data. Data mining has applications in multiple fields, like science and research. This helps businesses be closer to their objective and make better decisions. E-commerce is the easiest, most commodious way of organizing business over the internet for business experts and individuals. Managing business over the internet is simply surfing specific website for shopping products online or business related matters. E-commerce industries one of the fastest growing sectors with the expected market growth around USD 18 billion by 2017 in India. Recommendation system on e-commerce websites suggest product to consumers making it easier them to select the right products to meet their needs. Among shopping sites, there is huge competition in presenting their products, discount offer and buying experience they given to users. All these promotions and discounts are depending on analytics and business research done by professionals inside and outside of different firms. Consumer reviews and product ratings are the main parameters that companies used in the e-commerce sites in order to strategize the analysis. These reviews provide a crucial role in users to decide about buying an item or not. Thus, examining consumer feedbacks help shopping companies and manufactures who can identify specific area of improvement in their products. It is very difficult to read through each individual review of different item and make a good decision for an individual customer. For instance Flipkart E-Retailing company recently launched Moto G, 3<sup>rd</sup> generation mobile products has almost 15660 ratings and 5601 feedbacks. So it will be a tough task to read through these reviews before making decisions and also it is difficult to identify the important features of products which cannot be identifies by looking at the reviews. E-commerce sites and social media helps people to communicate with themselves and express their comments about different people, items or products, etc. These customer's generated comments are valuable sources of knowledge that can help users make proper judgment and assists developers of products in keeping track of user feedbacks. The E-commerce landscape is continually changing and evolving. One of the biggest changes in recent years has been the growing of online reviews. According to Spiegel Research Center, 2017, nearly 95% shopper read online reviews before making a purchase. Consumer don't just want the details of the products, as described in the seller's marketing, they also need to analyze reviews from other consumers before buying it. The user comments section highlights the product attributes as well as advantages and disadvantages after its use.

Recommendation system use Artificial intelligence concepts to provide users items recommendations. Recommender system divided into three main categories, depending on the information that recommendations are based on: Collaborative filtering, Content based filtering and Hybrid filtering.

Machine Learning is a sub branch of Artificial Intelligence. Machine Learning is sub fields of data science that focuses on designing algorithm that can learn from and make predictions on data. Both Data Mining and Machine Learning are area which have been inspired by each other, though they have many things common, yet they have different ends. Machine Learning is a technique which develops complex algorithms for processing large data and delivers results to its users. It uses a complex program which can learn through experience and make predictions. The algorithms are improved by itself through regular input of data. The goal of machine learning is to understand data and builds models from data that can be understood and used by humans. Machine learning algorithms will continuously run and improve the performance of system automatically, also analyze when the failure can occur.

## II. SYSTEM DESIGN

Algorithm steps:

Input: Dataset with user information and products ratings and reviews – Amazon Review Dataset or Home Depot's Dataset with product data set.

1. For a business without any user-item purchase history, a search engine based recommendation system can be designed for users
2. The Product recommendation can be based on textual clustering analysis given in a product description and reviews
3. Generate Utility Matrix for all possible user-item preference with ARM
4. Feature Extraction form product description using TFIDF vectorization
5. Converting a text from product description into numerical data for analysis
6. Layer the words in sorted order
7. Perform clustering(K-Means) for prediction
8. Choose the cluster with the highest frequency of occurrence of the word
9. Display the product based on corresponding cluster based
10. NB classification – (per user – classification with products)
11. Display classifies results



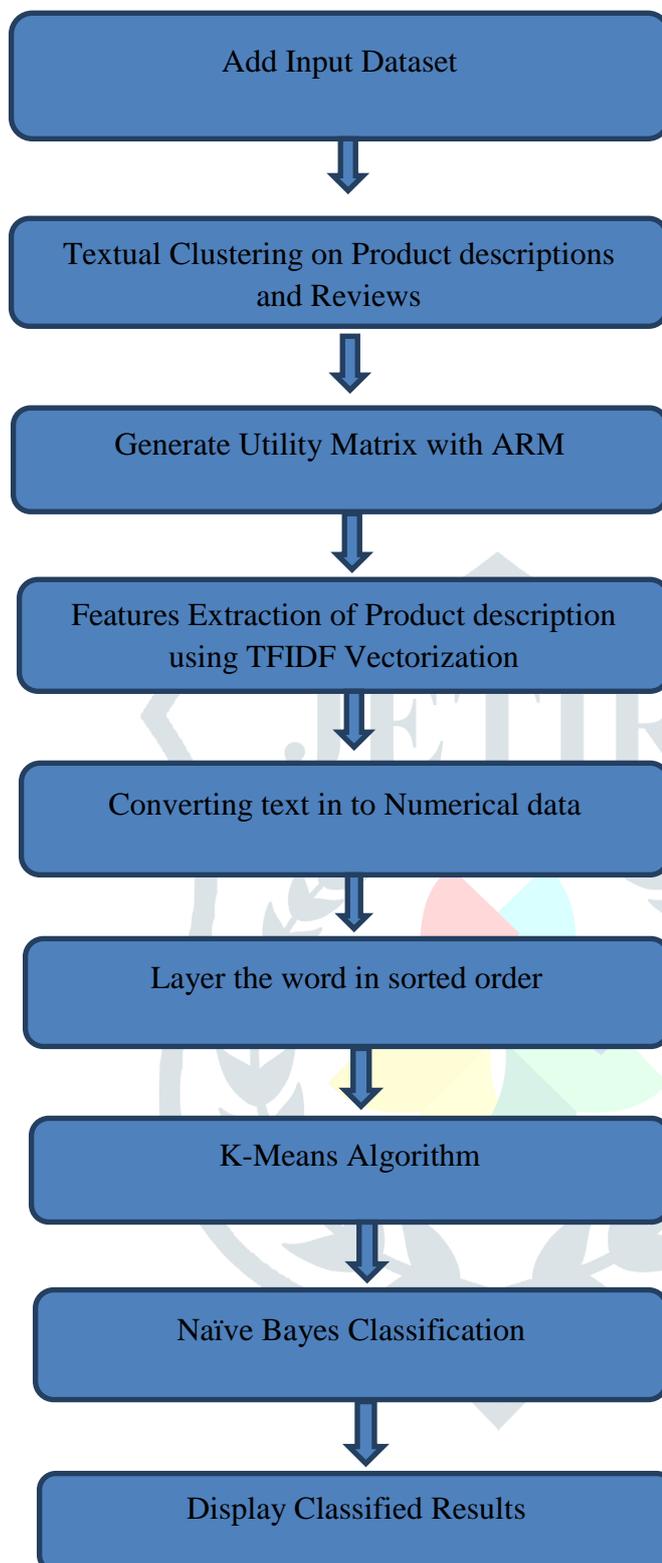
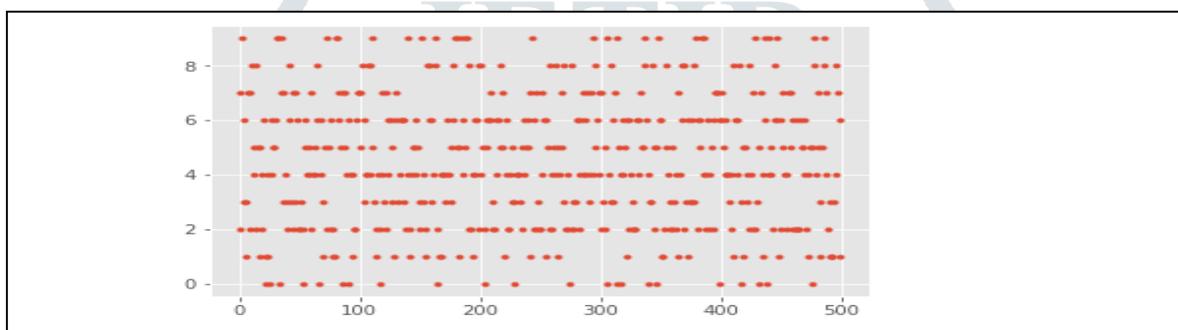


Figure 1. Flowchart

III. EXPERIMENTAL RESULT

coefficient correlation based on various conditions with given customized dataset from Amazon	
Amazon review	0.853
ratings beauty	0.828
Home depot's product reviews	0.874
Home depot's product ratings	0.954
Home depot's combined	0.921

1. Visualizing product clusters in subset of data



```

Top terms per cluster:
Cluster 0:
brass
end
soap
tube
ul
volt
caps
amp
fuses
lotion
Cluster 1:
power
tool
brush
easy
volt
free
    
```

## 2. NB based classification for product features

```

In [36]: 1 show_recommendations("cutting tool")

Cluster 1:
power
tool
brush
easy
volt
free
cutting
lithium
dust
work

• Keyword : spray paint

In [28]: 1 show_recommendations("spray paint")

Cluster 2:
used
trim
painted
65
proposition
shoe

• Keyword : water

In [39]: 1 show_recommendations("water")

Cluster 9:
water
handle
easy
tank
heater
toilet
flush
design
watering
gal

```

#### IV.CONCLUSION

The recommendation process depends on types and volume of data. Categories like frequently bought together, similar items, higher ratings category products, similar user choices based on area, gender or users who have recommended some products based on their experiences are used to filter products for suggestions. In this approach, two main categories are reviews with positive words and higher ratings are focused. These are the factors based on which a utility matrix is generated. Top-N products which are suggested are best suitable as recommendation. The results displayed are comparatively better than the earlier approach. In future, the product recommendation can be based on image displayed with camera just as Google lens. It can be used to find out not just similar products as Google lens but also frequently bought categories.

#### V.REFERENCES

- [1] Sentiment Analysis Approaches for Movie Reviews Forecasting: A Survey.
- [2] Machine Learning-based Product Recommendation using Apache Spark.
- [3] Product Recommendation Based on Customer Behavior using Data Mining.
- [4] RECOMMENDATION WITH DATA MINING ALGORITHMS FOR E-COMMERCE AND M-COMMERCE APPLICATIONS
- [5] Market size of IT industry in India. <http://www.ibrf.org/industry/information-technology-india.aspx>.
- [6] An Euclidian Similarity Measurement Approach for Hotel Rating Analysis
- [7] Singh, Rajni and Raj deep Kaur. "Sentiment Analysis on Social Media and Online Review. International journal of Computer Applications 121, no. 20(2015).
- [8] Pouransari, Hadi, and saman Ghili. "Deep learning for sentiment analysis of movie reviews". (2014)