International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 06 | June 2021 www.irjet.net p-ISSN: 2395-0072

# A Comprehensive Analysis of Recommendation System for E-Commerce Mobile App using various Algorithms

**Parth Roy[1], Prateek Rao[2], Abbas Sariya[3], Jay Gajre[4], Prof. Sumitra Pundlik[5]**

*[1-4]Student, Dept. Information Technology, MIT School of Engineering, Maharashtra, India*
*[5]Professor, Dept. Information Technology, MIT School of Engineering, Maharashtra, India*

---***---

**Abstract** - *This Paper gives an insight about the recommendation system used by the E-Commerce websites. Recommendation system is one of the important aspects of marketing and sales for an E-Commerce Website. These websites use this method to generally increase their sales and also to suggest customers with some great deals. Recommendation System is extensively used for providing programmed personalized proposals for information, products and services. The advantage of e-commerce over traditional Shopping is the user can browse, compare prices and order merchandise sitting at home on their PC. TF-IDF is a measure of originality of a word by comparing the number of times a word appears in a doc with the number of docs the word appears in.*

***Key Words*: Recommendation System, E-Commerce, content, collaborative & average weighted recommended algorithm, TF-IDF, k means, Truncated SVD.**

## 1.INTRODUCTION

Recommender systems are used by E-commerce sites to suggest products to their customers. The products can be recommended based on the top overall sellers on a site, based on the demographics of the customer, or based on an analysis of the past buying behavior of the customer as a prediction for future buying behaviour. Recommender systems are machine learning systems that help users discover new products and services. All stretch to shop online; a recommendation system is guiding you towards the most likely product you might purchase. Unlike traditional commerce that is carried out physically with the effort of a person to go & get products, ecommerce has made it easier for humans to reduce physical work and to save time. E-Commerce which was started in the early 1990's has taken a great leap in the world of computers, but the fact that has hindered the growth of e-commerce is security. Safety is the task facing e-commerce nowadays & there is still a lot of progress made in the field of security. To increase the use of e-commerce in developing countries, B2B e-commerce is implemented for improving access to global markets for firms in developing countries. Aimed at an evolving country advancement in the ground of e-commerce is essential. The research strategy shows the importance of e-commerce in developing countries for business applications.

This paper gives an insight to how machine learning can be implemented in e-commerce app by using K-means and Truncated SVD model. Following with its results and conclusion. Before going to these algorithms, we will have following discussions.

### 1.1 E-Commerce

Electronic commerce or ecommerce is a term for any type of business, or commercial transaction, that involves the transfer of information across the Internet. It protects a range of unlike types of businesses, from shopper-built retail sites, through sale or harmony spots, to business exchanges trading goods and services between corporations. It is now single most vital aspects of the Internet to occur.

The user moves complete the internet to the supplier's web site. From there, he decides that he wants to purchase something, so he is moved to the online transaction server, where all of the information he gives is encrypted. When the order is placed his order, the information moves through a private gateway to a Processing Network, where the issuing and acquiring banks complete or deny the transaction. This generally takes place in no more than 5-7 seconds.

### 1.2 Recommendation System

Recommendation systems are important and valuable tools for companies like Amazon and Netflix, who are both known for their personalized customer experiences. Each of these companies collects and analyses demographic data from customers and adds it to information from previous purchases, product ratings, and user behaviours.

These details are then used to predict how customers will rate sets of related products, or how likely a customer is to buy an additional product.

Before diving into specific recommendation engine applications from well-known retailers and online services, we think it's important for us to explain the different approaches to recommendation systems.

#### 1.2.1 Collaborative Filtering

Collaborative filtering systems work by collecting user remark in the form of ratings for items in a specified ground and exploiting similarities in rating actions amongst several users in determining how to recommend an item.

Collaborative filtering systems recommend an item to a user based on opinions of other users. This type of recommendation system makes predictions of what might interest a person based on the taste of many other users.

E.g., It assumes that if person X likes Snickers, and person Y likes Snickers and the Milky Way, then person X might like the Milky Way as well.

We have used a model-based collaborative filtering system using Singular Value Decomposition technique.

### 1.2.2 Content-Based Filtering

This type of recommendation system focuses on the products themselves and recommends other products that have similar attributes. Content-based filtering relies on the characteristics of the products themselves, so it doesn't depend on supplementary employers to interact with the goods before making a recommendation. a content-based recommendation approach analyses a set of documents and/or descriptions of items previously rated by a user, and builds a model or profile of user interests based on the features of the objects rated by that user. Techniques involved are:

- TF-IDF
- Naive-Bayes

### 1.2.3 Demographic Based Recommender System

This type of recommendation system categorizes users based on a set of demographic classes. This process requires shop research data to fully implement. The foremost assistance is that it doesn't need antiquity of user ratings.

### 1.2.4 Utility-Based Recommender System

This type of system makes recommendations based on a computation of its usefulness for each individual user. This relies on each industry's ability to decide on a user-specific utility function. The main advantage of this system is it can make recommendations that are unrelated to a product's attributes, such as availability and vendor reliability.

### 1.2.5 Knowledge-Based Recommender System

This type of system makes suggestions based on information relating to each user's preferences and needs. Using function knowledge, it can draw connections between a customer's need and a suitable product.

### 1.3 TF-IDF

In information retrieval, TF-IDF, short for term frequency–inverse document frequency, is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus. It is often used as a weighting factor in searches of information retrieval, text mining, and user modelling.

The TF-IDF value increases proportionally to the number of times a word appears in the document and is offset by the number of documents in the corpus that contain the word, which helps to adjust for the fact that some words appear more frequently in general. This technique can be well suited in a content-based recommendation system approach.

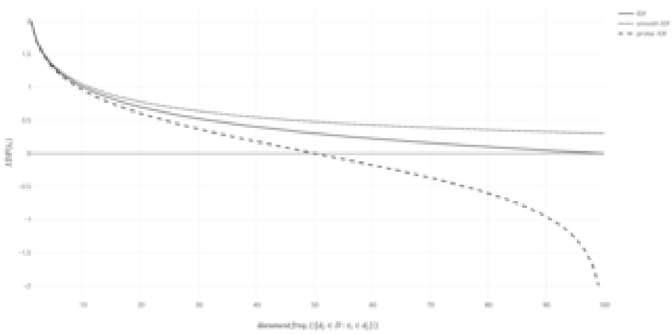

**Fig-1:** Plot of different inverse document frequency functions: standard, smooth, probabilistic.

Variations of the TF-IDF weighting scheme are often used by search engines as a central tool in scoring and ranking a document's relevance given a user query. It can be successfully used for stop-words filtering in various subject fields, including text summarization and classification.

## 2. METHODOLOGY

### 2.1 System Architecture Design

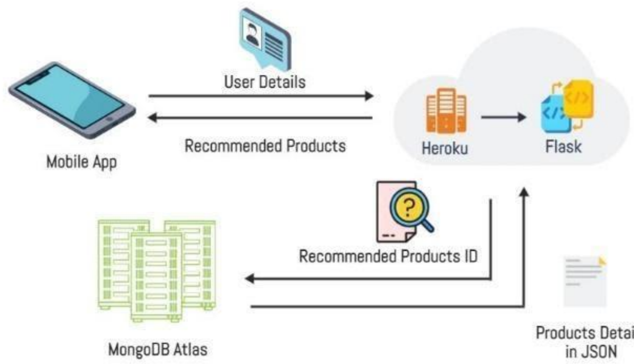


**Fig-2:** Architecture Diagram of our app

In the above figure, the user uses mobile app and his details are sent to the Heroku platform where the recommendation algorithm resides. Further, from the user details, User ID and his activity is captured and passed on to the algorithm. The algorithm analyses this input and processes the top 5 recommendation product's ID as an output. These product's ID are then sent to MongoDB Atlas Server for fetching each product's information like Name, Price, Image URL and

Category. After fetching the details in json format, this json file is sent to mobile app where it shows it as beautified output.

For our project, dataset was obtained from Kaggle. It had columns like Product ID, Ratings and user ID. Each row defines the rating of the specific product.

The second dataset we used was from scrapping Dmart's website where we fetched other attributes as Name, Price, category and Image URL. Further this data was imported to MongoDB Atlas and the following algorithms were applied to make Recommendations System.

- Popularity Based Recommendations.
- SVD in Collaboration Filtering.
- k-means in Content Based Filtering

### 2.1.1 Popularity Based Recommendations

It is a type of recommendation arrangement which all of it on the principle of popularity and or anything which is in trend. These systems check about the product which are in trend or are most popular among the users and directly recommend those.

The formula used was weighted average and is given by:

**WeightedRating(WR)=[vR/(v+m)]+[mC/(v+m)]**

where, v is the number of votes for the product; m is the minimum votes required to be listed in the chart; R is the average rating of the product and C is the mean vote across the whole.

Weighted Average was calculated of every product and then the highest top 10 products were displayed of each category.

The only limitations of the algorithm where it will show same recommended products to every new user. As each user have different tastes and interest, they might not like the recommended products

### 2.1.2 SVD in Collaboration Filtering

SVD is a method from linear algebra that has been used as dimensionality reduction technique in machine learning. In the perspective of Recommendation System, this technique is used in a collaborative filtering. It has a matrix like structure where each rows represents a user, and each column represents an item. The elements of the matrix are ratings of products given by the user. It finds factors of matrices from the factorisation of a high-level (user-item-rating) matrix. The singular value decomposition is a method of decomposing a matrix into three other matrices as given below:

$$A = USV^T$$

Where $A$ is a $m \times n$ utility matrix, $U$ is a $m \times r$ orthogonal left singular matrix, which represents the relationship amongst employers and covert factors, $S$ is a $r \times r$ diagonal matrix, which describes the strength of each latent factor and $V$ is a $r \times n$ diagonal right singular matrix, which indicates the similarity between items and latent factors.

```
Below are the recommended items for user(user_id = 9):

                   user_ratings  user_predictions
Recommended Items
B00004R91K                  0.0          0.989319
B00004S4TH                  0.0          0.953622
B00004R91I                  0.0          0.953622
B00004OCKS                  0.0          0.953622
B00004S1BZ                  0.0          0.953622
```

**Fig-3:** Recommended products using SVD Technique

The above figure represents the recommendations for user_id = 9 With the predictions that user may like this product.

Evaluation of this model is done RMSE metric. It measures the closeness of predicted rating to the true rating. RMSE squares the error before summing it, it tends to penalize large errors more heavily.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(r_{ui} - \hat{r}_{ui})^2}{N}}.$$

The lower value corresponds to higher accuracy of the recommendation system.

Out[53]:

| productId | Avg_actual_ratings | Avg_predicted_ratings | item_index |
|---|---|---|---|
| B00000JGRP | 0.454545 | 0.457474 | 0 |
| B00000JGRT | 1.363636 | 1.380520 | 1 |
| B00002N8BV | 0.363636 | 0.363565 | 2 |
| B00002ND67 | 0.818182 | 0.821845 | 3 |
| B00004OCIP | 0.909091 | 0.905089 | 4 |

**Fig-4: Average actual and predicted ratings of the products**

The above figure shows actual ratings and predicted_ratings of the recommended products. As we can infer, the difference between them is so small.

```
RMSE SVD Model = 0.01949
```

We got the RMSE as 0.01949 which is very low.

### 2.1.3 k-means in Content Based Filtering.

Content based Filtering is one of the approach of recommendations system. It requires no user_rating records and no assumptions. It based on a description of the item and a record of the user's preferences. The issue hires a sequence of distinct, pre-tagged characteristics of an item in order to recommend additional items with similar properties. This method is top suited when there is sufficient information available on the items but not on the users.

In our scenario, it is applied to find products which are similar to the user, cluster models divide the product description base into many segments using TFIDF and treat the task as a classification problem. The algorithm's goal is to assign the user to the segment containing the most similar products to customers. Once a cluster is identified based on the user's search words, the recommendation organization can presentation items from the agreeing product clusters based on the product descriptions.

```
In [40]:  show_recommendations("laptop")
          Cluster 13:
Out[40]:  ['bag',
           'women',
           'laptop',
           'makeup',
           'travel',
           'large',
           'backpack',
           'black',
           'duffle',
           'bags',
           '15',
           'inch',
           'messenger',
           'capacity',
           'storage']
```

**Fig-5: Related terms from product descriptions**

The above figure shows the related terms to laptop. All that terms are resultant of the cluster 13 /60 formed by the product descriptions.

## 3. REAL WORLD APPLICATION

Application of Recommendation System in E-Commerce Websites include:

### 3.1 Amazon

Amazon has single-handedly put a spotlight on the retail value of AI, and recommendations are part of what put the company on the map (in addition to their robotics initiatives, and AWS).

Amazon.com uses recommendations as a targeted marketing tool throughout its website. When a customer clicks on the "your recommendations" the link leads to another page where recommendations may be filtered even further by subject area, product types, and ratings of previous products and purchases. The customer can even see why a particular product has been recommended.

"At Amazon, we habit commendation algorithms to mark the online stockpile for each buyer. The stockpile drastically changes built on customer interests, showing programming titles to a software engineer and baby toys to a new mother," explain Greg Linden, Brent Smith, and Jeremy York in their paper Amazon.com Recommendations: Item-to-Item Collaborative Filtering.

In this instance, collaborative filtering doesn't just match each use to similar customers. The item-to-item connects each user's purchase to similar items and compiles a recommendation list from them. For example, if you're enthusiastic about the latest technology, you may find your Amazon web page suggests the latest device and devices, if food is your thing, you're sure to find plenty of recommendations for recipe books and cookware.

According to McKinsey & Company, 35% of Amazon income is generated by its recommendation engine.

### 3.2. Netflix

According to a paper written by Netflix executives Carlos A. Gomez-Uribe and Neil Hunt, the video streaming service's AI recommendation system saves the company around $1 billion each year. This agrees them towards investing surplus funds on new content which viewers will continue to view, giving them a good ROI.

"We have exposed through the years that there is marvellous value to our subscribers in incorporating references to personalize as much of Netflix as possible," say by Xavier Amatriain and Justin Basilico (Personalization Science and Engineering) in their Netflix Tech Blog.

Netflix uses RS personalized diversity to generate Top Ten recommendations for user households, so that it can offer videos that each member of the household may be interested in. The corporation too focuses on responsiveness and encouraging trust to help develop its personalized approach. Netflix implements these strategies by explaining why it makes video recommendations and encouraging members to give feedback, so no opportunities to personalize are unexploited.

Rendering to McKinsey, 75 out of a hundred of what users watch on Netflix come from product recommendations. We discuss freely about the general web trend of "personalization" and recommendation in our AI podcast with Adam Spector of Boost Lighter, which might be worth a listen for executives considering how references might be used in their private commercial.

**Next generation of recommendation systems:**

- **More relevant recommendations:** By digging deeper into customers' benefits and partialities, recommendation systems will be able to contemporary users with more-relevant, prognostic recommendations.

- **Integrate item effectiveness:** Instead of having reference based exclusively on a client's browsing history and past purchases, this would allow businesses to control how much a profit-based recommendation differs from the traditional recommendation and to set a balance so that customer trust would not be compromised.

- **Increase merchandise reach:** Each seller has an individual range of products; improved recommendation systems would be able towards admittance a bigger range of range in order to include new or niche items in shoppers' recommendations.

- **Reach shoppers through multiple channels:** Next generation recommendation systems should be able to reach customers across a range of stations including email, social media, on an shopping website, mobile applications, and the retail customer service centres.

## 4. CONCLUSIONS

We implemented 3 algorithms out of which popular recommendations algorithm gave same products based on weighted average to every user. Secondly, the collaborative filtering approach gave 5 different recommended products for every user. Lastly, the content-based approach gave us the related terms based on product descriptions. Issues faced for our project was to acquire desired dataset. So, that we can apply algorithms to make recommended engines works more accurately. The first and last algorithm weren't much of use at this initial stage due to lack in real-time data. So, the items recommended may be liked by the user or unliked. Therefore, we have proceeded for applying TFD model to collaboration filtering in our app.

Several recommendation systems have been anticipated based on collaborative filtering, average weight-based recommendation & many more methods and till now most of them have been able to resolve the problems while providing improved recommendations results. However, due to data burst, it is required to work on this investigate area to explore and provide new methods that can provide recommendation in a wide range of applications while considering the quality and privacy aspects. Thus, the current recommendation system needs enhancement for present and future requirements of better recommendation qualities.

## REFERENCES

[1] https://en.wikipedia.org/wiki/K-means_clustering

[2] https://en.wikipedia.org/wiki/Tf%E2%80%93idf

[3] https://towardsdatascience.com/tf-term-frequency-idf-inverse-document-frequency-from-scratch-in-python-6c2b61b78558

[4] https://emerj.com/ai-sector-overviews/use-cases-recommendation-systems/

[5] http://www.creativeworld9.com/2011/03/abstract-on-e-commerce.html

[6] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, Vol. 17(6), pp. 734–749

[7] G. Adomavicius and A. Tuzhilin. Multidimensional Recommender Systems: A Data Warehousing Approach. Proc. Second Int'l Workshop Electronic Commerce.

[8] X. Jiang, W. Song, and W. Feng, "Optimizing collaborative filtering by interpolating the individual and group behaviors," in APWeb, pp. 568–578.

[9] M Viswa Murali, Vishnu T G, Nancy Victor*," A Collaborative Filtering based Recommender System for Suggesting New Trends in Any Domain of Research", in IEEE

[10] Sangkeun Lee "A Generic Graph-based Multidimensional Recommendation Framework and Its Implementations."

[11] Agarwal, N., Haque, E., Liu, H., & Parsons, L. (2005, October). Research paper recommender systems: A subspace clustering approach. In International Conference on Web-Age Information Management (pp. 475-491). Springer, Berlin, Heidelberg.

[12] Farooq, U., Ganoe, C. H., Carroll, J. M., Councill, I. G., & Giles, C.L. (2008). Design and evaluation of awareness mechanisms in CiteSeer. Information Processing & Management, 44(2), 596-612.

[13] Rajpurkar, S., Bhatt, D., Malhotra, P., Rajpurkar, M. S. S., Bhatt, M.D. R., & Malhotra, M. P. (2015). Book recommendation system. International Journal for Innovative Research in Science And Technology, ISSN: 314–316, 1(11, 1)

[14] Mohammad Aamir, Mamta Bhusry, "Recommendation System: State of the Art Approach" in International Journal of Computer Applications (0975 – 8887) Volume 120 – No.12, June 2015

[15] S.Deepa Kanmani, Antony Taurshia.A / International Journal of Engineering Research and Applications (IJERA) Vol. 3, Issue 1, January -February 2013, pp.491-493

[16] Herlocker. J.L., Konstan. J.A., Terveen. L.G., and Riedl. J., "Evaluating Collaborative Filtering Recommender Systems," ACM Transactions on Information Systems, 22(1), pp. 5-53, 2004.

[17] Michael J. Pazzani and Daniel Billsus,"Content-Based Recommendation Systems. ", 325-341, 2007.