

Product Recommendation System for E-Commerce using Collaborative Filtering and Textual Clustering

Harsh Khatter¹

Shifa Arif²

Utsav Singh³

Sarthak Mathur⁴

Satvik Jain⁵

¹Assistant Professor, Department of Computer Science, KIET Group of Institutions, Delhi-NCR, Ghaziabad, India

^{2,3,4,5}Student, Department of Computer Science and Engineering, ABES Engineering College, Ghaziabad

Affiliated to Dr. APJ Abdul Kalam Technical University Lucknow, India

¹harsh.khatter@kiet.edu

²shifa.17bcs1092@abes.ac.in

³utsav.17bcs1108@abes.ac.in

⁴sarthak.17bcs1196@abes.ac.in

⁵satvik.17bcs1025@abes.ac.in

Abstract: In the past decade internet usage has seen an exponential growth. This growth has given opportunity for other businesses which rely on it to grow and prosper too. One such avenue is E-commerce. E-commerce is growing rapidly and with increase in popularity of E-Commerce, Recommendation has become very important too. Recommendation means providing relevant suggestions to the user as per his/her interest and need. Recommendations to users can be given based on price, area of living, wish listed items, cart items, searched items and previous purchased items. Recommendation systems improve user experience, boost sales and enhance user's engagement. Our proposed recommendation system will recommend products both for new users and existing users. This recommendation system uses model based collaborative filtering and recommends products based on rating and previous purchase history of the old users. Also the new users will get recommendations of new products, trending products and products on sale. Existing users will get recommendations based on recently viewed products, complementary products, etc. Since we're setting up a new e-commerce website, initially there are no user ratings for different products, so in this case, recommendations are made based on the textual clustering analysis of product description. Model based Collaborative Filtering along with Textual Clustering will help us in improving accuracy and targeting all types of users. E Commerce is gaining popularity and the recommendation system with E Commerce is like an icing on the cake.

Keywords: Collaborative Filtering, E-commerce, Information Retrieval, Recommendation system, Textual Clustering, Web services

I. INTRODUCTION

In the past decade internet usage has seen an exponential growth. This growth has given opportunity for other businesses which rely on it to grow and prosper too. E Commerce is one such business which is gaining popularity. People can now enjoy shopping with the comfort of their home. Sellers also have great advantages because of the large

reach of customers through online mode which is not possible in traditional offline shopping. With the increase in popularity of E-Commerce, Recommendation has become very important too. Recommender Systems are algorithms that aim at providing relevant suggestions to the user. Good recommendation improves user experience and thus enhances the sale. Recommender Systems are very crucial part of E Commerce as they can help us in boosting the sale. Depending upon the need one can use the recommendation algorithm. Some of the recommendation techniques are - Collaborative filtering based, content-based filtering, hybrid and etc. Websites provide recommendation systems to users for various reasons like Minimizing search time, Enhancing user experience, Boosting sales, Increasing user engagement, Personalization [1]

With the growing popularity of the Internet and availability of so many users online, E-commerce is gaining popularity. An E-commerce website with a powerful recommendation system can increase user's interest and enhance user's experience which will lead to higher sales and profit. Figure 1 shows the features of e-commerce websites.

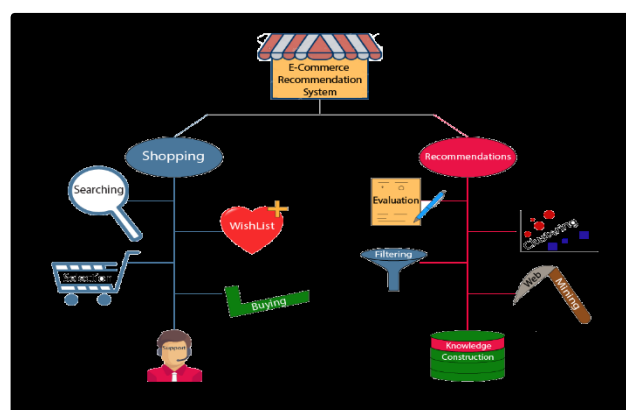


Fig 1: Features of our E-Commerce website

Everyone need things to be easy and convenient. Now people can sit at home and order their products without travelling and facing troubles. Also, E-commerce has a wide range of sellers and products which grasp the user's attention [2].

The powerful recommendation system will recommend products both for new users and existing users. This recommendation system will use collaborative filtering and will recommend products based on rating and previous purchase history of the user. Also, the new users will get recommendations of new products, trending products and products on sale. Existing users will get recommendations based on recently viewed products, complementary products, etc. Since we're setting up a new e-commerce website, initially we won't have user ratings on the different products, so in this case, we will be using textual clustering of the product description. Model based Collaborative Filtering and Textual Clustering together will make the system more efficient. Recommendation System will increase the sale and will be helpful to both the buyers and the sellers.

These are the techniques used generally for Recommendation systems but have used:

- Collaborative Filtering
- Textual Clustering

Collaborative Filtering targets the user's needs as it is related to the needs, purchase history, wish list or seen items by the users. It takes into account all these things and recommends the highest correlated product or a list of products to users to reduce time and increase efficiency. It also depends on the similar user's interest who is interested in similar items or purchased similar products as us.[16][17]

Textual Clustering is a clustering technique to cluster text-related documents. It targets the new users as our system will be unable to get any information about the user at the very beginning but our recommendation system should show some products even then, so the user will get the recommendations related to the search he/she has done.

Let us have a look at some of the previous works related to this field:

1.1. Ebay website

- Buy old and new products as per your need.
- Sell old products.
- Contact panel for queries and doubts.
- Add it to the bag or wish list.
- Easy cancellation, exchange and refund.
- eBay uses average rating, text comments as recommendation interface. Recommendation technology is aggregated rating.

1.2. Amazon website

- SMS & Payment Gateway Integrations
- Wish list or buy later.
- Product comparison with other sellers
- AmazonSmile donates 0.5% of the money to a charity.
- Amazon uses Collaborative filtering for recommendation systems. It uses Top N list, average rating, text comments, email, similar items as recommendation interface.

Table 1: Recommendation System Techniques

Recommendation Technique	Description
Collaborative Filtering	Recommends items to the user based upon the recommendations of similar users
Content Based Filtering	Recommends items to the user based on the purchase history and profile of user
Demographic	Similar users are find out based on this demographic information like age, gender, location etc.
Hybrid Recommendation	Recommends items based on combination of two or more recommendation algorithm

II. LITERATURE REVIEW

The manuscript carries the proposed approach which is defined with reference to the various related work. **Schafer J.B., Frankowski D., Herlocker J. and Sen S. [3]** have shared their ideas and analysis on one of the influencing technologies stated as "Collaborative filtering". It is basically the evaluation and filtration of the items that one will provide through the opinions of their users. They have discussed the concepts of CF, its evaluation, primary use of adaptive web, and algorithms. With all this theoretical plus implementation probability statics design rating systems. The alternative of CF, content-based personalization can be proved more effective but it would take decades for our systems(hardware/software) to automatically provide the information that is important to the person.

Zhijun Zhang, Gongwen Xu and Pengfei Zhang [4] have discussed Personalized recommendation systems. It increases efficiency, reduces time consumption of the user, and integrates the model to rate and time. Browsing for a long time and even then can't solve all your purposes doesn't make sense. Personalized recommendation systems display items according to user needs, behavior, and also apply some intelligence which it took in training. It is distributed in all fields. There are many methods like Collaborative filtering that are responsible for working on these models. **Z. Fayyaz, M. Ebrahimian, D. Nawara, A. Ibrahim, and R. Kashef [5]** have discussed that a perfect Portrait of a E-commerce website is directly proportional to the recommendation system it provides. Delivering accurate, relevant and fast results or recommendations to the users can validate the E-commerce website and its operations. Prediction is not always a bullseye but should revolve around perfection. The use of various Machine learning algorithms for trainings and make them coherent with the system is the idea.

K.Wei, J. Huang and S. Fu [6] In this paper, Authors have

discussed the recommendation system and their increasing demand. Personalization and product recommendation are being used massively in online markets. Authors have discussed more than 60 RS methods developed for E Commerce to understand the advancements and improvements that are needed in recommendation systems. After the research done by authors, authors have concluded that CF and hybrid techniques are used massively. Authors have also discussed how researchers are trying to solve the limitations of CF algorithms.

Hyunwoo Hwangbo, Yang Sok Kim and Kyung Jin Cha [7] have discussed K-RecSys which enhances and improves typical item-based collaborative filtering algorithms. K-RecSys method provides way better performance than traditional collaborative filtering. In this method authors have taken into consideration the fact that products are sold online as well as offline. Authors have also compared the performance of the new system(K-RecSys) with the traditional system. Authors through various experiments have reached the conclusion that the new system is better than the traditional. Also experiments proved that recommendations related to substitute products are better as compared to complementary products. **S. Priya and D. Mansoor [8]** have discussed three important techniques of creating a recommendation system- Collaborative Filtering , Content based and Hybrid. They have also discussed the existing E-Commerce websites and the type of Recommendation they use. Each implementation methodology has its own advantages and disadvantages and it's clearly mentioned in this paper to opt which one as per situation.[13]

Harsh Khatter and Anil Kumar Ahlawat [9][14] have discussed the importance of Recommendation and personalization and how they help in grasping the interest of users. Recommendations are very challenging because of the huge amount of data present on the internet. Clustering with Association mining rules and hybrid computing can help us in overcoming this challenge. **S. Sivapalan, A. Sadeghian, H. Rahnama and A. M. Madni [10]** have discussed the importance of the Recommendation System in E-Commerce. According to the authors , this recommendation system software in the future can even predict the demands of products. They have discussed various examples of Recommendation Systems which are used in E-Commerce websites. [14] Different recommendation systems can be implemented to

increase revenue differently. **H. Aditya, I. Budi and Q. Munajat [11]** have discussed the two approaches of Collaborative Filtering which are- Model-Based CF and Memory-Based CF. Memory based CF takes all user-item data and generates prediction, it is adaptive to data changes but takes more computational time if data size is large. Model-Based CF develops a model using user rating and then generates recommendations, it's computational time is constant regardless of the data size. Authors have discussed in details the comparative analysis of Model based and Memory based CF and have concluded that Model Based CF is better than Memory Based CF in terms of the accuracy of recommendation, computation time, and the relevance of recommendation.[15]

G. M. Dakhel and M. Mahdavi [12] have discussed that Collaborative filtering is one the best algorithm in terms of recommendations but it also suffers from poor accuracy and high computation time. Authors have discussed an approach that takes the rating dataset as input and generates cluster centers and cluster membership as output. This method uses k-means clustering to categorize users and voting algorithm to develop recommendations. On the particular data set this algorithm takes 161 sec, Item based CF takes 386 sec and User based CF takes 15030 sec. It also helps overcome some challenges faced by traditional algorithms like Scalability problem, Sparsity problem, Cold-start problem.

Learnings and Useful Extraction from these papers

From various papers we got to know the importance of recommendation system and how a good E Commerce means a good recommender system. E Commerce can be successful only if we provide relevant recommendations. Various recommendation techniques like Collaborative Filtering, Content based Filtering, Hybrid, etc. are discussed and checked. These papers also helped in deciding which recommendation technique is best for our system. Various papers helped us in deciding that Collaborative Filtering is one the massively used technique and it is also best for our system. [3][6] We also got to know that Model Based CF is better than Memory Based CF in terms of accuracy, computation time and relevance. [11] A paper also helped us in knowing that K Means clustering is one the efficient approaches that can even solve the drawbacks of CF.[12] These papers have helped us massively in reaching a conclusion and implementing an efficient system

Table 2: Literature Survey and related work done

TITLE	YEAR	IMPORTANT POINTS	USEFUL EXTRAC TION
Research on E-Commerce plat form-based personalized recommendation algorithm	2016	1: In this paper, Authors have discussed the Personalized recommendation system and its advantages in boasting the sale. 2: Personalized recommendation systems display items according to user needs. There are many methods like Collaborative filtering that are responsible for working on these models.	From this paper, we got to know about the importance of recommendation and personalization in E Commerce.
Recommendation systems:		1: In this paper authors have discussed that a perfect Portrait of a E-commerce website is directly proportional to the recommendation system it provides.	From this paper, we got to know about the importance of recommendation. We also got to know that a good E Commerce

algorithms, challenges, metrics, and business opportunities	2020	2: The use of various Machine learning algorithms for trainings and make them coherent with the system is the main idea. There is more to business needs than the consumer needs. Good Recommendation systems attract many users and thus boost the sale.	website means a good recommendation system.
Recommendation System Development for Fashion Retail E-Commerce	2018	1: In this paper authors have discussed K-RecSys which enhances and improves typical item-based collaborative filtering algorithms. 2: K-RecSys method provides way better performance than traditional collaborative filtering. 3: Authors have also compared the performance of the new system(K-RecSys) with the traditional system. Authors through various experiments have reached the conclusion that the new system is better than the traditional.	From this paper, we got to know that K-RecSys enhances and improves typical item-based collaborative filtering algorithms and is much better than many traditional algorithms.
Recommender Systems in E-Commerce: A Review	2017	1: In this paper, Authors have discussed three important techniques of creating a recommendation system- Collaborative Filtering, Content based and Hybrid. 2: Authors have also discussed the existing E-Commerce websites and the type of Recommendation they use. 3: Each implementation methodology has its own advantages and disadvantages and its clearly mentioned in this paper to opt which one as per situation.	From this paper, we got to know about different techniques of creating recommendation like Collaborative filtering, content based filtering and hybrid. We got the idea when to use these and when to use different filtering methods
Analysis of Content Curation Algorithms on Personalized Web Searching	2020	1: In this paper, Authors have discussed the importance of Recommendation and personalization and how they help in grasping the interest of users. 2: Clustering and Mining are used in this paper as a methodology to improve and enhance recommendation.	From this paper, we got to know how recommendation and personalization helps to lure the customers so that they can take interest in surfing and consider buying something from our website. Clustering is done and then mining is done due to which personalization also comes into picture.
Recommender Systems in E-Commerce	2018	1: In this paper, Authors have discussed the importance of the Recommendation System in E-Commerce. According to authors, this recommendation system software in future can even predict the demands of products 2: Authors have discussed various examples of Recommendation Systems which are used in E-Commerce websites. Different recommendation systems can be implemented to increase revenue differently.	From this paper, we got to know about the future demands of the customers to increase the revenue curve of the E-Commerce website by taking into consideration the chronology of the products bought more than expected recently.
A Comparative Analysis of Memory-Based and Model-Based Collaborative Filtering on the implementation of Recommender System for E-commerce in Indonesia: A Case Study PT X	2016	1: In this paper, Authors have discussed two approaches of Collaborative Filtering which are- Model-Based CF and Memory-Based CF. 2: Authors have discussed in details the comparative analysis of Model based and Memory based CF and have concluded that Model Based CF is better than Memory Based CF in terms of the accuracy of recommendation, computation time, and the relevance of recommendation.	From this paper, we got to know about two approaches of Collaborative Filtering- Model-Based CF and Memory-Based CF. We got to know that Model Based CF is better than Memory Based in terms of accuracy, computation time and relevance of recommendation.

III. Proposed Model:

The proposed model works for both new and old users. This recommendation system will firstly give recommendations to new users and as soon as they have purchase history, recommendations will be based on model based collaborative filtering.

This proposed recommendation system will use model based collaborative filtering and will recommend products based on rating and previous purchase history of the users. Also, the new users will get recommendations of new products, popular products and products on sale. Since we're setting up a new e-commerce website, initially we won't have user ratings on the different products, so in this case, we will be making recommendations based on the textual clustering analysis of product description. Proposed E Commerce website is having two views- one for the admin and other for the users or buyers

Operations that can be performed by Admin:

- Admin Login and Registration
- Add/Delete product

- Add/Delete category
- View ratings and orders by customers.
- Change status of order (Shipped, delivered, etc.)

Operations that can be performed by Users:

- User Login and Registration
- View Product and Add product to cart
- Wish-list product
- Rate a product
- Pay for an order (Payment Gateway)
- View recommended products

IV. IMPLEMENTATION

The website is made with MERN (MongoDB, Express, React, Node) stack. The website has both admin and user login. The heart of this project is the recommendation system and it is implemented using Collaborative Filtering algorithm and Textual Clustering. The implementation of the E Commerce recommendation system is done in three

subparts:

- 1: Popularity based recommendation system targeted at new users.
- 2: Model based collaborative filtering
- 3: Recommendation based on textual clustering of product description.

The basic steps before beginning will be loading the data sets from Kaggle and importing libraries.

The data sets are from i. Amazon product ratings by multiple users, <https://www.kaggle.com/skillsmugger/amazon-ratings>; ii. Home Depot products with descriptions, Data Source: <https://www.kaggle.com/c/home-depot-product-search-relevance/data>

1: Popularity based are the best ways to recommend products to new users. We are choosing top 10 products and recommending them to new users.

We are using the Amazon Product Review dataset for this purpose. First step will be loading amazon product review dataset and importing libraries and then fetching the top rated products.

2: Model based collaborative filtering recommends products based on the user's past purchase history and similarity of ratings provided by other users. We will be choosing top 10 highly correlated products and recommending them to users. Steps in model based collaborative filtering are-

Step 1: First step is loading the dataset and importing the required libraries. The required libraries are-

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

```
import sklearn
from sklearn.decomposition import TruncatedSVD
```

Step 2: Next we prepare a utility matrix. A utility matrix consists of all possible user-item preferences (ratings) details represented as a matrix. The utility matrix is sparse as none of the users would buy all the items in the list, hence, most of the values are unknown. Unknown values will be filled with 0. After that we transpose the utility matrix.

Step 3: Next step is to decompose the matrix for which we are using Singular Value Decomposition. SVD is a dimensionality reduction technique in machine learning. It finds factors of matrices from the factorization of a high-level 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, S is a $r \times r$ diagonal matrix.

Step 4: Next step is building a correlation matrix. Correlation for all items with the item purchased by a particular customer based on items rated by other customers who bought the same product.

Step 5: At last the top 10 products are displayed by the recommendation system to the above customer based on the purchase history of other customers in the website.

Figure 2 shows the collaborative filtering process.

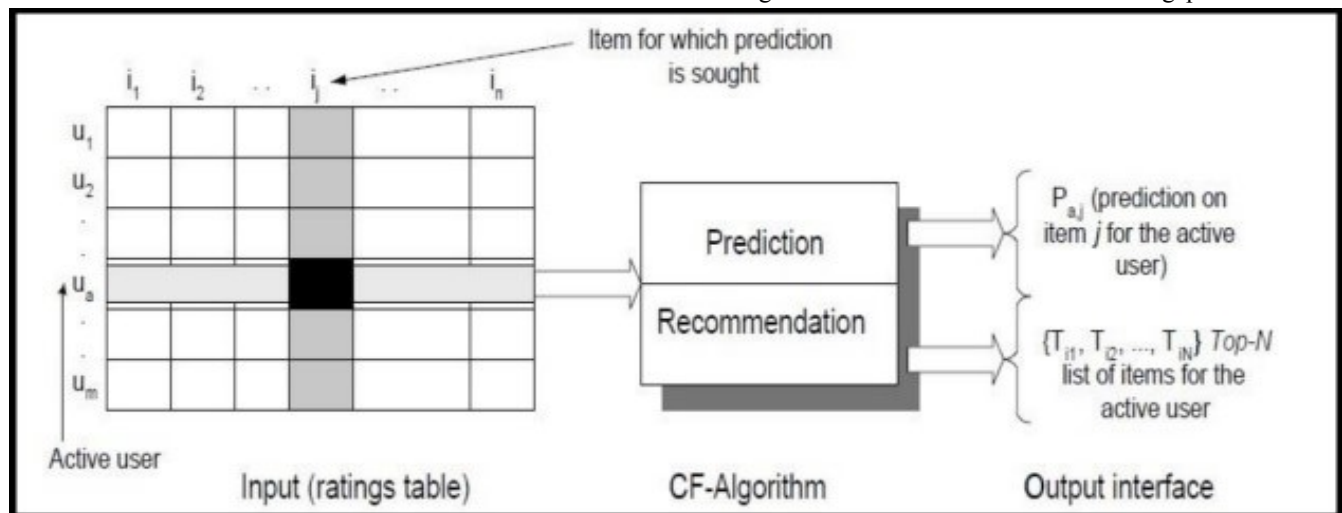


Fig 2 General working of a Collaborative Filtering Process

Textual Clustering of product description will be used when there is no or very less purchase history. Recommendations will be based on textual clustering. Based on what the user has typed in the search box, a specified cluster of products will be recommended. Steps in Textual Clustering:

Step 1: First step is loading the dataset and importing the required libraries. We are using Home Depot products with descriptions dataset. The required libraries are-

Step 2: Next step is feature extraction from product descriptions. Conversion of text in product description into numerical data for analysis takes place.

Step 3: Next step is visualizing product clusters in subsets of the dataset by using the K-means clustering algorithm.

Step 4: Last step is to predict clusters based on the words searched. Once a cluster is identified based on the user's search words, the recommendation system will display all the items from the corresponding clusters based on the product descriptions.

Model-based collaborative filtering works best when a user has bought and rated many products, but this is not possible for new users or new products. In this case, we get irrelevant and erroneous recommendations. To overcome this we have used textual clustering of the description of product. Using both model-based collaborative systems and textual clustering simultaneously we obtain higher accuracy in terms of meaningful recommendation and higher optimization in terms of computing time.

V. CONCLUSION AND FUTURE WORK

The proposed model is useful for recommending products both for new as well as existing customers. When a new user login, recommendations will be based on popular products, new products, products on sale and textual clustering of the item description. After the user has purchased history and ratings then recommendations will be based on model based collaborative filtering. With advancement in internet and technologies, recommender systems are also gaining popularity and we have to take care of both new as well as existing users which is the key motive of this proposed model.

E-commerce is not only a trend nowadays but also a necessity in the current times which is expanding day by day i.e. the number and variety of products increases on a daily basis as new customers and sellers join the network to get some profit and follow the trend. Not only customers are attracted due to the large variety, sellers also anticipate profit or growth in business and launch their products as well, as a result, the same product is sold by different sellers which increases the competition among different sellers. When there is a competition and you can't do anything like reducing the price etc so how to attract the customers to buy their product? So, the idea behind doing this is to make an equally likely platform for all the sellers and customers too. So, for this we can do multiple things and this is the future work which will make an E Commerce much better.

- Firstly, the most important thing, the user should be logged in to be eligible to give a review.
- When someone wants to give a review about a product, he/she should have bought the same item (validated by item code) by the same seller to make sure the review is not fabricated or false. This will defeat the purpose of

the seller who is trying to increase their sales by the flaws of the system.

References

- [1] Article by Becky Thames, Dated: August 25, 2020. <http://beckythames.com/motivation-for-e-commerce/>
- [2] Article by Josh Reyes, Dated: Mar 31, 2018. <https://medium.com/ecommerce-marketing-blog/7-types-of-product-recommendation-engines-your-e-commerce-store-needs-468627157d0>
- [3] Harsh Khatter, Anil K Ahlawat, "An intelligent personalized web blog searching technique using fuzzy-based feedback recurrent neural network". *Soft Comput* 24, 9321–9333 (2020).
- [4] Zhijun Zhang, Gongwen Xu, Pengfei Zhang, "Research on E-Commerce Platform-Based Personalized Recommendation Algorithm", *Applied Computational Intelligence and Soft Computing*, vol. 2016, pp. 1 – 6.
- [5] Z. Fayyaz, M. Ebrahimian, D. Nawara, A. Ibrahim, and R. Kashef, "Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities," *Applied Sciences*, vol. 10, no. 21, p. 7748, Nov. 2020.
- [6] Harsh Khatter, Brij Kalra, "A new Approach to Blog Information Searching and Curating", *Proceedings in CSI 6th International Conference on Software Engineering, IEEE*, Indore, India, September 5-7, 2012, pp. 372-377
- [7] Hyunwoo Hwangbo, Yang Sok Kim, Kyung Jin Cha, Recommendation system development for fashion retail e-commerce, *Electronic Commerce Research and Applications*, Volume 28, 2018, pp 94-101.
- [8] S. Priya & D. Mansoor. "Recommendation Systems for E-Commerce: A Review. *International Journal of Advanced Research in Computer and Communication Engineering*, Vol 6, Issue 4, 2017, pp 500 – 504.
- [9] Khatter, Harsh and Kumar Ahlawat, Anil, "Analysis of Content Curation Algorithms on Personalized Web Searching", *proceedings in International Conference on Innovative Computing & Communications*, Mar 2020.
- [10] S. Sivapalan, A. Sadeghian, H. Rahnama and A. M. Madni, "Recommender systems in e-commerce," *2014 World Automation Congress (WAC)*, Waikoloa, HI, USA, 2014, pp. 179-184.
- [11] P. H. Aditya, I. Budi and Q. Munajat, "A comparative analysis of memory-based and model-based collaborative filtering on the implementation of recommender system for E-commerce in Indonesia: A case study PT X," *2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 2016, pp. 303-308, doi: 10.1109/ICACSIS.2016.7872755
- [12] G. M. Dakhel and M. Mahdavi, "A new collaborative filtering algorithm using K-means clustering and neighbors' voting," *2011 11th International Conference on Hybrid Intelligent Systems (HIS)*, 2011, pp. 179-184, doi: 10.1109/HIS.2011.6122101.
- [13] Dubey, G., Kumar, S., Navaney, P. "Extended Opinion Lexicon and ML based Sentiment Analysis of tweets: A novel Approach towards Accurate Classifier", *International Journal of Computational Vision and Robotics (IJCVR)*, Inderscience Publishers, 2020, 10(6), pp. 505-521.
- [14] H Khatter, AK Ahlawat, "An Algorithmic approach for recommendation systems for web blogs and microblogs", *Journal of Xi'an Shiyou University, Natural Science Edition*, Volume 16, Issue 9, pp. 347 - 350, <https://www.xisdxjxsu.asia/viewarticle.php?aid=142>
- [15] Anand Prasad, Harsh Khatter, Sunita Yadav, "Query Suggestion for Long Tail Queries by Content and Keywords Matching", *Proceedings in ACEEE Springer 2nd International Conference on Computational Intelligence and Information Technology CIIT 2012 LNCS*, Chennai India, December 3-4, 2012.
- [16] Can Li, Ling Xu, Meng Yan, Yan Lei, "TagDC: A tag recommendation method for software information sites with a combination of deep learning and collaborative filtering". *Journal of Systems and Software*. Volume 170, 2020, <https://doi.org/10.1016/j.jss.2020.110783>.

[17] Anand Shanker Tewari, “Generating Items Recommendations by Fusing Content and User-Item based Collaborative Filtering”,

Procedia Computer Science, Volume 167, 2020, Pages 1934-1940,
<https://doi.org/10.1016/j.procs.2020.03>