An Innovative Way Recommend Products inE-Commerce

Girish Kumar P[1], Akhilesh M[2]

[1]B. Tech Student, Dept of Computer Science and Engineering, SITE College, Tadepalligudem, Andhra Pradesh, India
[2]B. Tech Student, Dept of Computer Science, GITAM University, Hyderabad .

---***---

Abstract - *Electronic commerce (e-commerce) refers to the purchasing and selling of products via the web. E-commerce platforms have been used by people around the globe in some form or another because everything can be purchased online with just a few mouse clicks. Since, there is a huge amount of data on every e-commerce site; a consumer may struggle to identify the product they require. In this scenario, the Recommendation System is used. A product's recommendation might be based on a variety of variables, including past search or purchase history, user reviews, and the most popular product. We have several Machine Learning-based approaches for these recommendations. We shall implement Collaborative Filtering and K-Means clustering in this work. We utilized Jupyter Notebook to develop the recommendation system, and the Amazon-ratings dataset from Kaggle was used. We will also examine numerous other recommendation techniques in this paper*

Keywords— Machine Learning, K-Means Clustering, Collaborative Filtering, Recommendation System.

1.INTRODUCTION

With the increased use of the Internet, we are seeing a dramatic surge in online purchasing. Amazon, eBay, Flipkart, and other well-known e-commerce sites are among the most popular. Customers can use Internet banking, UPI, Cash on Delivery, and a variety of other alternatives tomake transactions on e-commerce sites. People prefer to buy things online instead of visiting several shopping malls in search of their desired item. They attempt to locate their desired item in online stores. E-commerce stores offer a variety of brands and models for a single product. Customers find it tough to determine which thing to buy because there is so much data about the same item. This could cause the buyer to become disinterested.

Recommendation Systems (RS) are used by the sites to keepusers occupied in their platforms. When a consumer looks for an item, Recommendations Systems endorse some moresimilar products. Product recommendation systems suggest a product to a consumer based on a variety of factors such asthe customer's previous browsing history, previous shoppinghabits, and user profile. Recommendation Systems are frequently employed in health departments, transportation, and agriculture, in addition to e-commerce.

The health-based recommender system is a culpable system that offers both medical practitioners and end-users with suitable medical information. To avoid health risks, patients are advised to receive effective disease treatment. Health professionals benefit from the acquisition of crucial insight for clinical guidelines as well as the provision of high- quality health treatments for patients using this system [1].

Using the Transport Recommendation System on his phone, the client may obtain alternatives for modes of transportation between two different locations in the city based on his inclinations [2].

The Agriculture Department Recommendation System displays the number of questions farmers have in a given sector, such as financial assistance and crop detection, and suggests different government programs so that farmers mayget help and understand the process [3].

In this paper, we will examine various methodologies used for Recommendation Systems. We will develop a Recommendation System for a new client on an established e-commerce platform, a customer with some previous history on the platform, and a new business.

2. RELATED WORK

The varieties of recommendation systems will be discuss ed in brief.

Demographic-based Recommendation System: It is based onthe demographic characteristics of the users, such as age, ethnicity, school, profession, location, and so on. Clustering algorithms are commonly used to classify target customers based on demographic data. The system will generate the same set of suggestions if the demographic factors in this RS stay the same. As a result, they may overlook some innovative and important recommendations.

Content-based Recommendation System (CBRS): It proposes goods based on the person's account and the description of the item using CBF (Content-Based Filtering). The user's past search or purchase history may beincluded on the profile page. The algorithm learns to suggestgoods that are analogous to those that the user has previously valued. The resemblance of objects is determined by the features shared by the items being compared.

Collaborative Filtering Recommendation System (CFRS): Collaborative-based filtering makes use of the system's interactions and information obtained from other users. It generates product feature recommendations based on customer interests. CF is divided into two types.

a. The prediction computation and similarity measure are the two basic processes of *Memory-based CRS*, which are further split into two categories based on their similarity computation. *Item-based CRS*: a set of items is used forsimilarity computation.

   User-based CRS: similarity computation is basedon user similarity values.

b. *Model-based CRS:* In model-based CRS, various machine learning procedures, including clustering, Bayesian networks, Markov decision processes, dimensionality reduction, sparse factor analysis, and rule-based approaches, are used to create a model for the proposal.

Knowledge-based Recommendation System: The system provides suggestions by establishing a knowledge-based criteria based on explicit knowledge about products and users. A knowledge-based RS does not require a large quantity of data at the outset because its suggestions are independent of the user's ratings. After assessing the goods that meet the user's criteria, it offers product suggestions based on the user's preferences.

Hybrid-based Recommendation System: Hybrid RS is the result of combining different filtering algorithms, as the name implies. CBS and CFRS are the most prevalent HRS combo. Combining several filtering procedures has the goal of improving suggestion accuracy while removing the drawbacks of the particular filtering approaches.

Utility-based Recommendation System: This RS generates a utility model of each article for the user before making suggestions. This method creates multi-attribute user utility functions and clearly recommends the item with the highest utility based on each item's determined user-utility.

Countless studies and efforts have been done on Recommendation Systems using various algorithms. Some of the works will be addressed here.

Katore L.S and Umale J.S at [4] used Naive Bayes, K Star, J48 (C4.5), and Simple Cart algorithms to investigate Recommendation Systems. Naive Bayes is a probabilistic model that allows us to express nonlinearity in a systematic way by calculating the probabilities of outcomes. It is possible to tackle diagnostic and predictive issues. K star is a generalized closest neighbor approach based on transformations. It offers a consistent method for dealing with symbolic attributes, real-valued characteristics, and missing values. For classification, the J48 classifier employsa simple C4.5 basic decision tree. The data is classifiedusing the Divide and Conquer method.

Y. Koren, R. Bell, and C. Volinsky at [5] have provided strategies for a Recommendation System based on feedback, both implicit and explicit, using Matrix Factorization.

Pedro G. Campos et al. at [6] found that time-biased methods outperform their baseline and ad-hoc strategies bya significant margin. This demonstrates the potential for temporal information to be a useful input in making improved forecasts of user preferences.

Xiao and Benbasat at [7] examined at empirical publications on e-commerce product offering agents published between 2007 and 2012, as well as e-commerce product recommendation agents. They looked at subjects likerecommendation agent type and their operative aspects, userperception factors like enjoyment, suggestion quality, and social presence.

Lu, et al. at [8] analyzed papers published between 2013 and2015 in the areas of online government, online business, online shopping, online library, online learning, online tourism, online resource services, and online group activitiesin the application development of RS.

Adomavicius and Tuzhilin [9] looked at three different types of recommendation techniques: content-based filtering, hybrid filtering, and collaborative filtering. They discussed the methodology's constraints and limitations, as well as potential ideas for improving suggestion performance.

2. METHODOLOGY

We used the Recommendation System in three separate categories in this investigation. The first is for new users, and the products are suggested depending on their popularity. The second is based on the customer's previous purchases, searches, and user reviews. The third and the last solution is for a new firm that does not have any consumer feedback. We made use of two datasets for this study.

The first set of data is a collection of user ratings for various products. The latter is comprised of descriptions of products, which can be used in new firms.

We'll need the python libraries numpy, pandas, matplotlib, and sklearn to continue with this project. 'numpy' is utilised for efficient computations, 'pandas' for reading and writing spreadsheets, 'matplotlib' for data visualisation, and 'sklearn' for machine learning and statistical modelling.

A. *Popularity based Recommendation System:*

Customers who visit an e-commerce site for the first time will be guided to the most popular products on that site. We grouped the products based on the count of ratings after reading the dataset (shown in Fig1). Then we sorted the rating values in decreasing order, starting with the highest rated product. The *ratings* dataset consists of 2023070 tuples.

```
UserId,ProductId,Rating,Timestamp
A39HTATAQ9V7YF,0205616461,5.0,1369699200
A3JM6GV9MNOF9X,0558925278,3.0,1355443200
A1Z513UWSAAO0F,0558925278,5.0,1404691200
A1WMRR494NWEWV,0733001998,4.0,1382572800
A3IAAVS479H7M7,0737104473,1.0,1274227200
AKJHHD5VEH7VG,0762451459,5.0,1404518400
A1BG8QW55XHN6U,1304139212,5.0,1371945600
A22VW0P4VZHDE3,1304139220,5.0,1373068800
A3V3RE4132GKRO,130414089X,5.0,1401840000
A327B0I7CYTEJC,130414643X,4.0,1389052800
A1BG8QW55XHN6U,130414643X,5.0,1372032000
AIFAAVTUYHEHB,130414643X,4.0,1378252800
AVOGV98AYOFG2,1304146537,5.0,1372118400
A22VW0P4VZHDE3,130414674X,5.0,1371686400
AVOGV98AYOFG2,1304168522,5.0,1372118400
A6R426V4J7AOM,1304168522,5.0,1373414400
A22VW0P4VZHDE3,1304174778,5.0,1372896000
AKGB62WGF35J8,1304174778,5.0,1372896000
A22VW0P4VZHDE3,1304174867,5.0,1373068800
A1BG8QW55XHN6U,1304174867,5.0,1372291200
A1BG8QW55XHN6U,1304174905,5.0,1372291200
A22VW0P4VZHDE3,1304196046,5.0,1372896000
A22VW0P4VZHDE3,1304196062,5.0,1372896000
A3A4C2K3TWDAO5,1304196070,1.0,1378425600
```

Fig 1: The ratings dataset (first 25 values are shown)

Using the bar plot, we found which product has received the highest rating. On the new user's home page, this product isrecommended.

B. *Model-based Recommendation System:*

Because it assists in product prediction for a given user by identifying patterns based on preferences from various user data, a model-based collaborative filtering approach is used. It makes recommendations to consumers based on their previous purchases and the similarity of ratings offered by other users who bought comparable goods.

In this approach, we created a utility matrix with productId in columns and userId in rows. The values in the matrix represent the customer's ratings of the specific product. Zero is used to fill the unknown quantities. A utility matrix is used to express all possible consumer behavior (ratings). The utility matrix is weak (sparse), and most of the values are undefined, since none of the consumers would buy everything on the page.

The obtained utility matrix is now transposed. For our convenience, the matrix is now decomposed using SVD (Singular Value Decomposition) with 10 components. The decomposed matrix yields the correlation matrix.

Because the person has already looked for or purchased a product, the productId of the previously searched product will be separated. The correlation between the customer's purchase and all other things purchased by other customers buying for the same product is obtained. The recommendation system offers the user the top ten goods based on the purchase behaviour of other consumers on the website.

C. *Description-based Recommendation System:*

Without any user-item purchase history, a search engine-based recommendation system can be developed for users ina business. Product suggestions can be made using the textual clustering analysis provided in the product description. We used the product descriptions dataset, which contains 124428 descriptions as well as productids, for our technique.

```
"product_uid","product_description"
100001,"Not only do angles make joints stronger, they also provide more consistent, straight corners. Simpson Strong-Tie offers a wide variety of angles in various sizes and thicknesses to handle light-duty jobs or projects where a structural connection is needed. Some can be bent (skewed) to match the project. For outdoor projects or those where moisture is present, use our ZMAX zinc-coated connectors, which provide extra resistance against corrosion (look for a ""Z"" at the end of the model number).Versatile connector for various 90 connections and home repair projectsStronger than angled nailing or screw fastening aloneHelp ensure joints are consistently straight and strongDimensions: 3 in. x 3 in. x 1-1/2 in.Made from 12-Gauge steelGalvanized for extra corrosion resistanceInstall with 10d common nails or #9 x 1-1/2 in. Strong-Drive SD screws"
100002,"BEHR Premium Textured DECKOVER is an innovative solid color coating. It will bring your old, weathered wood or concrete back to life. The advanced 100% acrylic resin formula creates a durable coating for your tired and worn out deck, rejuvenating to a whole new look.  For the best results, be sure to properly prepare the surface using other applicable BEHR products displayed above.California residents: see Proposition 65 informationRevives wood and composite decks, railings, porches and boat docks, also great for concrete pool decks, patios and sidewalks100% acrylic solid color coatingResists cracking and peeling and conceals splinters and cracks up to 1/4 in.Provides a durable, mildew resistant finishCovers up to 75 sq. ft. in 2 coats per gallonCreates a textured, slip-resistant finishFor best results, prepare with the appropriate BEHR product for your wood or concrete surfaceActual paint colors may vary from on-screen and printer representationsColors available to be tinted in most storesOnline Price includes Paint Care fee in the following states: CA, CO, CT, ME, MN, OR, RI, VT"
100003,"Classic architecture meets contemporary design in the Ensemble Curve series, made of solid Vikrell material, blending sleek, clean lines with gentle curves. Corner shelving is perfect for storing bath accessories. Modular design allows it to be moved around corners and through doorways with ease. Curve wall with a smooth, contemporary look, featuring integrated storage shelves.Slightly narrower for tighter spacesDesigned with an 18 in. apronDurable high-gloss finish provides a smooth, shiny surface that is easy to cleanConforms to ANSI Z124.1.2 and CSA B45.1 national consensus standardsCurve wall with a smooth, contemporary look, featuring integrated storage shelves"
100004,"The Grape Solar 265-Watt Polycrystalline PV Solar Panel bonus pack bundles 4 Grape Solar 265-Watt solar panels for extra savings. The Grape Solar 265-Watt Polycrystalline PV Solar Panel uses high efficiency solar cells (approximately 19%) made from quality silicon material for high module conversion efficiency, long term output stability, and reliability. Virtually maintenance free. High transmittance, low iron tempered glass for durability and enhanced impact resistance.Positive power tolerance (0 to +5-Watt)Anti-reflective and anti-soiling surface reduces power loss from dirt and dustOutstanding performance in low-light irradiances environmentsCertified to withstand high wind and snow loads100% electroluminescence inspection ensures modules are defect freePositive and negative leads quipped with MC4 connectors"
```

Fig 2: The product description dataset with first 5 values

TfidfVectorizer from the sklearn library was used to extract features from product descriptions. Based on the product description, we utilized K-Means clustering to evaluate the data and find the top keywords in each cluster. If a word appears in more than one cluster, the algorithm chooses the cluster with the highest occurrence frequency.

The recommendation system can present items from the respective clusters based on the product descriptions once they are recognized based on the user's search phrases.

3. RESULTS

This section will present the outcomes of the three various strategies.

A. Popularity *based Recommendation System:*

The decreasing order of the dataset and the barplot for the popularity-based recommendation system are presentedbelow.

ProductId	Rating
B001MA0QY2	7533
B0009V1YR8	2869
B0043OYFKU	2477
B0000YUXI0	2143
B003V265QW	2088
B000ZMBSPE	2041
B003BQ6QXK	1918
B004OHQR1Q	1885
B00121UVU0	1838
B000FS05VG	1589

Fig 3: The descending order of ratings

<AxesSubplot:xlabel='ProductId'>

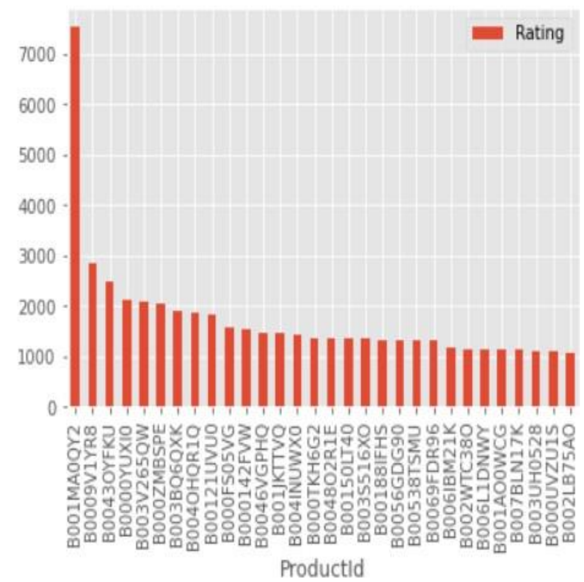

Fig 4: The popularity of the products in increasing value

The graph above shows the most popular goods sold by the company, in descending order.

B. *Model-based Recommendation System:*

The utility matrix obtained is as follows.

ProductId UserId	0205616461	0558925278	0733001998	0737104473	0762451459	1304139212	1304139220	130414089X	130414643X	1304146537	...
A00205921JHJK5X9LNP42	0	0	0	0	0	0	0	0	0	0	...
A024581134CV80ZBLIZTZ	0	0	0	0	0	0	0	0	0	0	...
A03056581JJIOL5FSKJY7	0	0	0	0	0	0	0	0	0	0	...
A03099101ZRK4K607JVHH	0	0	0	0	0	0	0	0	0	0	...
A0505229A7NSH3FRXRR4	0	0	0	0	0	0	0	0	0	0	...

5 rows × 886 columns

Fig 5: The utility matrix

The utility matrix that has been transposed by the model- based recommendation system is shown as follows.

UserId ProductId	A00205921JHJK5X9LNP42	A024581134CV80ZBLIZTZ	A03056581JJIOL5FSKJY7	A03099101ZRK4K607JVHH	A0505229A7NSH3FRXRR4	A05492663T95K
0205616461	0	0	0	0	0	
0558925278	0	0	0	0	0	
0733001998	0	0	0	0	0	
0737104473	0	0	0	0	0	
0762451459	0	0	0	0	0	

5 rows × 9697 columns

Fig 6: The transposed utility matrix

Assuming that a consumer purchased a product with the productId #6117036094, the top 10 productIds listed for thiscustomer are as follows.

```
Out[52]: ['1304482634',
          '322700075X',
          '360211600X',
          '4057362886',
          '4057363823',
          '5357955751',
          '5357955832',
          '5357955972',
          '5357956014']
```

Fig 7: The recommended items for the customer who has bought the product #6117036094.

C. *Description-based Recommendation System:*

For the technique of Description-based Recommendation System, the visualization of product clusters using K-Means is shown here.

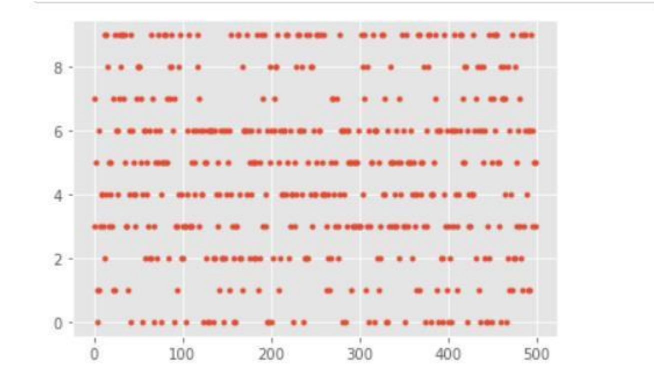

Fig 8: Visualising product clusters

The recommendations of cluster when we tried to give somewords as input for this technique are shown below.

```
In [59]: print("Cluster ID:")
         Y = vectorizer.transform(["fan"])
         prediction = model.predict(Y)
         print(prediction)

         Cluster ID:
         [11]
```

Fig 9: Output for the key word *fan*

The words present in cluster ID 11 are as below.

```
Cluster 11:
 air
 window
 cooling
 control
 power
 fan
 room
 filter
 ft
 snow
 blower
 unit
 installation
 shades
 cool
 helps
 conditioner
 000
```

Fig 10: Cluster 11

```
In [60]: print("Cluster ID:")
         Y = vectorizer.transform(["spray paint"])
         prediction = model.predict(Y)
         print(prediction)

         Cluster ID:
         [5]
```

Fig 11: Output for the key word *spray paint*

The words present in cluster ID 11 are as below.

```
Cluster 5:
 concrete
 water
 use
 paint
 metal
 brush
 ft
 seal
 provides
 coating
 garage
 watering
 spray
 easy
 cement
 coverage
 proposition
 nbsp
 residents
```

Fig 12: Cluster 5

4. CONCLUSION AND FUTURE ENHANCEMENTS

The variety of products available on e-commerce platforms is a factor to consider because it might impact a user's decision to purchase a product. As a result, the items on offermust cater to the needs of the users.

This article presents Recommendation System solutions for firms who are establishing their first e-commerce website and do not yet have any user-item purchase/rating history.

This recommendation system will aid consumers in obtaining a good suggestion, and after the buyers have a purchase history, the recommendation engine will be able to apply the model-based collaborative filtering approach.

We have implemented all these three techniques and future work may extend to achieve the same through the perfect accuracies by combining this with sentimental analysis to find the positive and negative reviews given the users to recommend a product which may give rise to the sale of a quality product.

REFERENCES

[1] Sahoo, Abhaya K., Chittaranjan Pradhan, Rabindra K. Barik, and Harishchandra Dubey. 2019. "DeepReco: Deep Learning Based Health Recommender System Using Collaborative Filtering" Computation 7,no.2:25. https://doi.org/10.3390/computation7020025

[2] K. Chopra, P. Moun and Kapil, "Smart transport recommender system," 2017 2nd International Conference on Telecommunication and Networks (TEL- NET), 2017, pp. 1-5, doi: 10.1109/TEL-NET.2017.834356

[3] Sapna Jaiswal1,, Tejaswi Kharade, Nikita Kotambe, and Dr.Shilpa Shinde. 2020. "Collaborative Recommendation System For Agriculture Sector" ITM Web of Conferences 32, 03034 (2020). https://doi.org/10.1051/itmconf/20203203034

[4] Katore L.S., and Umale J.S. "Comparative Study of Recommendation Algorithms and Systems using WEKA", International Journal of Computer Applications, Volume 110 – No. 3, pp14-17, 2015

[5] Y. Koren, R. Bell and C. Volinsky, "Matrix Factorization Techniques for Recommender Systems," in Computer, vol. 42, no. 8, pp. 30-37, Aug. 2009, doi: 10.1109/MC.2009.263.

[6] Pedro G. Campos, Alejandro Bellogín, Fernando Díez, and J. Enrique Chavarriaga 2010 " Simple time-biased KNN-based recommendations", In Proceedings of the Workshop on Context-Aware Movie Recommendation(CAMRa '10), Association for Computing Machinery,New York, NY, USA, 20–23. DOI:https://doi.org/10.1145/1869652.1869655

[7] B. Xiao and I. Benbasat, "Research on the use, characteristics, and impact of e-commerce product recommendation agents: A review and update for 2007– 2012," in Handbook of Strategic e Business Management: Springer, 2014, pp. 403-431.

[8] J . Lu, D. Wu, M. Mao, W. Wang, and G. Zhang, "Recommender system application developments: a survey," Decision Support Systems, vol. 74, pp. 12-32, 2015

[9] Zeshan Fayyaz, Mahsa Ebrahimian, Dina Nawara, Ahmed Ibrahim and Rasha Kashef 2020 "Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities" in MDPI.

[10] Jatin Sharma et al 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1022 012021

[11] Singh, Pradeep & Dutta Pramanik, Pijush & Dey, Avick & Choudhury, Prasenjit. (2021). "Recommender Systems: An Overview, Research Trends, and Future Directions. International Journal of Business and Systems Research." 15. 14–52.

[12] P. M. Alamdari, N. J. Navimipour, M. Hosseinzadeh, A. A. Safaei and A. Darwesh, "A Systematic Study

on the Recommender Systems in the E-Commerce," in IEEE Access, vol. 8, pp. 115694-115716, 2020, doi: 10.1109/ACCESS.2020.3002803.

[13] Sohail, Shahab & siddiqui, jamshed & Ali, Rashid. (2012). "Product Recommendation Techniques for Ecommerce - past, present and future". ijarcet. 1. 219- 225.