ISSN 2277-2685
IJESR/Apr-Jun. 2024/ Vol-14/Issue-2/404-415
Thota Shiva Kumar *et. al.,* / International Journal of Engineering & Science Research

RECOMMENDATION SYSTEM FOR E-COMMERCE USING MACHINE LEARNING ALGORITHMS

Thota Shiva Kumar[1], C. Rohith Kumar[2], Kondabathula Maruthi[3], Kannam Adithya[4], Ms. K. Pooja[5]
[1,2,3,4]B.Tech Students, Department of CSE, J.B. Institute of Engineering & Technology, Hyderabad, India.
[5]Assistant Professor, Department of CSE, J.B. Institute of Engineering & Technology, Hyderabad, India.

ABSTRACT: In our day-to-day life every one of us happen to use the e-commerce websites to shop many items. This leads to an increasing diversity of consumers' demand, turning into a challenge for a retail store to provide the right products accordingly to customer preferences. These e-commerce websites use different types of recommendation systems to provide a positive shopping experience to the user. Recommender systems are a tool to cope with this challenge, through product recommendation it is possible to fulfill customers' needs and expectations, helping maintaining loyal customers while attracting new customers. In this project we have used the ratings given to a product by other users to make recommendations to the current user. We have created a matrix to represent the ratings given to the product by the different users. Next, we have created the user id and item id and provide recommendations depending on the similarity to the other users. Then we apply Cosine Similarity which is a part of machine learning to make recommendations to the user.

Keywords: collaborative, hybrid, content-based, recommendation system, machine learning.

INTRODUCTION

The history of recommendation systems dates back to the early 1990s. Since then, the recommendation systems have been undergoing continual evolution. Recommendation systems have become an essential component of online applications. This is mostly due to the extensive range of product inventories accessible on e-commerce websites, which may result in overwhelming amounts of data to be shown to the consumer. Consequently, the consumer may have challenges in locating the specific product they are seeking. At this juncture, the use of recommendation systems becomes relevant. According to Harvard Business Review, recommendation algorithms are the primary distinguishing factor between "born digital" organizations and traditional legacy companies. The primary reason for this is because the recommendation engine may provide tailored recommendations of items to consumers based on their browsing behavior and past purchasing records. By offering individualized suggestions, our technology creates an engaging environment that significantly increases user interest and encourages repeat purchases. By increasing the frequency of user visits to our website, we may get a larger amount of data about the user and the items. This, in turn, offers an opportunity to identify areas for product development. The higher the quality of our goods, the more likely we are to attract a larger number of people. This is a complete, uninterrupted, and repetitive cycle.

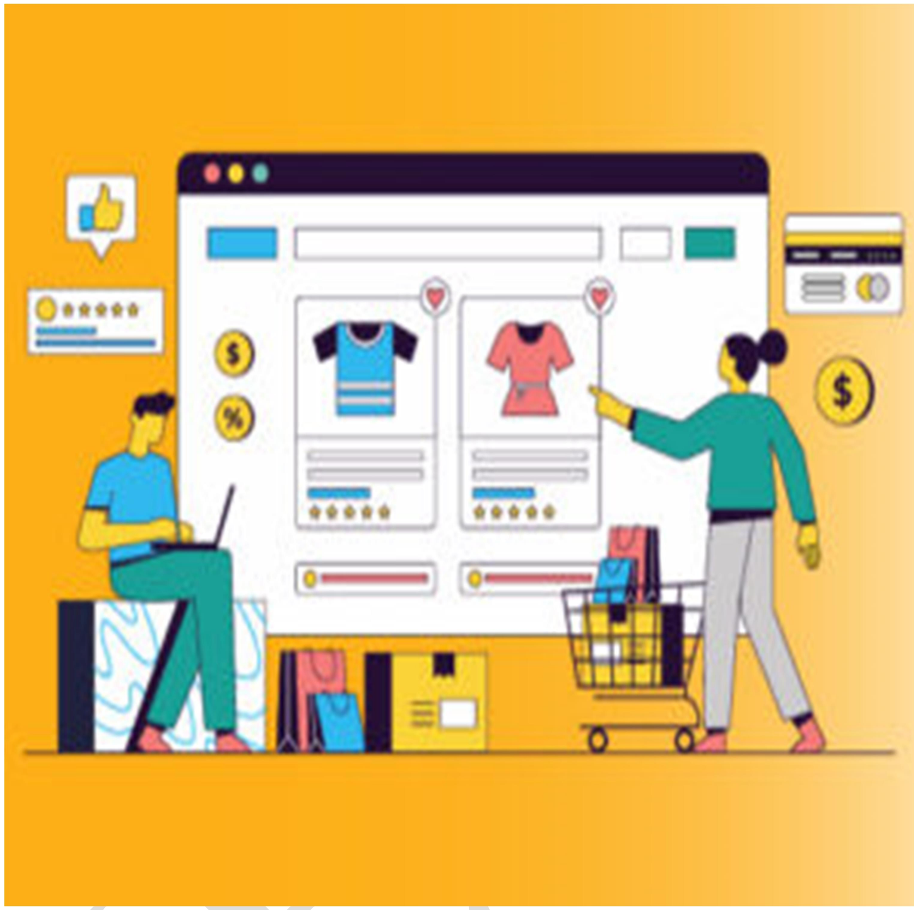


Fig 1.1 Product Recommendation

The recommendation system helps to increase the sales of items related to a particular product too. For example, if a user buys a monitor from the website, our recommendation system will suggest him different types of keyboards and mouse. In this way it helps in increasing the sales of certain items which the user by himself may not be searching. This recommendation system reduces the load on the database as they provide a personal recommendation to every individual user instead of displaying whole inventory. These recommendation systems usually are based on two different filtering techniques namely Collaborative filtering and Content based filtering. content based filtering is based on the content i.e. item-item relationship. In this method we form a relationship or cluster between group of items and display the relevant items to the customer when the customer searches for one item from the cluster. Collaborative filtering is based on the user behaviour. Depending on the use behaviour we do the further recommendations.

PROBLEM STATEMENT

Using only content-based filtering or only collaborative filtering for product recommendations can present a variety of challenges and limitations:

Content-Based Filtering

- Limited Data: Content-based filtering relies on information about the features of items (e.g., description, genre, attributes). If there is insufficient data about the features of an item, recommendations may be less effective.

- Narrow Recommendations: This method can lead to recommendations that are too similar to what the user has already seen. This can limit the diversity of recommendations and may cause user fatigue.
- New Item Cold Start: Content-based filtering may struggle with recommending new items if there is no historical data available for them.

Collaborative Filtering

- Sparsity: Collaborative filtering relies on user interaction data such as ratings or purchases. If the data is sparse, it can be difficult to find meaningful patterns, leading to less accurate recommendations.
- Cold Start: Collaborative filtering can struggle with recommending items to new users who have not yet interacted with many items.
- Bias Toward Popular Items: The method may Favor items with more interactions, potentially biasing recommendations toward popular items and ignoring niche interests.

LITERATURE SURVEY

Content, collaborative filtering, demographic, etc. are recommendation methods. Those recommendation methods and machine learning may improve movie recommendations. Different machine learning approaches may be applied in the recommendation system to improve user suggestions. Different recommendation methods have pros and cons that might affect system accuracy and efficacy.

M Viswa Murali, Vishnu T G, et al. [1] developed a collaborative filtering-based recommender system for emerging research trends. The suggested recommender system uses datasets, user prediction ratings, and cosine similarity. User accuracy is determined by the number of correct ratings. Cosine similarity orders the results. Ramni Harbir Singh, Sargam Maurya, et al. [2] used cosine similarity and KNN to propose movies. This research proposes a technique that suggests films based on popularity and genre. Deep learning is used to build the Content-Based Recommender System. This research also shows content-based recommendation system challenges and our solutions.

KNN and cosine similarity were used to produce a recommendation system by Shivganga Gavhane, Jayesh Patil, and others [3]. They developed machine learning-based technologies to understand customer needs and propose products. This study compares machine learning methods to recommend user product buying habits and improve search results.

In this study, Shubham Pawar, Pritesh Patne, and others constructed a cosine similarity-based recommendation system. As well as movie suggestions, the system provides movie information. Additional information includes the film's rating, debut date, actors, and genres. The system also provides cast data. The technology also analyzes movie reviews and categorizes them as "Good" or "Bad," saving users time.

Chen et al. [5] used CCAM (co-clustering with augmented matrices) to develop a recommendation system using heuristic scoring, conventional classification, machine learning, and content-based hybrid recommendation systems in conjunction with collaborative filtering models.

Zhou et al. devised the collaborative filtering-based ALS Algorithm for the Netflix Prize. ALS addresses huge dataset scalability. The ALS algorithm was used to estimate user ratings for this movie recommendation system. The Restricted Boltzmann Machine (RBM) has not been improved, hence this system cannot enhance outcomes. Tiantian He et al. [7] developed a graph clustering approach that employs contextual correlation to detect multiview vertex groupings. Before this paradigm, techniques concentrated on single-view attributes and overlooked contextual links. Their technique uses graph clustering and multiview learning to cluster the multiview highlighted graph. The model's unsupervised learning base prevents vertex embedding for attributed graphs.

Wu et al.[8], They recommended integrating user reputation and recommendation systems. Online recommendation systems generate suggestions based on user interests or kinds. Recommendation systems can promote products without verifying their quality. Suggest reducing distorted user ratings by checking user trustworthiness and rating history. One might use the cumulative sum technique algorithm to detect prejudiced buyers. recommended collaborative filtering. They offer legitimate users more reputation value and fraudulent ones less. If this value is the same for both sorts of users, their system cannot differentiate between them.

PROPOSED SYSTEM

A suggested eCommerce recommendation system uses hybrid filtering and cosine similarity to provide consumers more tailored and diversified suggestions. A system like this could be created. The system uses content-based and collaborative filtering to maximize their capabilities. The system uses descriptions, categories, and keywords for content-based filtering. Item profiles with product characteristics are created using this data.

Purchase history, ratings, and browsing tendencies are used for collaborative filtering. From this data, the system may identify comparable users based on their actions and interests.

The method estimates cosine similarity between item profiles to enhance suggestions. Using feature vectors, this metric compares two things. Higher cosine similarity suggests comparable qualities, enabling the algorithm to offer goods that match a user's prior tastes.

Content-based and collaborative filtering results are combined in the hybrid technique. The system may employ collaborative filtering to find things that people with similar interests have enjoyed, then content-based filtering and cosine similarity to refine the list and favor items with attributes similar to the user's past purchases. The eCommerce recommendation system may better comprehend consumer preferences by combining these strategies. It uses content-based qualities and collaborative user data to recommend a variety of items to suit individual interests. This hybrid strategy may also solve new user and item cold starts by filling gaps using each method's strengths.

The suggested approach may provide more complex and tailored suggestions, improving user engagement and satisfaction.

SYSTEM DESIGN

5.1 SYSTEM ARCHITECTURE:

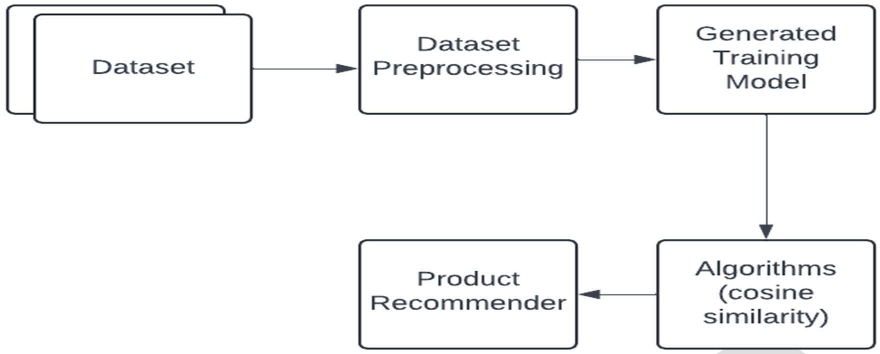
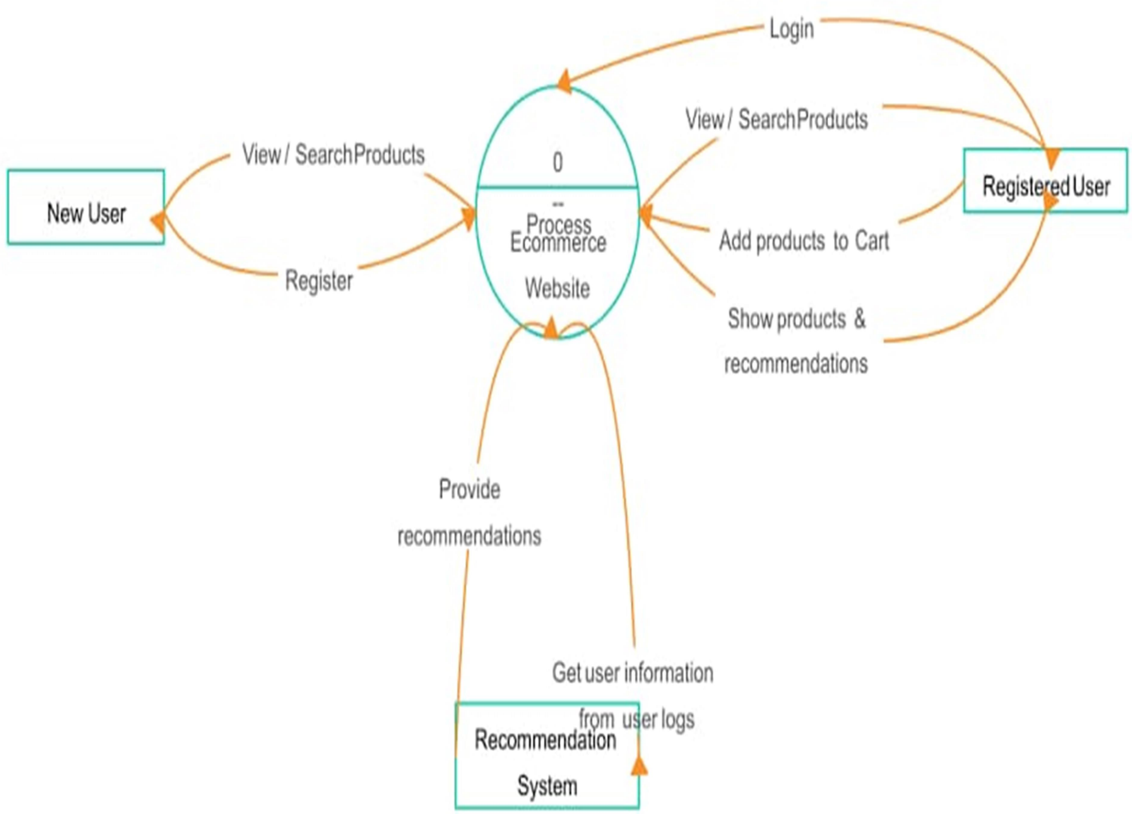


Fig 5.1 System Architecture

The dataset we used is Amazon products dataset. The raw data undergoes through the process of preprocessing. After pre-processing the data, it generates a new file contains only necessary information. This file is used to generate the training model. The algorithms we used are cosine similarity. The algorithm predicts the similar products. The algorithm creates a similarity score between the products that helps to recommend the product.

2 Sequence diagram:



5.2.2 Sequence diagram

The sequence diagram refers to the steps performed in an e-commerce site that helps in recommending the best products and similar products to the customer.

ALGORITHMS

Cosine similarity is a fundamental concept in the realm of natural language processing, information retrieval, and machine learning. It's a measure used to determine the similarity between two vectors in a multidimensional space, often employed in tasks like text mining, document similarity analysis, and recommendation systems. Understanding cosine similarity requires delving into vector spaces, mathematical computations, and real-world applications.

At its core, cosine similarity quantifies the cosine of the angle between two vectors. Imagine each vector as a direction in space, with the magnitude of the vector representing its length or strength in that direction. When we compute the cosine similarity between two vectors, we're essentially evaluating how similar their directions are, irrespective of their magnitudes.

To grasp this concept better, let's break down the mathematical formula for cosine similarity. Given two vectors \(A \) and \(B \), the cosine similarity \(\text{similarity} \) is calculated as:

$$similarity(A, B) = \frac{A \cdot B}{|A| \cdot |B|}$$

Here, \(A \cdot B \) denotes the dot product of vectors \(A \) and \(B \), while \(\|A\| \) and \(\|B\| \) represent their respective magnitudes or norms. The dot product captures the similarity in direction, while dividing by the product of magnitudes normalizes the result, ensuring that the similarity metric ranges from -1 to 1, where 1 indicates perfect similarity, 0 indicates orthogonality (no similarity), and -1 indicates perfect dissimilarity.

One of the key advantages of cosine similarity is its robustness to changes in vector magnitude. This means that even if the vectors representing documents or text passages vary in length (due to different word counts, for instance), cosine similarity can still effectively measure their similarity based on the underlying directional alignment of words or features.

In natural language processing (NLP), cosine similarity finds extensive application in tasks like document clustering, information retrieval, and text classification. For instance, in document clustering, cosine similarity helps group similar documents together by comparing their content vectors, which can be derived from word embeddings, TF-IDF (Term Frequency-Inverse Document Frequency) representations, or other numerical encodings.

OUTPUT SCREENSHOTS

The below image shows the product id with a total rating.

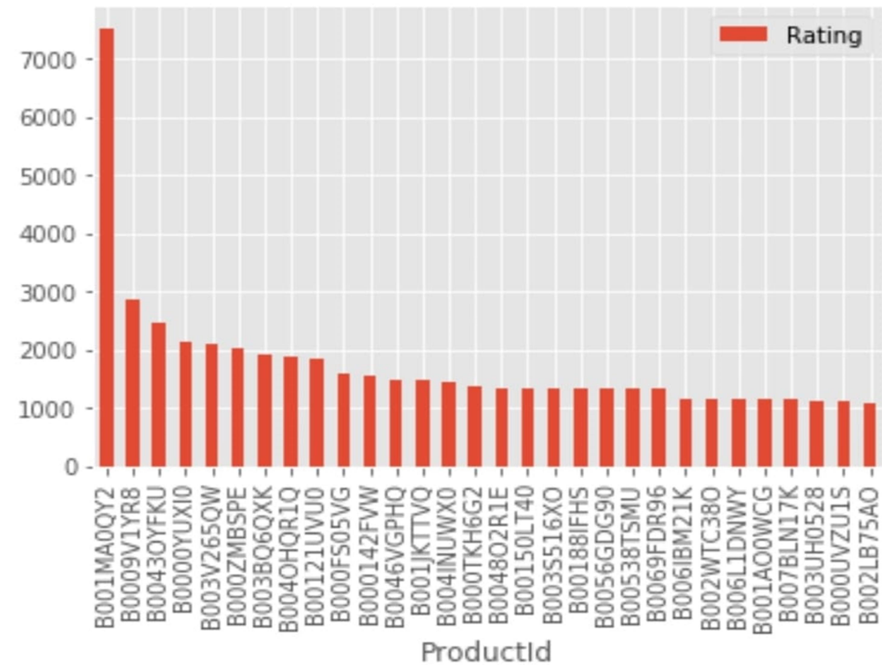


Fig: 8.1 Ratings

Below are the top 10 products to be displayed by the recommendation system to the customer based on the purchase history of other customers in the website.

```
Out[17]:

['0733001998',
 '1304139212',
 '1304139220',
 '130414089X',
 '130414643X',
 '130414674X',
 '1304174778',
 '1304174867',
 '1304174905']
```

Fig:8.2 Top 10 Recommended products

ISSN 2277-2685
IJESR/Apr-Jun. 2024/ Vol-14/Issue-2/404-415
Thota Shiva Kumar *et. al.,* / International Journal of Engineering & Science Research

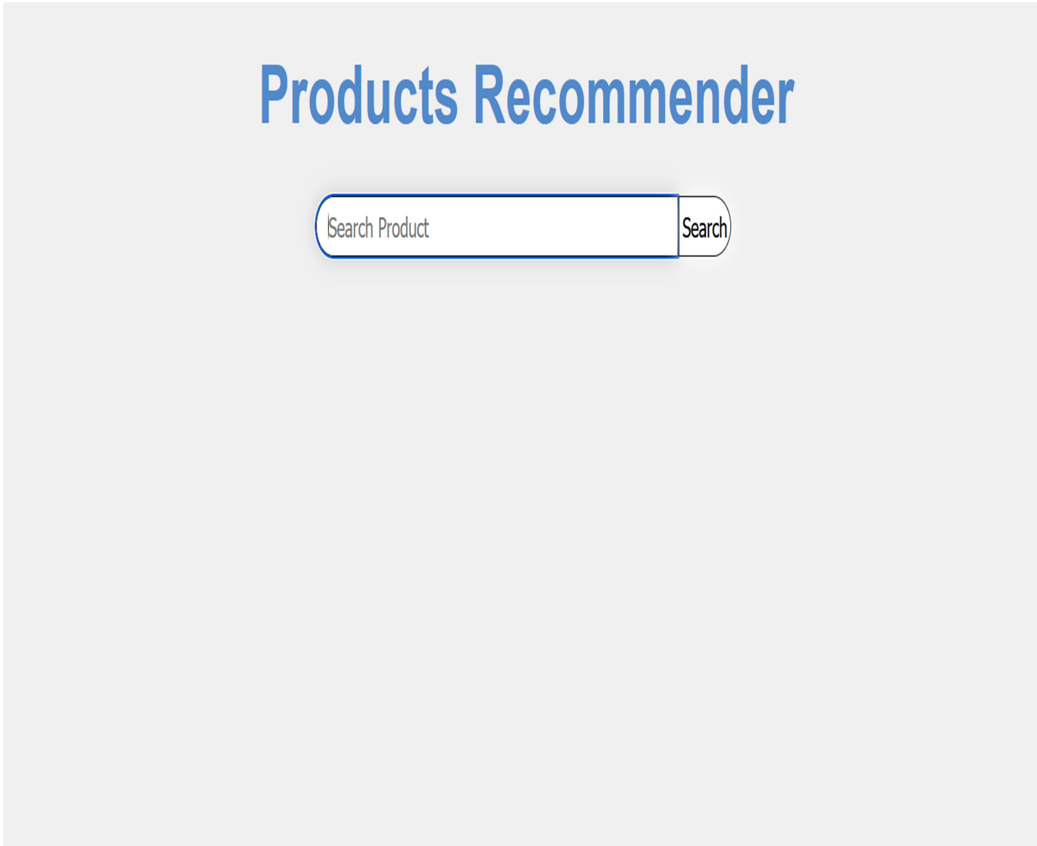


Fig: 8.3 User Interface

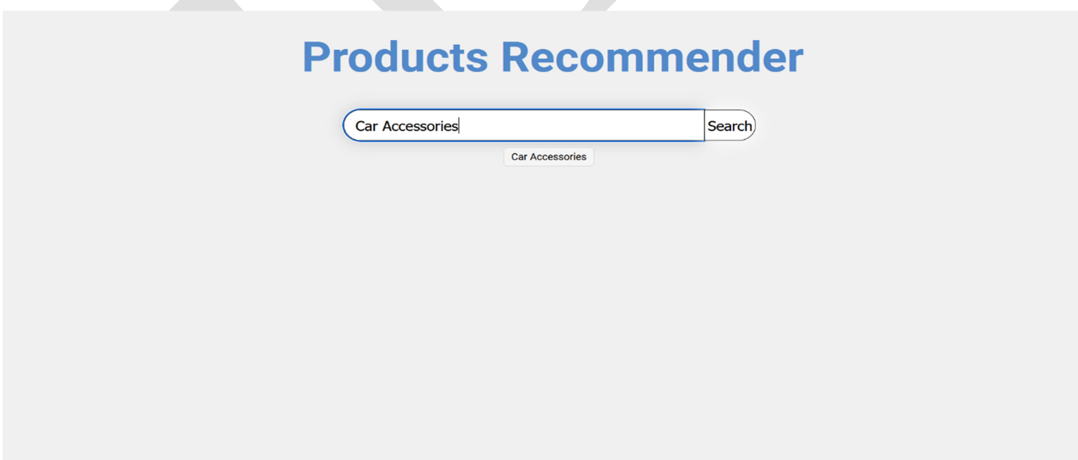


Fig: 8.4 Searching for a Product

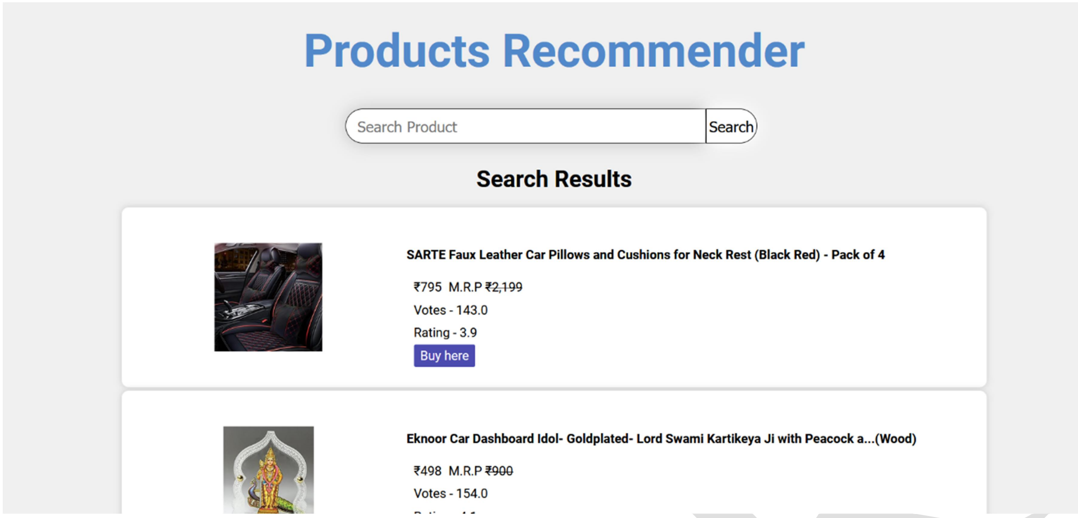


Fig 8.5 Search Results

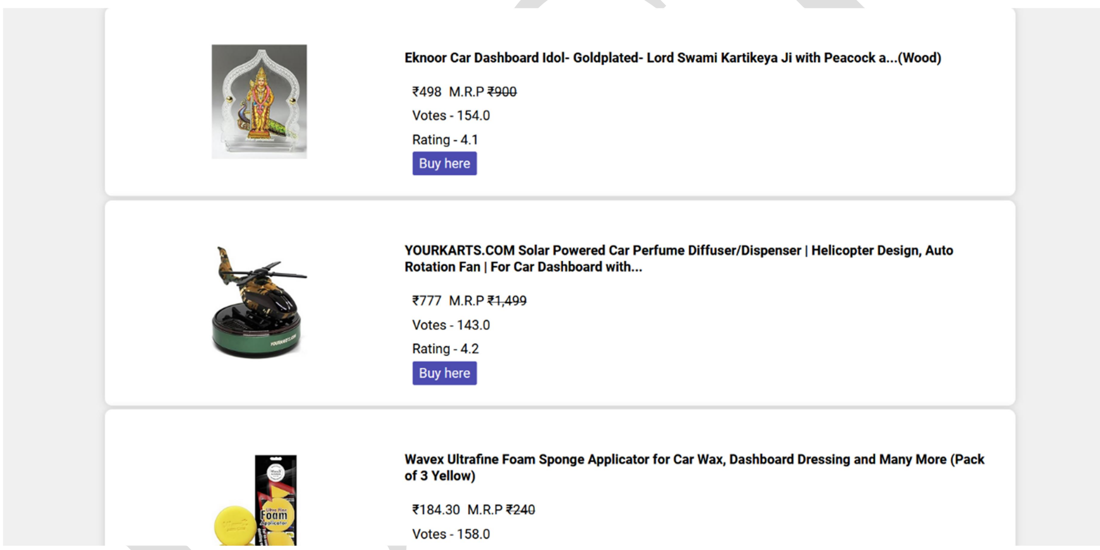


Fig: 8.6 Top 5 Search Results

CONCLUSION

E-commerce platforms may improve user experience and revenues using a hybrid product recommendation system that uses cosine similarity. Content-based and collaborative filtering are used to provide individualized and relevant product recommendations. This system relies on cosine similarity to compare products or people based on their qualities and preferences. This method uses both recommendation systems' strengths to overcome their weaknesses, creating a more robust and versatile solution.

Providing more diversified and accurate suggestions is a major benefit of a hybrid system. Content-based filtering lets the system compare product qualities and user preferences to provide suggestions. This method is excellent for proposing new or niche products that few people have reviewed. Cosine similarity quantifies item similarity, allowing the system to provide accurate suggestions even with little data.

However, collaborative filtering emphasizes user-product ties based on past encounters. This approach may recognize user activity trends and propose things comparable users liked. Collaborative filtering and cosine similarity allow the system to leverage user interactions and product ratings to provide suggestions that match user preferences.

Scalability and flexibility are further advantages of the hybrid system. By weighing diverse recommendation algorithms, the system can manage complexity as the platform and data evolve. The system may favor collaborative filtering during busy shopping seasons to capitalize on current trends and content-based filtering during calmer times to expose consumers to new goods.

Also, cosine similarity helps the system provide precise choices. The method finds the best matches by computing the cosine of the angle between vectors representing user preferences or product features. This accuracy boosts sales, customer loyalty, and user happiness.

Finally, the hybrid product recommendation system for an e-commerce site that uses cosine similarity improves purchasing. Content-based and collaborative filtering allow the system to provide tailored and diversified suggestions based on individual tastes and developing trends. The link between objects and users may be quantified using cosine similarity, enabling accurate and appropriate recommendations. This holistic approach creates a more efficient and effective recommendation system that boosts user happiness, platform loyalty, and company success.

FUTURE SCOPE

An e-commerce platform's hybrid product recommendation system uses many suggestion methods to improve accuracy and personalization. Collaborative filtering, content-based filtering, and cosine similarity may create an advanced system that knows user behavior and product characteristics. This hybrid system may be improved using many ways to increase performance and user experience.

1. Use advanced NLP:

Modern NLP algorithms may extract subtle information from product descriptions and reviews. Sentiment analysis and named entity recognition may reveal consumer views and product characteristics. This clarifies product content and user choices.

A hybrid system may use various recommendation techniques, such as collaborative filtering and content-based filtering, with dynamic weighting of algorithms. Dynamic weighing may balance these strategies depending on context or user response. Due to increased user interactions during peak seasons, collaborative filtering may be more relevant, whereas content-based filtering may be more relevant during new product launches.

3. User and Product Clustering: Grouping people and goods by attributes and preferences enhances recommendation accuracy. Clustering identifies people with similar preferences and interests, enabling the algorithm to offer popular goods. Content-based recommendations may be improved by grouping goods by characteristics.

4. User profile enrichment personalization:

Social media activity and browser history may offer context to user profiles and recommendations. This method helps the system adapt to user preferences and provide more tailored recommendations.

5. Real-Time suggestions: Improving user experience with real-time suggestions. The software can instantly react to changing user behavior and preferences by updating suggestions as people engage with it.

6. Multimodal Similarity:

Consider employing additional similarity metrics for photos, audio, and videos than cosine similarity for text-based comparisons. Image recognition methods like convolutional neural networks (CNNs) may uncover product visual similarities for a more holistic recommendation.

7. Explainability and Transparency: Offering suggestions with explanations boosts trust and satisfaction. The algorithm may show "Recommended because you purchased a similar item" or "Popular among users with similar interests."

8. Context-Aware Recommendations: Use user context, such as location, time of day, or current trends, to enhance relevance. This method may adjust recommendations to where and when the consumer shops. Monitoring user interactions and gathering feedback on suggestions helps enhance the system over time (9. Continuous Monitoring and Feedback Loops). The system may spot patterns, alter algorithms, and improve suggestions by examining this data.

10. Scalability and Performance Optimization: Optimize system for efficient handling of huge data and user traffic. Distributed computing and caching speed up the system.

A hybrid product recommendation system may make e-commerce buying more customized, efficient, and fun by adding these advantages. Multiple suggestion methods, real-time flexibility, context-awareness, and continual feedback will maintain the system adaptable and successful in addressing client demands.

REFERENCES

[1] M Viswa Murali, Vishnu T G, Nancy Victor," A Collaborative Filtering based Recommender System for Suggesting New Trends in Any Domain of Research",2019, (ICACCS),DOI:10.1109/ICACCS.2019.8728409

[2] Ramni Harbir Singh, Sargam Maurya, Tanisha Tripathi, Tushar Narula, Gaurav Srivastav," Movie Recommendation System using Cosine Similarity and KNN",2020,((IJEAT), DOI: 10.35940/ijeat.E9666.069520

[3] Shivganga Gavhane,Jayesh Patil,Harshal Kadwe,Projwal Thackrey,Sushovan Manna, "Recommendation System using KNN and Cosine Similarity",2020.

[4] Shubham Pawar, Pritesh Patne, Priya Ratanghayra, Simran Dadhich, Shree Jaswal, "Movies Recommendation System using Cosine Similarity", (IJISRT), Volume 7, Issue 4, April – 2022, 342-346, April 2022.

[5] Y. C Chen, "User behavior analysis and commodity recommendation for pointearning apps," In 2016 Conference on Technologies and Applications of Artificial Intelligence (TAAI). IEEE,2016.

[6] Y.H Zhou, D. Wilkinson, R. Schreiber, "Large scale parallel collaborative filtering for the Netflix prize," In Proceedings of 4th International Conference on Algorithmic Aspects in Information and Management (pp. 337–348). Shanghai: Springer,2008.

[7] Tiantian He , Yang Liu, Tobey H. Ko , Keith C. C. Chan , and Yew-Soon Ong "Contextual Correlation Preserving Multiview Featured Graph Clustering",(2019),(IEEE transactions).

[8] Zhiheng Wu,Jinglin Li,Qibo Sun,Ao Zhou,"Service recommendation with context-aware user reputation evaluation",(2017),(IEEE conf).

[9] Khamael Raqim Raheem; Israa Hadi Ali, "Content-based Recommender System Improvement using Hybrid Technique", (2020) (IEEE Xplore).

[10] Shailesh Kalkar, Prof. Pramila Chawan, "A Survey on Recommendation System based on Knowledge Graphand Machine Learning", (2022) (IRJET), Volume: 09 Issue: 06 | Jun 2022.

[11] A. A. Ewees, Mohamed Eisa, M. M. Refaat, "Comparison of cosine similarity and k-NN for automated essays scoring",(2014),(IJARCCE), DOI10.17148/IJARCCE

[12]GregLinden,BrentSmitch,JeremyYork,Amazon.com Recommendations item-to-item Collaborative Filtering.Online source: IEEE Computer Society
(https://www.cs.umd.edu/~samir/498/AmazonRecommendations.pdf)

[13]K.Yogeswara Rao,G.S.N.Murthy,S.Adhinarayana "Product Recommendation System from Users Reviews using Sentiment Analysis"Research Gate