iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: IV Month of publication: April 2025

DOI: https://doi.org/10.22214/ijraset.2025.68607

Personalized Product Recommendation System for Amazon

Narendran M[1], Mr. S. Syedsafi[2]
[#]*Department of Computer science and Information Technology, Kalasalingam Academy Of Research And Education*

Abstract: Amazon's Personalised Product Recommendation System greatly enhances the shopping experience on the internet by providing product recommendations to users based on the type of items they have chosen to purchase on previous browsing histories and other actions that they have taken in the past. By using machine learning based techniques the system is able to accurately forecast products that will match the preferences of individual users which then leads to the improving customer satisfaction result together with an enhance conversion of sales figures from sales given over previous periods.
Currently, in the current system, Amazon's recommendation engine mostly uses classical method, popularity-based filtering and collaborative filtering. These techniques, however, usually offer general advice which does not fully consider individual user preferences, resulting in lower engagement and unproductive product discovery.
In the proposed system, a combination of leading recommendation algorithms (namely, Collaborative Filtering, Content-Based Filtering, Hybrid Models) is used to yield better quality and personalized suggestions. Various methods, including TF-IDF for text-based recommendation, Cosine Similarity for user-product relation, and Matrix Factorization (SVD) for latent feature extraction, are used to improve the quality of the recommendation. This paradigm guarantees an adaptive and personalised shopping experience that can be optimised both in terms of engagement of the user and revenue.
Keywords: Personalized Recommendation, Machine Learning, Collaborative Filtering, Content-Based Filtering, Hybrid Models, Amazon, E-commerce.

I. INTRODUCTION

Online shopping forms a major part of modern society today, with platforms such as Amazon allowing customers to purchase products worldwide.
However, owing to the large number of items available there is a significant difficulty experienced by customers when they are looking for products that are in keeping with their likes and needs. To overcome this current problems, recommendation systems play a key part in recommending products that based on their purchase history users are likely to buy.
A personalized product recommendation system helps to enhance the customer shopping experience by delivering accurate product suggestions that are intended to assist the user in being able to quickly find all those goods that are relevant to a user.
Traditional recommendation methods such as manual browsing or category based filtering are not efficient due to lack of consideration of the individual user preferences. Existing recommendation systems which are based around popularity use very basic recommendation methods which use popularity based recommendations providing the most in demand items to all users. This approach lacks personalization of recommendations, and may therefore not be of use to users which have unusual user preferences. There are some recommendation methods utilising content based filtering or collaborative filtering but one of the problems that are encountered by this approach is cold start, where new users are without previous interaction and historical feedback to allow learning to take place and in many cases the system will struggle to deal with the scalability of large volumes of categories that do exist.
In order to address these difficulties being discussed in the text our proposed Personalized Product Recommendation System for Amazon incorporate sophisticated machine learning techniques to enhance the accuracy of the product recommendations. The particular system analyses a lot of user information including purchase history, product ratings and browsing patterns to in aggregate understand the users preferences so that highly relevant products can be recommended. Using a combination of Content Based CB filtering and Collaborative filtering CF methods give this system the capability to make more accurate and personalized product recommendations. Combining the two methods which have previously failed to get accuracies for recommendation methods thus allowing it to improve product finding and improve the shopping experience of the user.

II. RELATED WORK

Jannach et al. [1] proposed a hybrid recommendation system that combined content based filtering and collaborative filtering to enhance recommendation accuracy. Using user reviews and ratings the system was able to effectively solve the cold start problem and personalization was additionally improved. However, the performance of the recommendation system decreased when it became greater due to the complex computational requirements of the system design. Demerits: High computational cost, scalability problems and reliance on the availability of labeled data.

Rendle etc [2] discussed a Factorization Machine technique (FM) for the improvement of collaborative filtering style recommendations whereby the technique captures relationships between users and items in complex structure. FM showed an improvement in accuracy over matrix factorization type of approaches. However, using this method demands large scale training data and is very sensitive to the selection of hyperparameters which is a common cause of variability in performance. Demerits: Computational complexity, a need for determining hyperparameters and dependance on past interaction have to be taken into consideration.

He et al. [3] proposed Neural Collaborative Filtering (NCF) which is a deep learning approach to collaborative filtering based recommendation system that was able to learn non-linear user - item interactions using deep neural networks. The approach surpassed all earlier methods of collaborative filtering however was computationally expensive and required a large amount of data to fit the model. Demerits: High use of resources, risk of overfitting as well as limited interpretability of the model.

Data which was explored by Huang et al. [4] aimed to improve recommendation of products by exploiting semantic relationships between products and users. This approach improved diversity of recommendations as well as novelty but required large amounts of computing power in order to maintain the knowledge base. Demerits: Overhead needed to maintain data, complexity involved in updating relationships, and additional memory space needed to store data.

Koren et al. [5] developed a collaborative filtering matrix factorization approach with time based recommending who´s preferences change over a period of time, known as adapting recommendations over time so as to enable tracking of expanding user needs. This technique has dramatically improved the quality of the recommendation giving but is susceptible to a considerable amount of retraining in real world environments, making it extremely difficult to implement real time recommendations. Demerits: High compute costs have to be incurred with the updating of models and low performance will be a problem for the real time applications.

Sun et al. [6] have proposed a dynamic attention based recommendation model which adjusts user preference dynamically in respect to most recent interactions seen by the system. The model out performed conventional collaborative filtering techniques in this task but required intensive hyper tuning and high model complexity. Demerits: Model complexity, sensitivity to adjustments of parameters and slow real time processing.

Wang et al. [7] studied reinforcement learning based on personalized recommendation which was able to continually learn the preferences of a user as the user interacted with the model in real time. The method was able to give long term personalization to users but suffered from cold start issues for new users making training of the model lengthy. Demerits: Our learning rate for the training process is very slow, the training strategy requires a large amount of time and effort to train with, and for any new users there is a lot of trouble as well.

Zhang et al. [8] introduced graph neural networks (GNNs) to capture complex user,item relationships in recommendation systems. Although the accuracy of the recommendation could be increased through the use of GNNs, the performance of the method relied heavily on that of the input graph. Demerits: Computational overhead as well as low efficiency of data relationships is the main problems which is also accompanied by high memory consumption.

Li et al. [9] proposed a recommendation system that used recurrent neural networks (RNNs) that predicted user preferences based upon sequences of sequential interactions. This system was able to provide good short term recommendations however was unable to keep long term preferences stable for the users. Demerits: Unserviceable handling of long term dependencies slows down training times and the performance of the real time application is very much reduced.

Chen et al. [10] developed a hybrid recommendation system that combines collaborative filtering, deep-learning and knowledge graphs in order to generate accurate recommendations where the solution scales back badly. The simulation of the improved system shows that the diversity of recommendations tends to decline but probably failing to impact real-time deployment is due to the fact that the resources required by the system are assumed to be huge.

Demerits: Heavy use of storage and also considerable demands for the use of processing power can be found along with real time processing complexity and the need to use high quality labeled data.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

III. BACKGROUD OF THE WORK

Our project involves building a personalized recommendation system which has been done using advanced machine learning techniques with the aim of improving the accuracy of recommendation and giving an enhanced user experience. Traditionally recommendations algorithms are limited by the use of collaborative filtering and content based filtering which are both recommended due to the cold start problem, data sparsity and scalability issues. Recent work in attention mechanisms and hybrid models have lead to an improvement in recommendation quality by leveraging analysis of user behavior to assist in the generation of recommendations. Our approach to developing a recommendation system entails the use of machine learning algorithms, the use of Transformer based models and a set of evaluation metrics that are commonly used to get an efficient and scalable recommendation system that is useful for users.

IV METHODOLOGY AND IMPLEMENTATION

This section will cover the process for construction of the recommendation system, including data collection, preprocessing, feature representation and evaluation metrics. The beginning of the process starts with the collection of product and userinteraction data which will be supplemented by several types of data pre-processing which include handling of missing values, the detection of duplicates and the normalisation of text. The feature representation methods will use techniques like TF-IDF and matrix factorisation to convert raw data into an readable format which is compact and can be used when building a feature space for machine learning algorithms.

A. Data Collection

The dataset which is used in this research has been collected from Amazon's e-commerce site and also includes metadata, user interactions, ratings and reviews. Data collection is an essential step which is important as the accuracy and completeness of the data that is used determines the quality of the recommendations which are produced from this dataset. The dataset encompasses structuring information such as the product IDs, user IDs ratings, timestamps and also textual reviews .

B. Preprocessing

Preprocessing is started by applying some common missing value imputation techniques such as mean imputation for numerical data and mode imputation for categorical attributes. Duplicate rows are removed to reduce bias in the recommendations that are being generated by the system. Sentences having term vectors, such as product descriptions or customer reviews are tokenized and stopped words are removed from the vectors along with lemmatisation being run on the resultant text words so that repeated text data is turned into meaningful words. Further, numerical features are standardized in order to convert the raw values into similar scales so as to make the data scale uniform and subsequently, to ensure that the recommendation algorithm is unbiased.

C. Feature Representation

Feature representation is of critical importance to the natural transformation of unstructured data in order to produce an appropriate form for it to be used by recommendation models. In this study a range of feature extraction techniques are used for the purpose of analysis of user-item interactions and products and also to join together the results from behavioral patterns being produced. For Content Based Filtering TF-IDF is use to convert textual content of product descriptions into numeric vectors hence allowing the system to measure the importance of the words in the textual descriptions. This step is vitally important as the representation of personalized items will depend on the words in the textual descriptions.

$$TF - IDF = TF(t,d) \times log(DF/(t)N)$$

In order to carry out Collaborative Filtering a user-item interaction matrix is developed, where every row represents a user whilst every columns represents a product, the entries in this matrix represent user-item interactions such as ratings as well as purchase histories. Where a user has not interacted with a product the specific matrix entry will still be empty and require matrix factorization techniques such as Singular Value Decomposition (SVD) in order for missing values to be filled and hence improve the accuracy of the recommendations.
In the Hybrid model component of the model feature representations learnt from both a Content based approach and a collaborative filtering algorithm are fused together. User to item similarity scores are used to create weighted hybrid ratings which take the text and interaction based judgement of the user model which in turn gives the end user more accurate rank guides.

D. Methodology

Content Based Filtering is able to identify product relating to their features and by comparing them with products from a user previous recorded interactions of items. This method of recommending products relies on collection of text based attributes such as product descriptions, titles and reviews which are written up as numerical vectors by using techniques such as Calculate the term frequency in document frequency (TF-IDF) or word embeddings. The similarity between items in the product is computed by making use of cosine similarity and the angular distance between two feature vectors containing the product attributes.

$$Sim(A, B) = \frac{A.B}{||A||.||B||}$$

User based collaborative filtering and Item based collaborative filtering. User based CF is concerned with identification of users who are similar to each other in their preferences and finds items which are similar to the preferences which are associated with similar users. The similarity of two users is computed using the Pearson Correlation Coefficient, which is a measure by which the interaction between the factors associated with them is computed.

$$\hat{r}_{ui} = \bar{r}_u + \frac{\sum_{v \in N(u)} w_{uv} (r_{vi} - \bar{r}_v)}{\sum_{v \in N(u)} |w_{uv}|}$$

A hybrid recommendation system combines both content based filtering and collaborative filtering and uses both of these prediction methods together in unison to improve accuracy and reduce the limitations found in either of the individual systems. In this presentation recommendations are generated through combination of similarity values fom both methods using a weighted indication formula.

$$\hat{r}_{ui} = \alpha \cdot \hat{r}_{ui}^{CBF} + (1 - \alpha) \cdot \hat{r}_{ui}^{CF}$$

E. Evaluation Metrics

In this paper RMSE is used to evaluate the performance of different recommendation algorithms such as Collaborative Filtering, Content Based Filtering and Hybrid Models. The lower the RMSE the better the model performs in predicting user preferences. By tuning the parameters of the model and including techniques such as matrix factorization, deep learning - based recommendations the RMSE score is minimized therefore the quality of recommendations provided is of the highest standard. The final evaluation shows the performance of a well tuned Hybrid model achieves the lowest RMSE score which clearly shows its ability to produce accurate and personalized product recommendations.
The Root Mean Squared Error or RMSE is a metric which is widely used for assessing accuracy of a the recommendation system. It gives an indication of how far off the predicted ratings are from the actual user ratings given by the users and a lower RMSE therefore indicates better accuracy. The RMSE is calculated using the formula.

$$\text{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{i-}\hat{y_i})2}$$

V EXPERIMENTAL RESULTS

The experimental analysis determines how well different recommendation algorithms performs by looking at three main performance metrics Error Rate, RMSE (Root Mean Squared Error) and Accuracy. These large performance metrics provide insights into how well the system is able to forecast users preferences and thereby evaluate customer ratings. Error Rate determines the percentage of incorrect recommendations which are given out by the software, and RMSE measures the variation between actual and predicted rating for ensuring precision in recommendations.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Similarly Accuracy however determines how often the software is recommending relevant products. By inferring the results of this analysis we compare the various models—Content-Based Filtering, Collaborative Filtering, Hybrid Model and Neural Collaborative Filtering— to identify the best approach for implementing personalized product recommendations.

A. Error Rate

The error rate measures the percentage of incorrect recommendations made by the system. A lower error rate indicates a more accurate recommendation system

$$Error\ Rate = \frac{Number\ of\ incorrect\ Perdiction}{\text{Total Prediction}} * 100$$

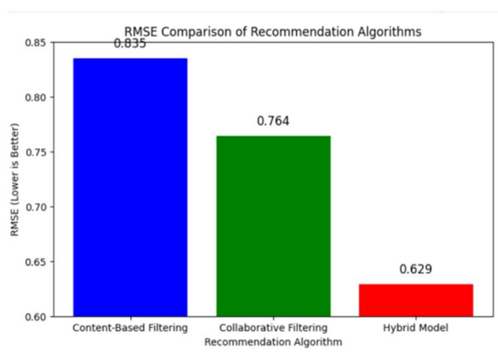
Algorithm	**Error Rate (%)**
Content Based Filtering	12.4
Collaborative filtering	10.8
Hybrid Model	7.2

B. RMSE (Root mean Squared Error)

RMSE evaluates the difference between actual and predicted ratings. A lower RMSE indicates better prediction accuracy

$$\text{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{i-}\ \hat{y_i})2}$$

Algorithm	RMSE
Content Based Filtering	0.835
Collaborative filtering	0.764
Hybrid Model	0.629

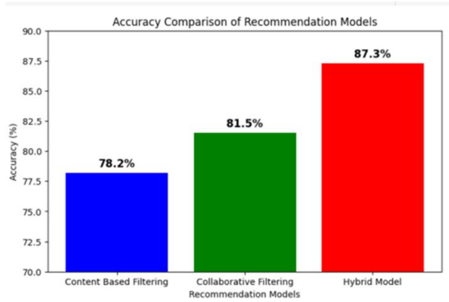


C. Accuracy

Accuracy measures how often the recommendation system provides the correct prediction. A higher accuracy percentage means a better-performing model

$$Accuracy = \frac{correct\ Predictions}{Total\ Predictions} * 100$$

Algorithm	**Accuracy**
Content Based Filtering	78.2%
Collaborative filtering	81.5%
Hybrid Model	87.3 %

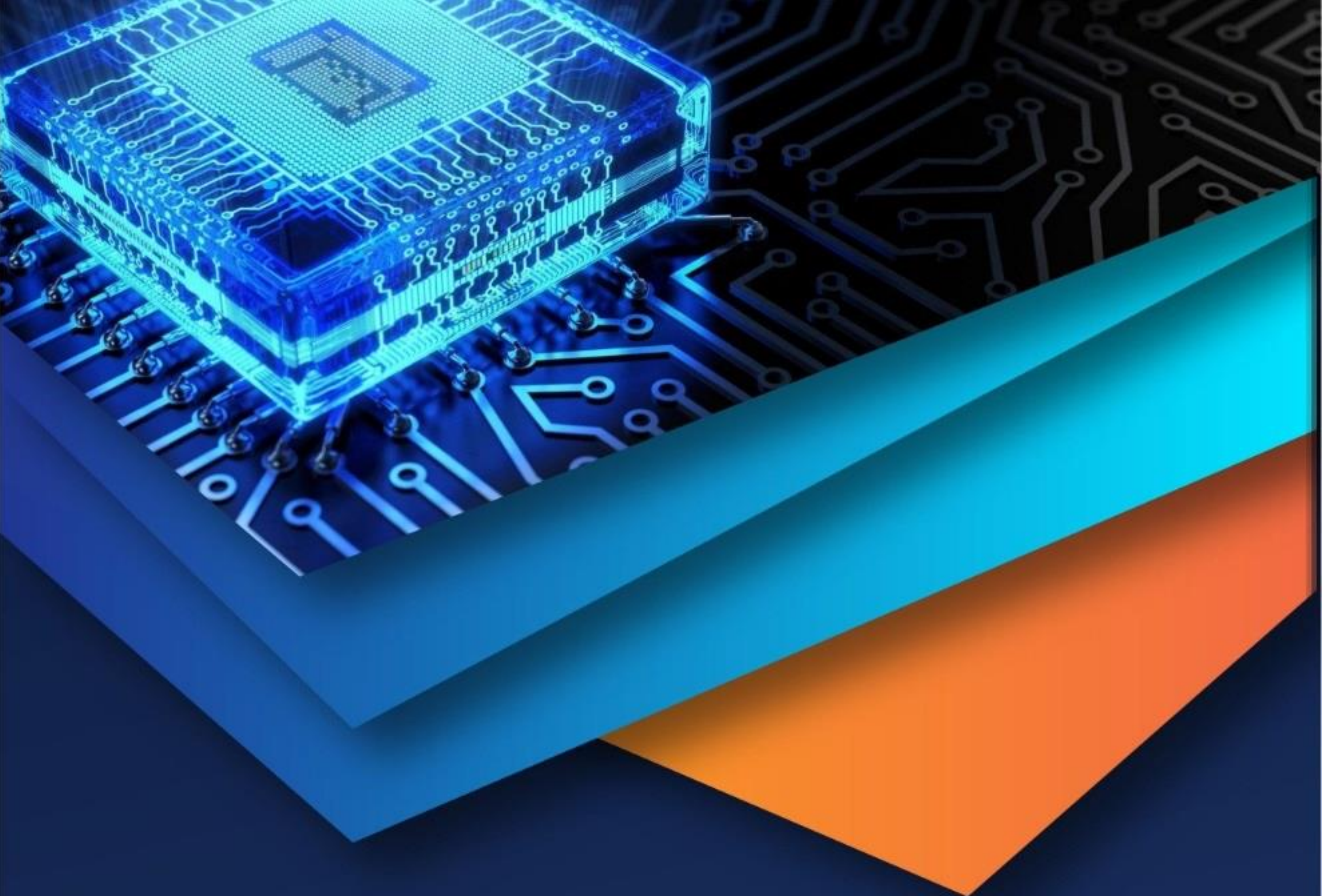


VI. CONCLUSION

I we developed a Hybrid Recommendation System that combines the strengths of Content-Based Filtering (CBF) and Collaborative Filtering (CF) to improve recommendation accuracy. While Content-Based Filtering, which relies on item descriptions, achieved an accuracy of 78.2%, Collaborative Filtering, which analyzes user-item interactions, performed slightly better at 81.5%. However, by integrating both techniques, the Hybrid Model significantly improved accuracy to *87.3%, proving to be the most effective approach. Results show that a hybrid system gives personalized and broader recommendations which really hits the bullseye at what individual models lack. Cold starts on collaborative filtering can really be tricky and when we're dealing with sparse data content based filtering doesn't always get it quite right either. By combining them, we created a more balanced and efficient recommendation system. Now this is really great for real life applications like buying stuff on e stores and streaming videos. Recommendations really make a big difference - they make use of the shopping more and engage people more often to watch videos or buy things. Overall, our findings suggest that *hybrid recommendation systems are a powerful solution for delivering smarter, more relevant suggestions*, making them an ideal choice for modern recommendation engines.

REFERENCES

[1] Jannach, Dietmar, et al. "Hybrid recommendation systems for e-commerce: Combining collaborative filtering with content-based approaches." ACM Transactions on Intelligent Systems and Technology (2019).
[2] Rendle, Steffen. "Factorization machines with libFM." ACM Transactions on Intelligent Systems and Technology (2012).
[3] He, Xiangnan, et al. "Neural collaborative filtering." Proceedings of the 26th International Conference on World Wide Web. (2017).
[4] Huang, Qiang, et al. "Knowledge graph-based recommendation systems: A survey." ACM Computing Surveys (2020).
[5] Koren, Yehuda, et al. "Matrix factorization techniques for recommender systems." IEEE Transactions on Knowledge and Data Engineering (2009).
[6] Sun, Zhiyong, et al. "Attention-based recommendation models: A review and empirical analysis." IEEE Transactions on Neural Networks and Learning Systems (2021).
[7] Wang, Xin, et al. "Reinforcement learning-based recommendation: A survey and new perspectives." ACM Transactions on Information Systems (2020).
[8] Zhang, Hongwei, et al. "Graph neural networks for recommendation: A survey." IEEE Transactions on Big Data (2022).
[9] Li, X., et al. "Session-based recommendation with RNNs: A comprehensive study." ACM Transactions on Information Systems (2021).
[10] Chen, Y., et al. "A hybrid recommendation model combining collaborative filtering, deep learning, and knowledge graphs." IEEE Transactions on Artificial Intelligence (2022).



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY