Product Recommendations using Data Mining and Machine Learning Algorithm

Patel Jaimini[1] , Jadav Dhaval [2]
[1]Assistant Professor, [2] Student
[1,2]Department of Computer Engineering
[1,2]Kalol Institute of Technology and Research Center, Kalol, Gandhinagar, Gujarat, India.

Abstract **-** Products and services are mostly bought by the user by looking at the reviews and ratings. The recommender is an important analysis for traders and buyers which takes multiple criteria in considerations while generating it. Recommender systems use algorithms from data mining to date. Nowadays recommender to provide users with product or service recommendations have been using machine learning algorithms from the field of artificial intelligence too. However, selecting an appropriate machine learning algorithm for a recommender system is difficult because of the number of algorithms. Also, the development of a recommender system using a machine learning algorithm often has problems and open questions that must be evaluated, so where to focus research efforts?. This paper presents a simple and efficient hybrid algorithm that uses a data mining algorithm apriori in a modified way and machine learning algorithm with content-based filtering. In recommender systems to identify opportunities of minimum efforts and yield accurate results is required. The paper concludes that the hybrid algorithm is used in recommender systems because of their relative simplicity, and that requirement and design phases of recommender system development appear to offer opportunities for further research.

Index Terms**: NLP, machine learning, apriori algorithm, recommendation system, hybrid approach.**

I. INTRODUCTION

Recommender systems (RS) use artificial intelligence (AI) concepts to provide users with item recommendations.
For example, an internet bookstore could use a machine learning (ML) formula to classify books by genre and advocate alternative books to a user World Health Organization desires to shop for a book.RSS was introduced in 1992 when Tapestry, the first RS, appeared. Its authors used the term 'collaborative filtering' to seek advice from the advice activity.
This term is still used to classify RSS.

Recommender systems square measure divided into 3 main classes, betting on the knowledge that recommendations square measure primarily based on: cooperative, content-based, and hybrid filtering [2]. RSs with a cooperative approach considers the user information once process info for a recommendation. as an example, by accessing user profiles on an internet music store, the RS have access to information, like the age, the country and town, and purchased songs of all users. With this info, the system will establish users that share identical music preference and so recommend users songs that similar users bought. RSs with a content-based filtering approach bases their recommendations on the item information they need access to. As associate example, contemplate a user World Health Organization is trying to find a replacement pc in an internet store.

When the user browses a selected pc (item), the RS gathers data that pc and searchers during a info for computers that have similar attributes, like value, CPU speed, and memory capability. The results of this search is then came to the user as recommendations. The third classification describes RSs that mixes the 2 previous classifications into a hybrid filtering approach, recommending things supported the user and also the item knowledge. as an example, on a social network, a RS might suggest profiles that area unit just like the user (collaborative filtering), by examination their interests. On a second step, he system might think about the suggested profiles as things and so access their knowledge to go looking for brand new similar profiles (content-based filtering).In the end,each sets of profiles area unit came as recommendations. once employing a cooperative or a hybrid filtering approach, RSs should gather data concerning the user to base recommendations on. This activity may be done expressly or implicitly. expressly user knowledge gathering happens once users provides their data and area unit awake to it. as an example, once registering for are placement on-line service, users typically fill during a kind that asks their name, age, and email. different types of specific user knowledge gathering area unit once users categorical their preferences by rating things either employing a numerical price or with a Facebook like. Conversely, implicitly user knowledge gathering is that the activity of accessing data concerning the user in associate degree indirect method. as an example, once visiting a web store, the server at the web store exchanges messages with the user's pc, and supported that, the store's RS might understand the browser the user is victimization, in addition as its country. a lot of advanced applications monitor user click and keystroke log. Besides the common recommendation method, within which users area unit given with things that they may have an interest in, recommendations may be wiped out alternative ways.

Trust-based recommendations take into thought the trust relationship that users have between them. A trust relationship may be a link in an exceedingly social network either by an addict or following connections. Recommendations supported sure friends are value over people who don't have links. Context-aware recommendations are done supported the context that the user is inserted. A context may be a set of data concerning this state of the user, like the time at the user location (morning, afternoon, evening), or their activity (idle, running, sleeping). the quantity of context info to be processed is high, creating context-aware recommendations a difficult analysis field. Risk-aware recommendations ara set of context-aware recommendations and take into thought a context during which essential info is out there, like user very

important info. it's risk-aware as a result of a wrong call could threat the user life or cause real-life damages. Some examples ar recommending pills a user ought to take or that stocks the user can buy, sell, or invest in.

II. SYSTEM DESIGN

A Algorithm

Algorithm steps :

1. Read transaction/Product data and User data
2. Using Modified-Apriori algorithm, build association rule set
    a. Min set size = 1
    b. Min confidence – 10%
    c. Min support is controlled by double input quick form node in %
3. Translate Antecedent collections into product name concatenations
4. Translate consequent item ID into consequent product name
5. Write association rule set
6. Calculate lift value in 2d matrix of order X product details table
7. Sort the product dataset details in lift value order (descending)
8. Generate a physical model
9. Form matrix based on new rules with dataset
10. Read the data from new matrix
11. Check the size and shape of two data sets (order and product details)
12. Check the data of one table
13. Check re-ordered level data from rules matrix and save it to matrix new
14. Check order count of every product with content based filtering technique
15. Filter the product details by sorting them
16. Compare prediction to measurements
17. Output estimate of state
18. Display result.

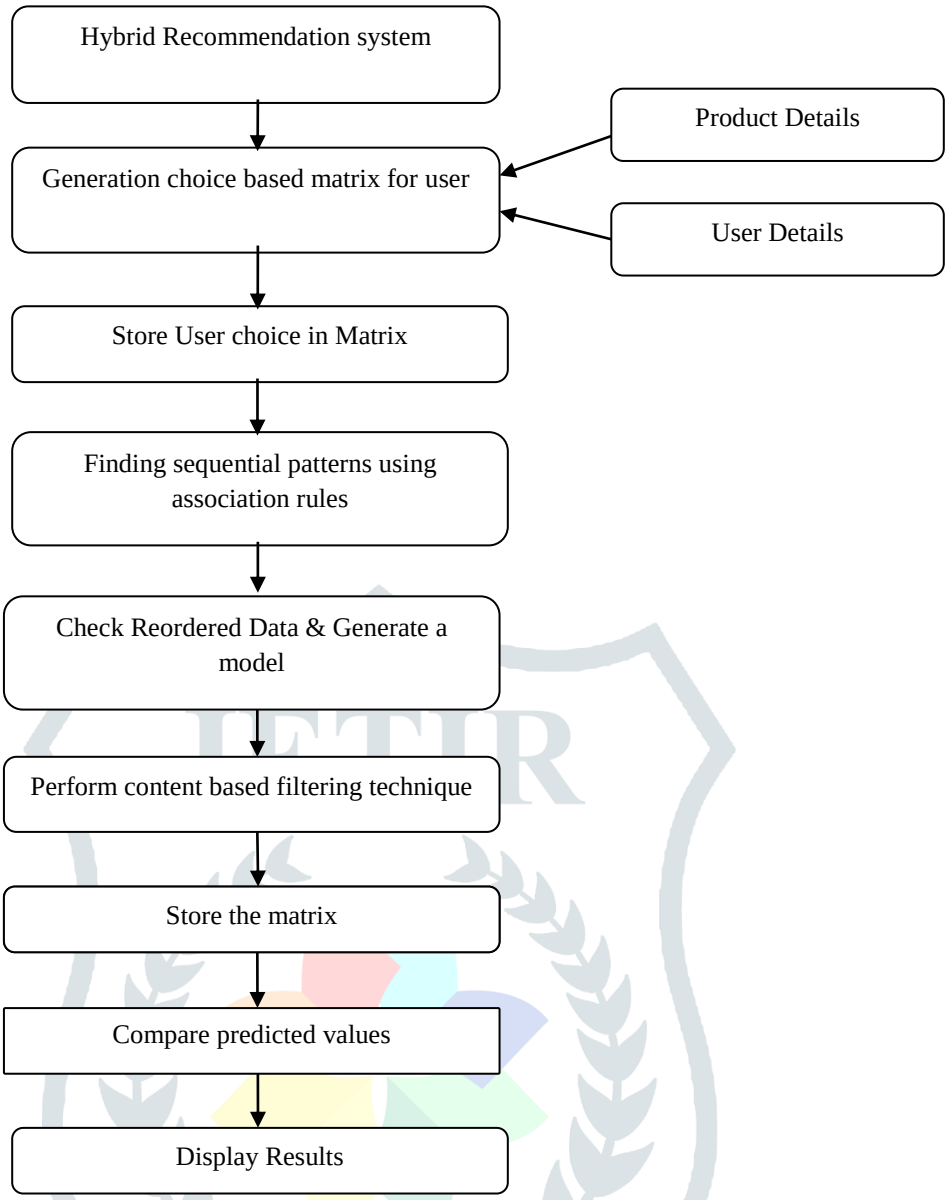


Figure 1 Flowchart

III. EXPERIMENTAL RESULT

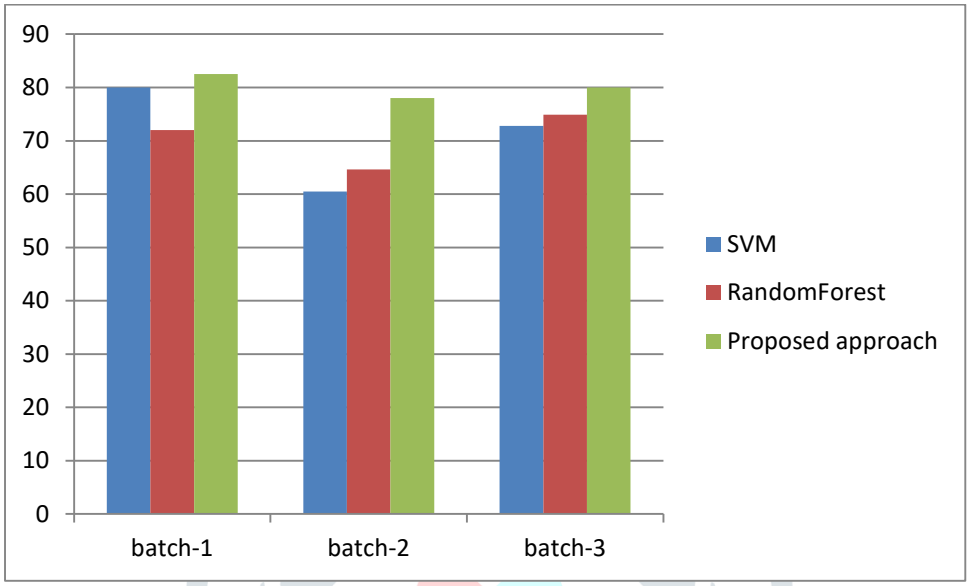
in our research, we have compared our results with existing algorithm for supervised learning approach. The results generated are better in terms of time efficiency. We have divided the database into 3 batches each batch with 1000 records from movie lens dataset. it has the same features as the movie-lens100k dataset. We have performed it over 3 times and found it less time consuming. We have also compared the accuracy in terms of percentage by comparing the results. The results founded 5 to 10% better than the earlier one.

	ML-100K	ML-1M
Number of users	943	6,040
Number of movies	1,682	3,952
Number of ratings	100,000	1,000,209
Number of all genres	19	18
Average number of genres	1.7	1.6
Rating scales	1–5	1–5

Fig 2: The result shows accuracy in percentage

algorithm	batch-1	batch-2	batch-3
SVM	80	60.5	72.8
Random Forest	72	64.6	74.9
Proposed approach	**82.5**	**78**	**80**

Fig 3: Accuracy in Percentage



Algorithm	batch-1	batch-2	batch-3
SVM	2.48	4.36	3.98
Random Forest	4.33	4.01	3.05
Proposed approach	**2.36**	**3**	**2.72**

Fig 4: Time efficiency comparison (in seconds)

IV. CONCLUSION

I have applied the hybrid algorithm by adding data mining algorithm with NLP algorithms with data mining modified Apriori with content-based filtering technique and created a new model it will generate better result than existing techniques used so far I have implemented partially algorithm which is giving results as expected In future, I will complete the implementation of the hybrid algorithm and will compare its result with state of the art technique.

V. REFERENCES

[1] Sentiment Analysis Approaches for Movie Reviews Forecasting: A Survey.
[2] Machine Learning-based Product Recommendation using Apache Spark.
[3] Product Recommendation Based on Customer Behavior using Data Mining.
[4] RECOMMENDATION WITH DATA MINING ALGORITHMS FOR E-COMMERCE AND M-COMMERCE APPLICATIONS
[5] PRODUCT RECOMMENDATIONS USING DATA MINING AND MACHINE LEARNING ALGORITHMS
[6] Ko-Jen Hsiao, Alex Kulesza, and Alfred O. Hero. 2014. Social Collaborative. IEEE Journal of Selected Topics in Signal Processing. 8(Disease: 4).
[7] D. Dubois and H.F'rade. 1990. Fuzzy Sets and System; Theory and Applications. New York: Academic.
8] Singh, Rajni, and Rajdeep Kaur. "Sentiment Analysis on Social Media and Online Review." International Journal of Computer Applications 121, no. 20 (2015).