IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 3, March 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903205|

Multiple Representation of Large Data Using ICFS Clustering

Bavana Sri Chandana G[1], Sharmila S[2], Sruti B[3], K. Balachander[4]

UG Student, Department of Computer Science and Engineering, Velammal Institute of Technology, Chennai, Tamil Nadu, India[1,2,3]

Associate Professor, Department of Computer Science and Engineering, Velammal Institute of Technology, Chennai, Tamil Nadu, India[4]

ABSTRACT: The Web creates excellent opportunities for businesses to provide personalized online services to their customers. Recommender systems aim to automatically generate personalized suggestions of products/services to customers (business or individual). Although recommender systems have been well studied, there are still two challenges in the development of a recommender system, particularly in real-world B2B e-services. In Proposed a recommendation technique utilizing the fast diffusion and information sharing capability of a large customer network. This system implemented a GRS based on opinion dynamics that considers these relationships using a smart weights matrix to drive the process. In GRSs, a recommendation is usually computed by a simple aggregation method for individual information the proposed method [described as the customer-driven recommender system (CRS)] follows the collaborative filtering (CF) principle but performs distributed and local searches for similar neighbors over a customer network in order to generate a recommendation list. Then we focuson the assignment of new arriving objects and dynamic adjustment of clusters, in incremental CFS (ICFS) clustering.

I. INTRODUCTION

Web mining or Knowledge Discovery is the process of analyzing data from different perspectives and summarizing it into useful information. This information can then be used to increase revenue, cuts costs, or both. A software created with web mining as its basic theme should allow users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, web mining is the process of finding correlations or patterns among dozens of fields in large relational databases.
Usually web mining tends to be misunderstood with terms like searching and exploration, the following table brings about the difference lucidly:-
This project is an extension of one of the popular sub-categories of web Mining: - "Market Basket Analysis (MBA)", which is a modeling technique providing insight into the customer purchasing patterns. A market basketis composed of the item-sets which are purchased in a single trip to the store. MBA basically seeks to find the relationship between the items purchased in this basket.

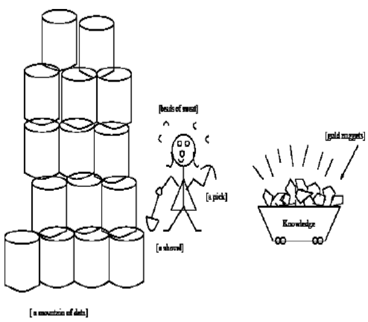

Figure 1:Data mining

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903205|

II. LITERATURE REVIEW

In the year 2014 Ruining He, Julian McAuley Modern recommender systems model people and items by discovering or 'teasing apart' the underlying dimensions that encode the properties of items and users' preferences toward them. Critically, such dimensions are uncovered based on user feedback, often in implicit form in addition, some recommender systems make use of side information, such as product attributes, temporal information, or review text. However one important feature that is typically ignored by existing personalized recommendation and ranking methods is the visual appearance of the items being considered.
In the year 2016 Matthew D. Zeiler and Rob Fergus proposed Large Convolution Network models have recently demonstrated impressive classification performance on the Image Net benchmark Krizhevskyet al. However thereis no clear understanding of why they perform so well, or how they might be improved. In this paper we explore both issues. We introduce a novel visualization technique that gives insight into the function of intermediate feature layers and the operation of the classifier. Used in a diagnostic role, these visualizations allow us to find model architectures that outperform Krizhevskyet al. on the Image Net classification benchmark.
In the year 2016 Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton Proposed and trained a large, deep convolution neural network to classify the 1.2 million high-resolution images in the Image Net LSVRC-2010 contest into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous state-of-the-art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolution layers, some of which are followed by max-poolinglayers, and three fully-connected layers with a final 1000-way softmax. To make training faster, we used non- saturating neurons and a very efficient GPU implementation of the convolution operation.
In the year 2015 Xiangnan He, Tao Chen, Min-Yen Kan, Xiao Chen proposed that Most existing collaborative ltering techniques have focused on modeling the binary relation of users to items by extracting from user ratings. Aside from users' ratings, their aliated reviews often provide the rationale for their ratings and identify what aspects of the item they cared most about. We explore the rich evidence source of aspects in user reviews to improve top-N recommendation. By extracting aspects (i.e., the specific properties of items) from textual reviews, we enrich the user item binary relation to a user item aspect ternary relation. We model the ternary relation as a heterogeneous tripartite graph, casting the recommendation task as one of vertex ranking.

III. PROPOSED METHODOLGY AND DISCUSSION

The items or user profiles often present complicated tree structures in business applications which cannot be handled by normal item similarity measures. Promising frequent item set assumes that the two thresholds minimum support and confidence doesn't change. Items which are neither bought frequently nor bought sparingly, which constitute the middle item infuse additional noise. This method will not be efficient if the transaction database turns out to be homogeneous. This type of clustering is not user controllable except for the modification of support values Fuzzy preference tree-based recommendation approach is tested and validated using an Australian business data set and the Movie Lens data set.

DRAWBACKS OF EXISTING SYSTEM:
This method will not be efficient if the transaction database turns out to be homogeneous. This type of clusteringis not user controllable except for the modification of support values. ▪ Time consuming ▪ Need more user interaction
This project aims to accomplish an optimized predicting algorithm to find the frequent items likely to be purchased by the customer. This algorithm has better running time than FUP incremental algorithm. It helps to find frequent items in a dynamically added transaction. The previous purchasing patterns of the customers information is procured, to arrive in conjunction with the purchasing mentality of particular sets of customers. Acts as a powerful predictive tool for the marketers in enhancement of their sales strategy. A step-wise elucidation of the process is as follows. Disintegrate the transaction history database into purposeful pattern separated clusters. Mapping the current customer to the best suited cluster. Sequencing of past purchases of the customers. Prediction of the purchase sequence of the current customer. Extracting the frequent item from the transactions. To adjust clustering patterns dynamically, one-time cluster adjustment strategy is employed in E_ICFSMR to improve the generality and effectiveness of ICFS. ICFS should be able to detect nonspherical and unbalanced clusters, and canautomatically adjust the number of clusters correctly .

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903205|

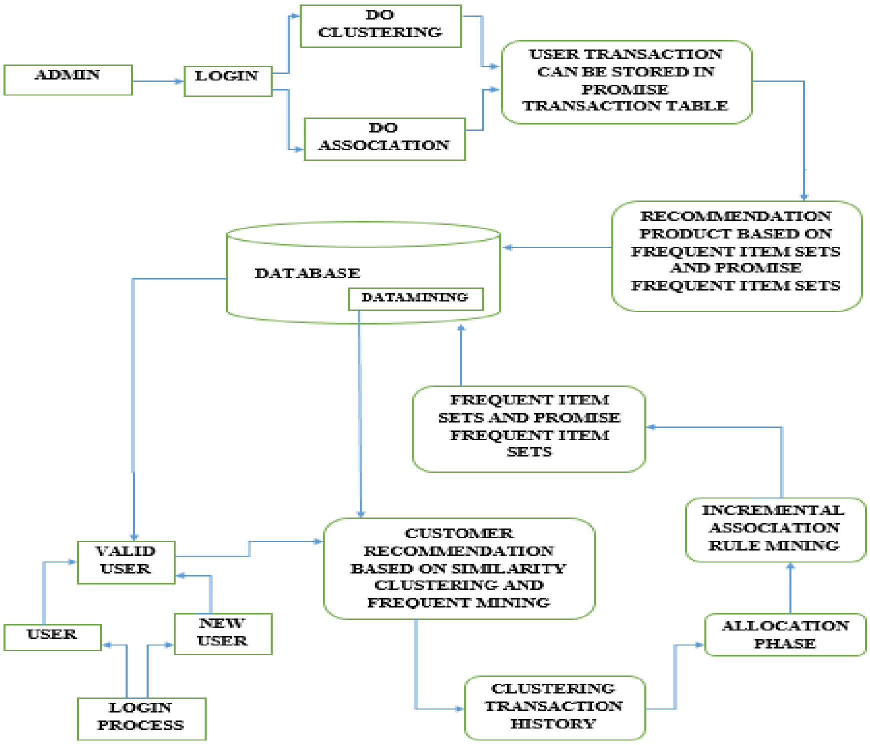

Fig 2:Architecture Diagram

User Interface:

In the industrial design field of human–machine interaction plays an important role. It is the space where interaction between humans and machines occurs. Its goal of interaction between a human and a machine at the user interface is effective operation. Input allowing the users to manipulate a system. The user will perform either login or registration operation. After this operations get over he will go to the next phase.

CLUSTERING TRANSACTION HISTORY

Input: Transaction history database Output: Clustered set of transactions
The initial phase in the process of finding the frequent item is to cluster the transaction history database. The transaction history database contains the previous transactions made by the customers. The details include customer id, the set of items bought along with the transaction id. This phase has two sub phases viz.

ALLOCATION PHASE

In the allocation phase, each transaction t is read in sequence. Each transaction t can be assigned to an existing cluster or a new cluster will be created to accommodate t for minimizing the total cost of clustering. For each transaction, the initially allocated cluster identifier is written back to the database. The decision of whether to include the transaction in one of the existing clusters or to create a new one is made by calculating the cost of clustering. The cost consists of intra-cluster dissimilarity and inter-cluster similarity which are calculated as follows.

Intra-cluster dissimilarity

Intra-cluster dissimilarity tells us how different the transactions are within a cluster.

Intra(U)= |Ukj=1 sm(Cj,E)|

where

Intra(U) – Intra cluster dissimilaritySm-small items
Cj – j th cluster

E – maximum ceiling

The maximum ceiling is the maximum number of transactions that might contain an item to call it a smallitem. Thus intra cluster dissimilarity is the union of distinct small items present in all the clusters.

Inter-cluster similarity

Inter-cluster similarity, on the other hand briefs us on the pair wise similarity between transactions present indifferent clusters. As their purpose simply, these parameters need to be kept to a minimum for the clustering to beefficient. The incoming transactions are first assigned to one of the existing clusters or a new cluster is created to accommodate the incoming transaction. The decision on whether or not to create a new cluster is based on the cost parameter i.e., a new cluster is created to accommodate the transaction if it reduces the overall cost of clustering.

REFINEMENT PHASE

In the refinement phase, the small large ratio (SL ratio) of all thetransactions are calculated as follows.
*SLR=(no. of small items)/(no. of large items)*The SL ratio of each transaction thus calculated is then compared with the SLR threshold. If the SLR of the transaction exceeds the threshold, then the transactions are moved to the excess pool. An attempt is then made to accommodate these transactions is a different cluster, if the SLR of these transactions in the new cluster doesn't exceed the threshold. If not these transactions are deemed outliers and are eliminated from consideration.
The clustering process is thus complete, incorporating both the allocation and refinement phases.

INCREMENTAL ASSOCIATION RULE MINING

Input : Transaction history database

Output: frequent itemsets and promised frequent itemsets

The transaction history database contains the previous transactions made by the customers. The details include customer id, the set of items bought along with the transaction id. This phase has two sub phases viz.,

1. Original database Discovery

2. Updating frequent and promising frequent itemsets

ORIGINAL DATABASE DISCOVERY

A dynamic database may allow insert new transactions. This may not only invalidate existing associationrules but also activate new association rules. Maintaining association rules for a dynamic database is an importantissue. Thus, a new algorithm to deal with such updating situation is proposed. Assumption for the new algorithmis that the statistics of new transactions slowly change from original transactions. According to the assumption, the statistics of old transactions, obtained from previous mining, can be utilized for approximating that of new transactions. Therefore, Support count of item sets obtained from previous mining may slightly different from support count of item sets after inserting new transactions into an original database that contains old transactions.The new algorithm uses maximum

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |
|| Volume 9, Issue 3, March 2021 ||
| DOI: 10.15680/IJIRCCE.2021.0903205|

support count of 1-itemsets obtained from previous mining to estimate infrequent item sets of an original database that will capable of being frequent item sets when new transactions are inserted into the original database. With maximum support count and maximum size of new transactions that allow insert into an original database, support count for infrequent item sets that will be qualified for frequent item sets, i.e. maniple, is shown in equation 1:

$$min\_sup_{DB} - \left(\frac{maxsupp}{total\ size}) \times inc\_size \right) \leq min\_PL < min\_sup_{DB} \quad (1)$$

wheremin_sup(DB) is minimum support count for an original database, mix-up is maximum support count of item sets, current size is a number of transaction of an original database and inc_size is a maximum number of new transactions. Here, a promising frequent item sets is defined as following definition:
A promising frequent item set is an infrequent item set that satisfies the equation 1. In this paper, apriority algorithm is applied to find all possible frequent k- item sets and promising frequent k-item sets. Apriority scans all transactions of an original database for each iteration with 2 steps processes are join and prune step. Unlike typical apriority algorithm, items in both frequent k- item sets and promising frequent k-item sets can be joined together in the join step. For a frequent item, its support count must be higher than a user-specified minimum support count threshold and for a promising frequent item, its support count must be higher than maniple but lessthan the user-specified minimum support count.

UPDATING FREQUENT AND PROMISING FREQUENT ITEMSETS

When new transactions are added to an original database, an old frequent k-item could become an infrequent k-item and an old promising frequent k-item could become a frequent k-item. This introduces new association rules and some existing association rules would become invalid. To deal with this problem, all k-items must be updated when new transactions are added to an original database. In this section, It explain how to update all old items. The size of an updated database increases when new transactions are inserted into an original database. Thus, min_PL must be recalculated in order to associate with the new size of an updated database. min_PL (update) is computed as the follows:

$$min\_PL_{DB\cup db} = min\_sup_{DB\cup db} - \left(\frac{max\,supp}{total\ size} \times inc\_size \right) \quad (3)$$

Then, If any k-item has support count greater than or equal to min_sup(DBUdb), this itemset is moved to a frequent k-item of an updated database. In the other case, if any k-item has support count less than min_sup(DBUdb) but it is greater or equal to min_PL(update) , this k-item is moved to a promise frequent itemset of an updated database. The following algorithms are developed to update frequent and promising frequent k- terms of an updated database.

Product Recommendation:

- Finally recommended items are filtered.
- Including when purchase rate or like a new item, as well as changes in the interests of othercustomers like.
- Items that interest Wish List or Shopping Cart
- Finally recommended items are provided by the customer.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903205|

IV. EXPERIMENTAL RESULTS

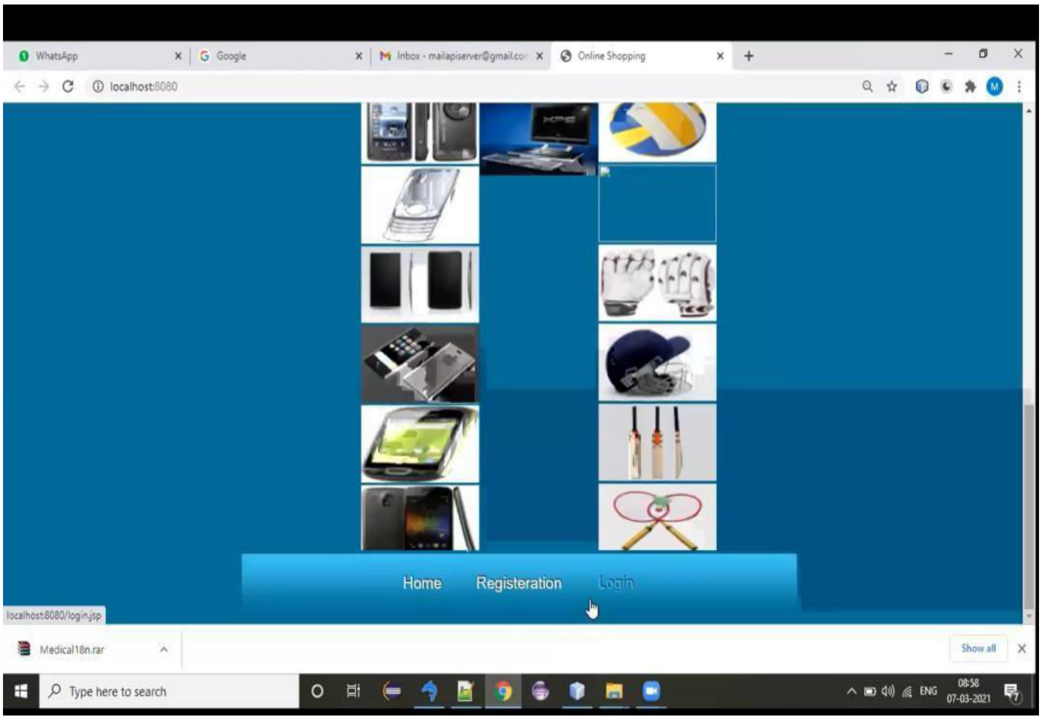

Figure 3: User will login in the web page and then user details will collected

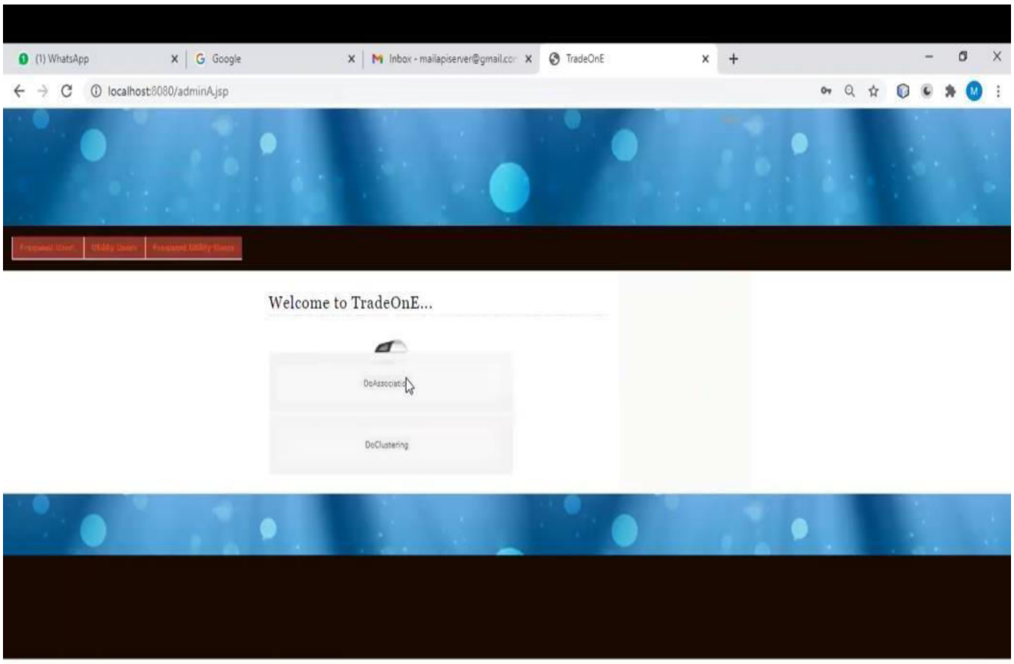

Figure 4: After user login through the account and buying the products the association is done.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903205|

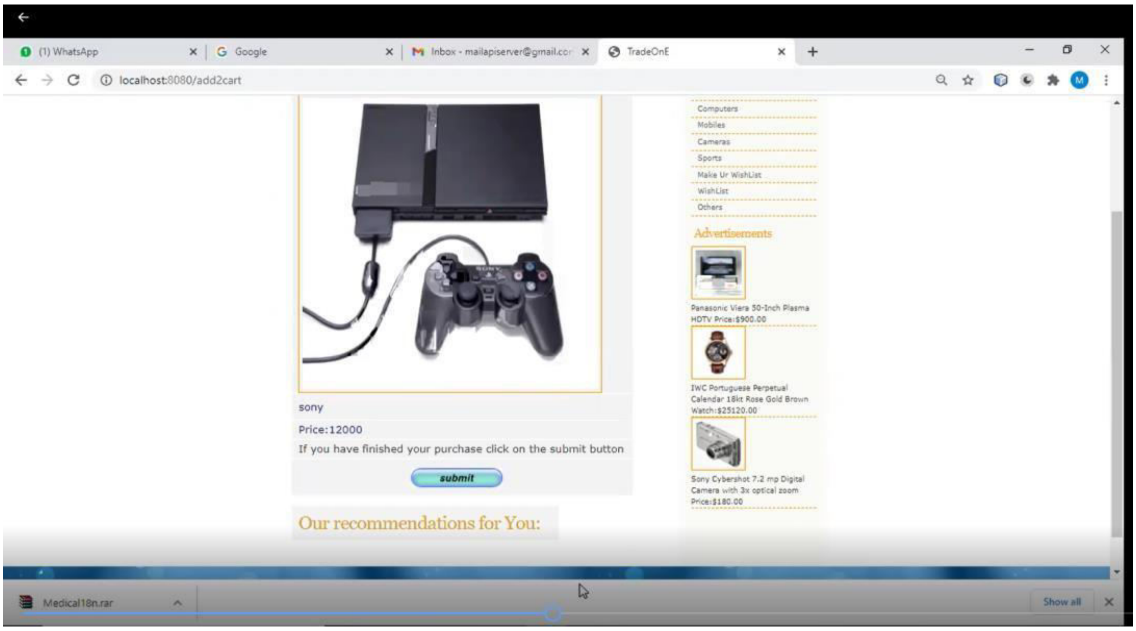

Figure 4: Viewing the products and adding to card

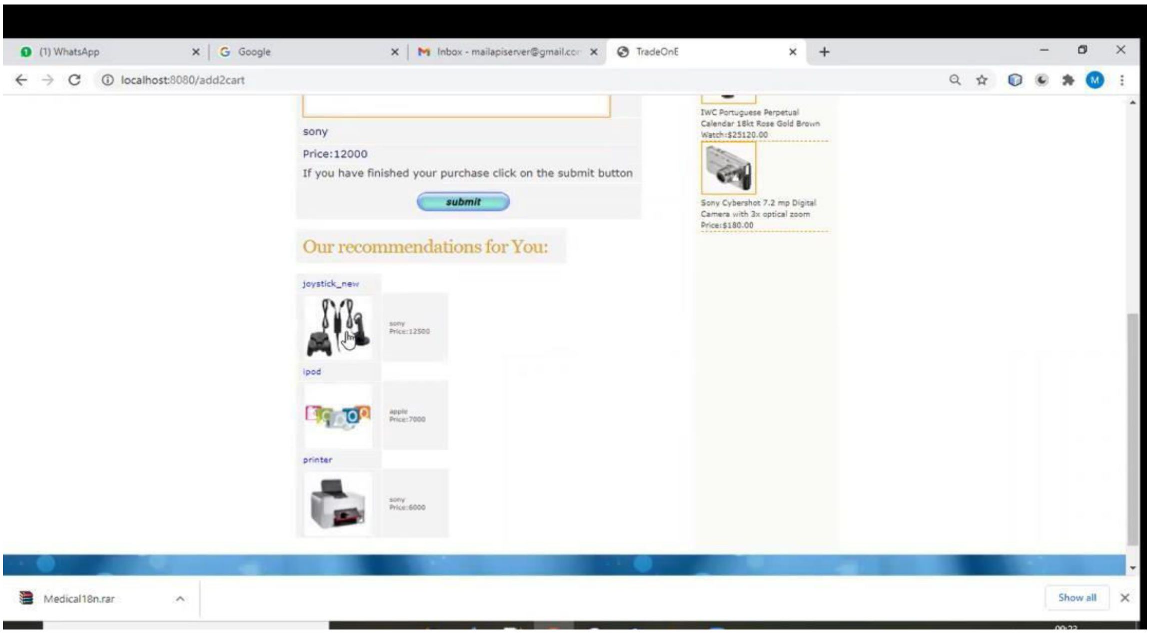

Figure 5: After buying the products recommendations will be generated.

V. CONCLUSION

With the help of Incremental Association Rule Mining and Transaction Clustering, It introduced a method to designan improved and well-structured website design for an E-shop in the design phase. Assuming that the two thresholds, minimum support and confidence, do not change, the promising frequent algorithm can guarantee to discover frequent item sets. It have used an efficient clustering algorithm for data items to minimize the SL ratio in each group. The algorithm is able to cluster the data items very efficiently. This algorithm not only incurs an execution time but also leads to the clustering results of very good quality.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903205|

VI. FUTURE ENHANCEMENTS

With the help of Incremental Association Rule Mining and Transaction Clustering, It introduced a method to designan improved and well-structured website design for an E-shop in the design phase. Assuming that the two thresholds, minimum support and confidence, do not change, the promising frequent algorithm can guarantee to discover frequent item sets. It have used an efficient clustering algorithm for data items to minimize the SL ratio in each group. The algorithm is able to cluster the data items very efficiently. This algorithm not only incurs an execution time but also leads to the clustering results of very good quality.

REFERENCES

[1] J. W. Lu, Y. Guo, Z. Q. Mi, and Y. Yang, Trust-enhanced matrix factorization using PageRank for recommender system, in Proc. 2017 Int. Conf. Computer, Information and Telecommunication Systems, Dalian, China, 2017, pp. 123– 127.

[2] G. Linden, B. Smith, and J. York, Amazon.com recommendations: Item-to-item collaborative filtering, IEEE Internet Compute., vol. 7, no. 1, pp. 76–80, 2003.

[3] Y. Koren, R. Bell, and C. Volinsky, Matrix factorization techniques for recommender systems, Computer, vol.42, no. 8, pp. 30–37, 2009.

[4]P. Vincent, H. Larochelle, Y. Bengio, and P. A. Manzagol, Extracting and composing robust features with denoising
auto encoders, in Proc. 25th Int. Conf. Machine Learning, Helsinki, Finland, 2008, pp. 1096–1103.

[5] R. N. He and J. McAuley, VBPR: Visual Bayesian personalized ranking from implicit feedback, in Proc. 30th AAAI Conf. Artificial Intelligence, Phoenix, AZ, USA, 2016, pp. 144–150.

[6] M. D. Zeiler and R. Fergus, Visualizing and understanding convolutional networks, in Proc. European Conf.Computer Vision, Zurich, Switzerland, 2014, pp. 818–833.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, Image Net classification with deep convolutional neural networks,
in Proc. 25th Int. Conf. Neural Information Processing Systems, Lake Tahoe, NV, USA, 2012, pp. 1097–1105.

[8] X. N. He, T. Chen, M. Y. Kan, and X. Chen, TriRank: Review-aware explainable recommendation by modeling aspects, in Proc. 24th ACM Int. Conf. Information and Knowledge Management, Melbourne, Australia,2015, pp. 1661–1670.

[9] S. Rendle, Factorization machines, in Proc. 10th Int. Conf. Data Mining, Sydney, Australia, 2010, pp. 995– 1000.

[10] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt Thieme, BPR: Bayesian personalized ranking from implicit feedback, in Proc. 25th Conf. Uncertainty in Artificial intelligence,Montreal,Canada,2009,pp.452-461.