International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 09 | Sep 2020 www.irjet.net p-ISSN: 2395-0072

A Real Time Application to Improvise Business Profits in E Commerce

Aparna M S[1], Dr. Srinath S[2]

[1,2]*Department of Computer Science and Engineering, JSS S&T U, Mysuru, Karnataka, India*

Abstract - *In current day large number of businesses in market are transforming from offline mode to online mode especially as an impact of COVID-19 pandemic. Business experts are discovering smart ways to enhance their sales. Products in the market are mainly classified into two groups namely frequent items and rare items considering minimum support threshold as a factor.*

In this paper, we consider concept called borderline rare items, which are available just below the minimum support threshold. These borderline rare items tend to have strong correlation with frequent items. In this paper, we have implemented two algorithms i.e. MS Apriori and Eclat algorithm to discover borderline rare items and their strong correlation with frequent items. In this paper, we have also developed a real time e-commerce application, which utilizes above two algorithms to find strong association rules. E-commerce application as three main actors, admin, guest and member. Guest as permission only to browse products whereas member as access to recommendation system which is been developed using above two algorithms. Result analysis part includes performance comparison of MS Apriori and Eclat algorithm w.r.t. transactions obtained from real time e-commerce application developed in this project.

Key Words: E-commerce, MS Apriori, Eclat, performance, comparison

1. INTRODUCTION

E-commerce applications are gaining popularity because online shopping saves both money and time. Business experts emphasis on strategically placing products in the market which can enhance their sales and resulting in profits increase. Items which are present above the minimum support threshold are frequent items and items below minimum support threshold are rare items.

Instead of concentrating on either frequent items or rare items, here in this paper we consider a group of items called as borderline rare items [7]. Border rare items are found in an area, which is between minimum support threshold and a parameter called margin. For example consider minimum support threshold as 50% and we define margin as 10%, then items whose minimum support is between 50% to 40% (40% is minimum support threshold - margin) are defined as borderline rare items. In this paper, we have implemented two algorithms, which discover borderline rare items and their strong correlation with the frequent items. MS-Apriori and Eclat algoritms are compared against their performance as the part of result analysis.

In this work, we have also developed a real time E-Commerce website, transactions are performed in this website and stored in database. Transactions from this website will be used to compare performance of algorithms, MS-Apriori and Eclat. Above two algorithms and E-commerce website have been implemented using C#, .Net framework.

2. LITERATURE SURVEY

Survey on recommendation system using data mining techniques [1] states that "recommendation systems are used online to propose items that users discover interesting, thereby, benefiting both the user and merchant". This work introduces various approaches applied by the recommendation system they are Item based, User-based and hybrid recommendation approaches. Further, it explains challenges and issues in data mining.

The work "Application of Data Mining to E-Commerce Recommendation Systems" [2] express that "data analysis has become more important because information is growing very rapid along with its complexness". This paper aims at recommending products to the user, which have more probablity of buying. This work explains different techniques available with respect to recommendation and also research on recommendation system. This paper concludes with the suggestion i.e. to obtain the better recommendation list we need to combine k-means clustering and apriori algorithm on transactional dataset.

Each and every system has its own favourable circumstances and impediments. This work [3] provides survey of applications and also provides importance of research in the area of recommendation problems and helps researchers to adopt a better algorithm strategy for recommendation considering the requirement as one factor and another factor is input sets of the system.

This work [4] provides an overview in the field of recommendation systems. It further classifies the recommendation methods into three main categories: collaborative, content-based and hybrid recommendation approaches. This paper also suggests improvement in few aspects which results in improve of recommendation capabilities, they are understanding of customers and products, adopting the contextual information by the recommendation system, also option for providing ratings

for different criterias, and a provision of more flexible and less intrusive types of recommendation.

Survey [5] classifies rare pattern mining algorithms into Apriori and tree based taking into account various constraints such as variable support threshold, without support threshold, consequent constraint based algorithms or tree algorithms.

The paper "Taxonomy of Sequential Pattern Mining Techniques" [6] describes existing sequential pattern mining algorithms, their comparative study and also research work in this domain. This work classifies the algorithms into three major categories, viz. Pattern-Growth, Early-Pruning and Apriori-based. This paper also lists applications of the algorithms in various domains.

Most of the above works classify the items into frequent items and rare items depending on the minimum support threshold. A work [7] then considers a concept called as borderline rare items, these are the items which are present just below the minimum support threshold. This paper also proposed MSD Apriori algorithm to discover border line rare items and their correlation with frequent items.

Our proposed system implements MS Apriori and also Eclat algorithm to discover borderline rare items and their strong correlation with frequent items. We also compare the efficiency of the two algorithms implemented. Some part of functionality of both the algorithms are modified in such a way that algorithms discovers borderline rare items along with frequent items and rare items. Work presented in this paper also consists of development of a realtime E-Commerce website, transactions are performed in this website and stored in database. Transactions from this website is used to compare performance of algorithms implemented in this work.

3. PROPOSED SYSTEM

Proposed system is mainly divided into three parts i.e. development of real time e-commerce application, MS Apriori algorithm and Eclat algorithm for discovery of borderline rare items and also their correlation with frequent items. Performance comparision of the two algorithms implemented in this work is graphically represented. Proposed system is developed keeping in mind the following factors, availability, reliability, scalability, security, performance and quality of service.

3.1 Architecture Design

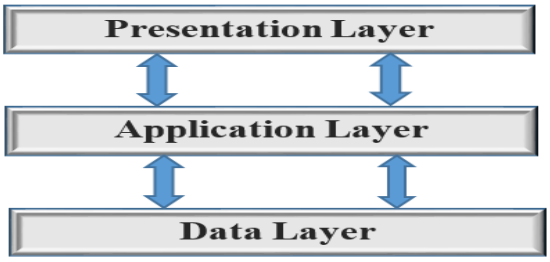

Fig -1: Three tier Architecture

Fig -1 shows the architecture adopted in this work. The presentation tier contains the UI (User Interface) which is mainly responsible for communication between the application and the user. Components of the UI should be user friendly because presentation layer is responsible of convey of results to the end user. ASP.NET Web Forms, Web User Controls and ASP.NET Master Pages are used in the development of UI.

The business tier processes the requests obtained from the presentation tier and depending on the business logic it contains(C# Classes), returns a result to the presentation tier.

The data tier is responsible for storing the application's data and forwarding it to the business tier when requested (SQL).

3.2 Database Design

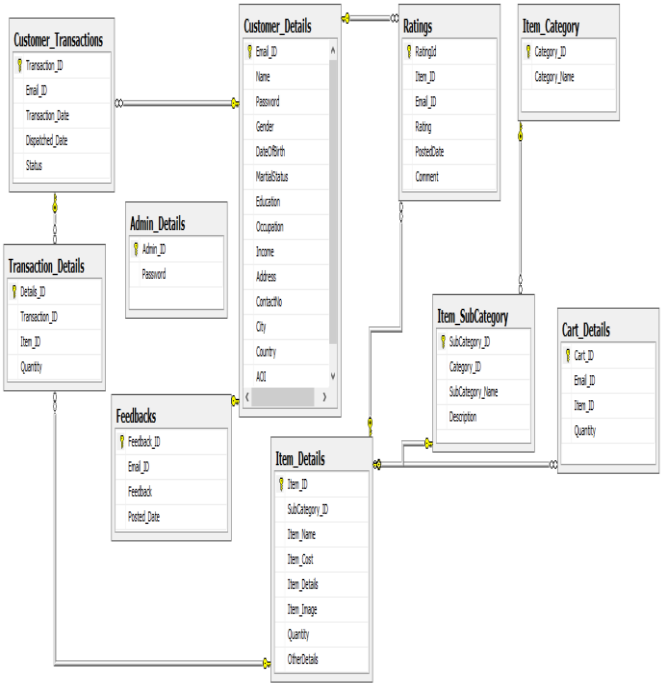

Fig -2: Database design of e-commerce application

Microsoft SQL Server is a full-featured relational database management system (RDBMS) that facilitate different kinds of administrative tools, which makes database development, maintenance and administration easier. Some of the more often utilized tools are Query Analyzer, Enterprise Manager, SQL Profiler, Data Transformation Services and Service Manager. SQL Server 2005 is the database used in this work and Fig -2 illustrates database design of e-commerce application.

3.3 Alogrithms

3.3.1 MS Apriori Algorithm for border line rare items

E commerce data is stored in database, we fetch data from database and input to MS Apriori Algorithm. Before we input to the algorithm data is preprocessed, where we remove irrelevant data and extract relevant data. Only transactions inputted to the algorithm.

Initially set minimum support threshold and minimum confidence.
To find border line rare products we provide margin (Example: 25%).

Steps as Follows;

Step 1: To determine the support counts of each item we need to scan the data set.

Step 2: While generating L1 i.e. frequent one item set, we need to consider minimum support threshold with margin for border line rare items. For remaining items we consider just the minimum support threshold.

Step 3: To obtain the set of candidate k - item set join Lk-1(frequent item set from last step).

Step 4: To determine support of each candidate k – item set again scan the transaction database.

Step 5: Until C i.e set of candidate k - item set is equal to null set add to frequent item set. While generating frequent item set, we need to consider minimum support threshold with margin for border line rare items. For remaining items we consider just the minimum support threshold.

Step 6: Obtain all possible non empty subsets for each item in the frequent item set.

Step 7: Confidence is calculated for each subset obtained in the last step. Confidence is compared against specified confidence. Items in the subset are added to Strong Association Rule list if confidence is greater than or equal to the specified confidence.

3.3.2 Eclat Algorithm for border line rare items

E commerce data is stored in database, we fetch data from database and input to Eclat Algorithm. Before we input to the algorithm data is preprocessed, where we remove irrelevant data and extract relevant data. Only transactions inputted to the algorithm.

Initially set minimum support threshold and minimum confidence.
To find border line rare products we provide margin (Example: 25%).

Steps as Follows;

Step 1: To determine the support counts of each item we need to scan the data set. Add the transaction ids instead of specifying the actual support.

Step 2: While generating L1 i.e. frequent one item set, we need to consider minimum support threshold with margin for border line rare items. For remaining items we consider just the minimum support threshold.

Step 3: To obtain the set of candidate k - item set join Lk-1(frequent item set from last step).

Step 4: To calculate the support of each candidate k – item set, scan the item set in the previous step. While calculating support count of candidate items, we compare with previous step, no need to again scan the database and compare with original data-set.

Step 5: Until C i.e set of candidate k - item set is equal to null set add to frequent item set. While generating frequent item set, we need to consider minimum support threshold with margin for border line rare items. For remaining items we consider just the minimum support threshold.

Step 6: Obtain all possible non empty subsets for each item in the frequent item set.

Step 7: Confidence is calculated for each subset obtained in the last step. Confidence is compared against specified confidence. Items in the subset are added to Strong Association Rule list if confidence is greater than or equal to the specified confidence.

3.4 Implementation Details

The proposed work is implemented using C#, which is an object oriented programming language.

Presentation layer which is the front end(User Interface) is implemented using Asp.NET. Business logic is invoked by presentation layer, when an event occurs due to action such as click of an radio button, selecting items from a

drop down list etc. An object from the Business logic class then invokes the method of table Adopter. Database connection establishment is the responsibility of table adopter. SQL server 2005 is the database used in this application and hence SqlDataSource is responsible for communication with the database. SqlDataSource will execute the method and returns the result to the business logic. Business Logic method will return the result to the Presentation Layer.

4. RESULT ANALYSIS

Proposed system is implemented and tested in the PC with the following configurations, Intel Core i5 processor, 2.50 GHz processor speed and 4 GB RAM.

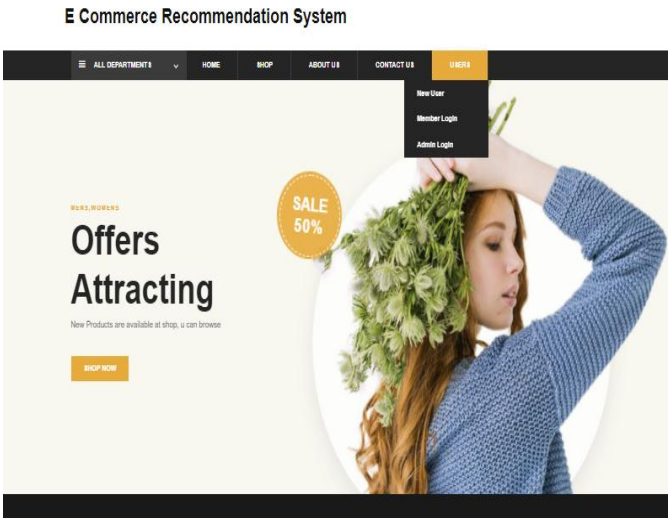

Fig -3: Home page of the e-commerce website

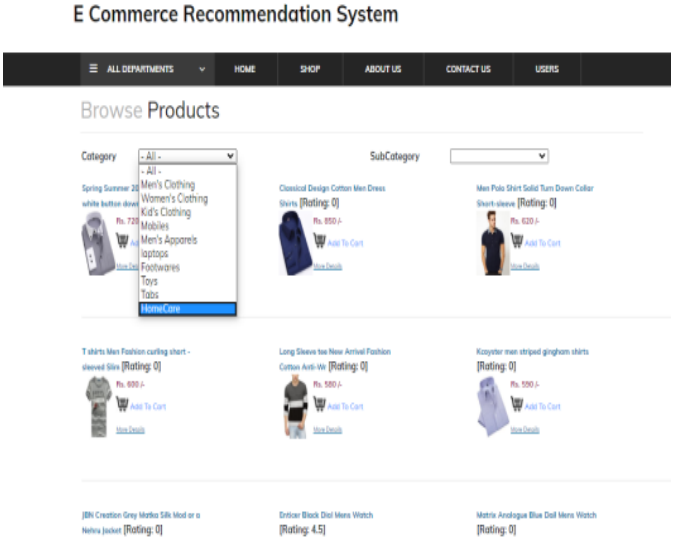

Fig -4: Browse products page for member and guest

Fig -3 shows the home page of the e-commerce website, "shop" tab allow members to shop the product and "users" tab provides option for new user registration, admin/member login. "About us" and "contact us" page gives respective information. Fig -4 shows product browsing page.

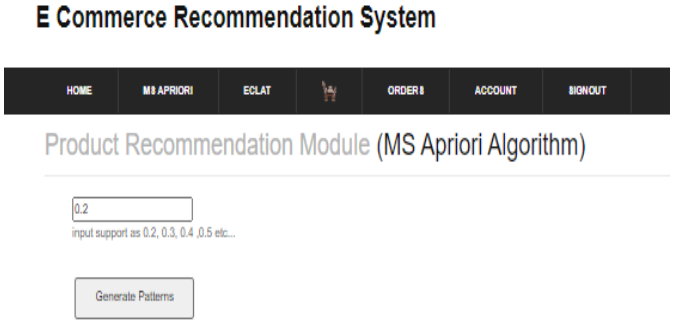

Fig -5: Recommendation system using MS Apriori algorithm for registered user/member

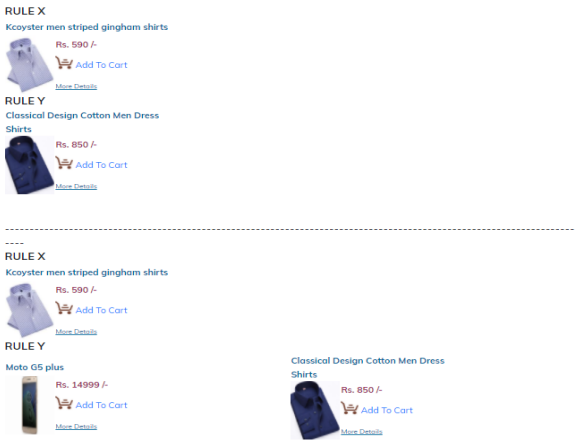

Fig -6: Strong association rules generated by MS Apriori algorithm

Fig-5 shows the recommendation system using MS Apriori algorithm, available only for the registered users. We need to provide minimum support threshold for borderline rare items considering the margin parameter, for example if minimum support threshold is defined as 0.3, consider margin as 10% then minimum support threshold for borderline rare item will be 0.2(0.3-0.1). Fig -6 shows strong association rules generated by MS Apriori algorithm.

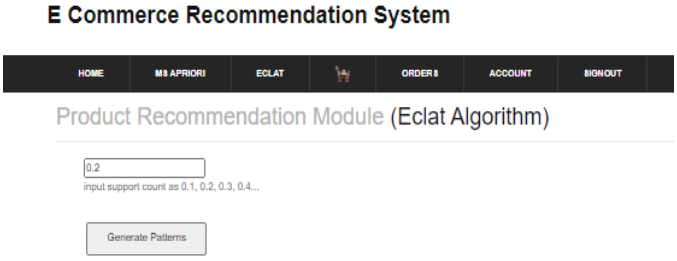

Fig -7: Recommendation system using Eclat algorithm for registered user/member

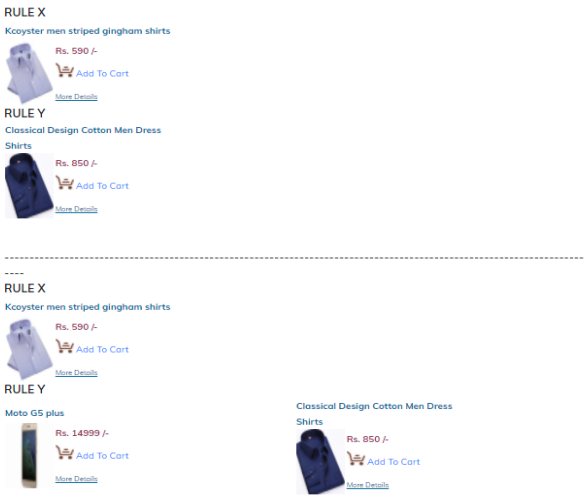

Fig -8: Strong association rules generated by Eclat algorithm

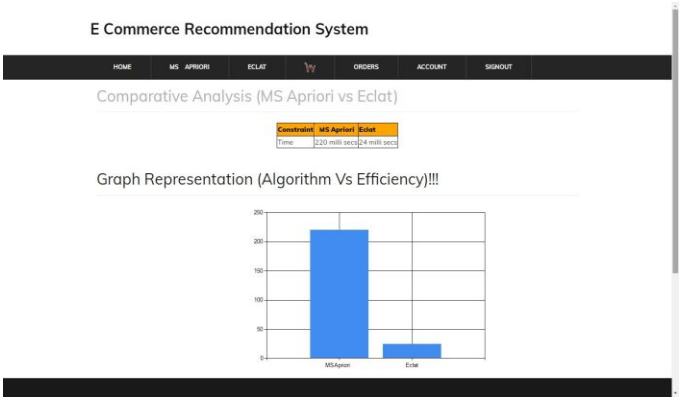

Fig 9: Graphical representation of comparative analysis of MS Apriori and Eclat algorithm

Fig -7 shows the recommendation system using Eclat algorithm, available only for the registered users. We need to provide minimum support threshold for borderline rare items, this value has to be same as the one provided to MS Apriori algorithm since we carry out performance analysis. Fig -8 shows the Strong association rules generated by Eclat algorithm. Fig -9 shows the graphical representation of comparative analysis of MS Apriori and Eclat algorithm.

4.1 Performance Comparison between MS Apriori Algorithm and Eclat Algorithm

Performance Comparison between MS Apriori Algorithm and Eclat Algorithm is been carried out with database size (number of transactions) ranging from 50 – 400. We considered minimum support threshold=0.3, minimum confidence=0.7 and margin=10%, as a part of prerequisite for comparative analysis. Graphs in the Fig -10, Fig -11, Fig -12 and Fig -13 represents algorithms name (x-axis) against algorithms execution time in milliseconds (y-axis). From all the graphs it is evident that Eclat algorithm is more time efficient than MS Apriori algorithm.

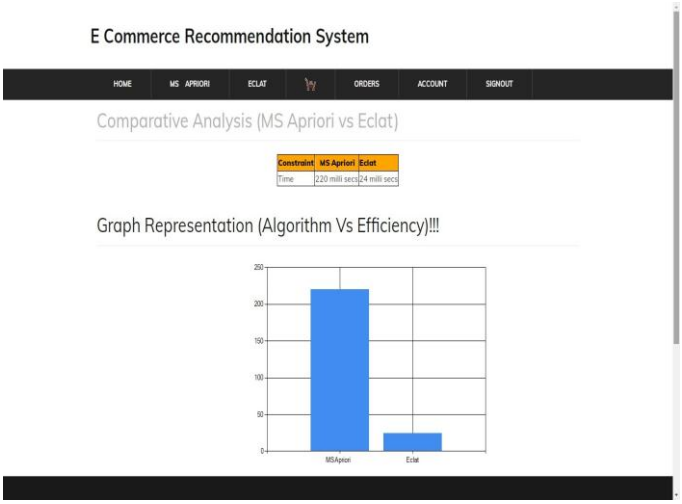

Fig -10: Comparative analysis for 50 transactions

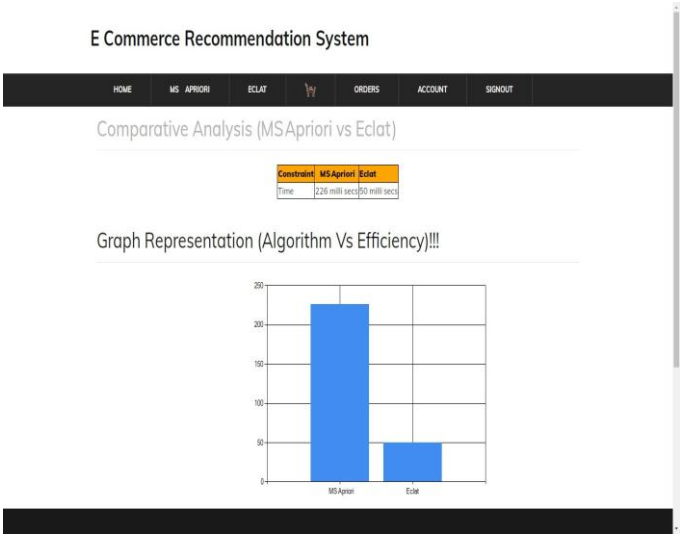

Fig -11: Comparative analysis for 100 transactions

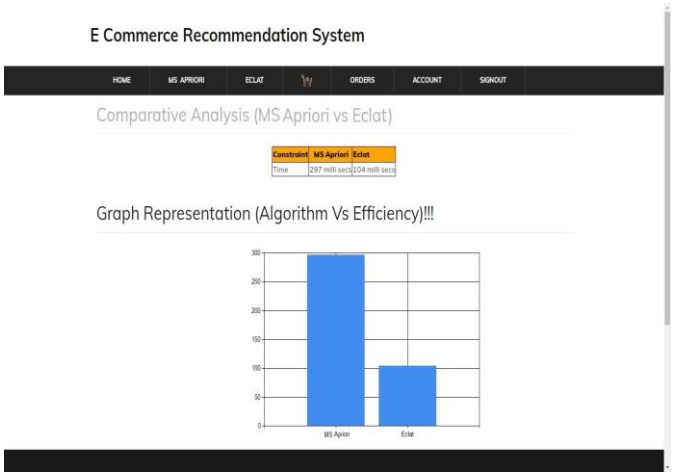

Fig -12: Comparative analysis for 200 transactions

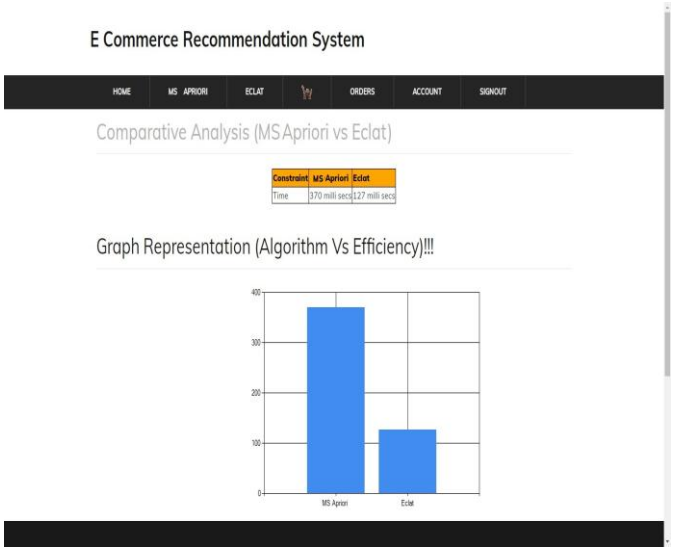

Fig -13: Comparative analysis for 400 transactions

5. CONCLUSION AND FUTURE ENHANCEMENT

5.1 Conclusion

In this work a real-time E-Commerce application is been developed which as three main actors Admin, Visitor and Member. Visitor can only browse products where as member has permission to access recommendation system. MS Apriori Algorithm and Eclat Algorithm that have been implemented in this paper to discover strong correlation between borderline rare items and frequent items are compared against their performance. Performance Comparison between MS Apriori Algorithm and Eclat Algorithm is been carried out with database size (number of transactions) ranging from 50 – 400. In all the cases, it is observed that Eclat algorithm is time efficient than MS Apriori algorithm.

5.2 Future Enhancement

• Online money payment is not implemented in this work, this can also be an enhancement to the current application.

• Proposed system uses MS apriori and Eclat algorithm to find the relationship between border line rare items and frequently purchased items, in data science we have many algorithms to find patterns, in future we can add more algorithms to find patterns and algorithms can be compared to find the efficient algorithm.

• In this work only few categories of products considered in E-Commerce application such as laptops, tabs, Men's Apparels, foot-wares, mobiles and clothing, in future we can add more product categories.

REFERENCES

[1] Hovale, Sony & Ghuli, Poonam. (2016). Survey Paper on Recommendation System using Data Mining Techniques. International Journal Of Engineering And Computer Science. 10.18535/ijecs/v5i5.60.

[2] Rao, Dr & Varakumari, S & B, Vineetha & Satish, V. (2018). Application of Data Mining to E-Commerce Recommendation Systems. International Journal of Engineering & Technology. 7. 420. 10.14419/ijet.v7i2.32.15730.

[3] Godbole, Devdatta and Manish Narnaware. "A Survey on Personalized Service Recommendation Systems." International journal of engineering research and technology 5 (2016): n. pag.

[4] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," in IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 6, pp. 734-749, June 2005, doi: 10.1109/TKDE.2005.99.

[5] YUN SING KOH, SRI DEVI RAVANA, "Unsupervised Rare Pattern Mining: A Survey" in ACM Transactions on Knowledge Discovery from Data, Vol. 10, No. 4, Article 45, 2016

[6] NIZAR R. MABROUKEH and C. I. EZEIFE, "A Taxonomy of Sequential Pattern Mining Algorithms" in ACM Computing Surveys, Vol. 43, No. 1, Article 3, 2010

[7] S. Kesarwani, A. Goel and N. Sardana, "MSD-Apriori: Discovering borderline-rare items using association mining," 2017 Tenth International Conference on Contemporary Computing (IC3), Noida, 2017, pp. 1-4, doi: 10.1109/IC3.2017.8284319.

[8] Ameera M. Almasoud et al., "Recent Developments in Data Mining Applications and Techniques" in The Tenth International Conference on Digital Information Management, 2015

[9] RuPeng Luan et al., "A Dynamic Improved Apriori Algorithm and Its Experiments in Web Log Mining" in 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2012), 2012

[10] Fournier-Viger, P., Lin, C.W., Gomariz, A., Gueniche, T., Soltani, A., Deng, Z., Lam, H. T. (2016). The SPMF Open-Source Data Mining Library Version 2. Proc. 19th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD 2016) Part III, Springer LNCS 9853, pp. 36-40.

[11] Bing Liu et al., "Mining Association Rules with Multiple Minimum Supports" in ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-99),August 15-18,1999,San Diego,CA,USA

[12] AGRAWAL, R. AND SRIKANT, R. 1994. Fast algorithms for mining association rules. In Proceedings of the 1994 International Conference on Very Large Data Bases (VLDB'94). 487– 499.