

# A Study on the Digital Wallet Usage among Citizens of Kochi using FP-Growth Algorithm

Aparna H, Karthika S, Rajalakshmi V R

**Abstract:** Digital transaction has gained popularity among the Indian citizens after the demonetization of 2016. People have started to prefer cashless payment methods for their financial transactions and digital wallets has emerged as one of the significant tools. A digital wallet is a smart phone application provided by a service provider which enables the user to keep money and to make electronic transactions. They are an alternative version of the traditional leather wallets. The money can be loaded to the wallet from bank account by using any electronic payment methods such as internet banking, mobile banking, debit cards etc. as well as it can be transferred to other accounts as well. There are many types of transactions for which digital wallets are used such as paying utility bills, mobile recharge, fund transfer, donations, etc. This paper focuses on mining association rules on the types of transaction for which people use digital wallet the most using FP-growth algorithm. The algorithm is best suited for mining out interesting patterns from large datasets. A survey is conducted among the citizens of Kochi whose results is used as dataset for the algorithm. RapidMiner Studio is used for analyzing the data. Through our research we found that citizens of Kochi use their app the most for mobile recharging, train/bus/cab/airline booking and movie ticket booking. We have also found out the results based on gender. This data can be used by both private and public sector companies in building an efficient app having a wider public approval and thus promoting cashless transactions in our country

**Keywords:** Association rules, Data mining, Digital Wallet, FP-Growth algorithm

## I. INTRODUCTION

We are inundated with varieties of data and analyzing them has become a necessity these days. Data Mining is the process of converting a large assemblage of data into knowledge [12]. Commonly known as Knowledge discovery from data (KDD) or data archeology, it is the extraction of interesting patterns from large quantity of data using intelligent methods. There are various steps associated with interpreting these patterns which includes data cleaning, integration, selection, transformation, mining, evaluation and presentation. Data mining is performed by various enterprises with an aim of generating effective and efficient business strategies. Association rule mining is intended to identify

frequently occurring patterns, associations and correlations from large data repositories. A Market Basket Analysis is often performed by retailers by making use of association rules to uncover combination of items that are purchased together by the customers or for finding out the group of services used by them. FP-growth algorithm is an effective and scalable method for discovering frequent patterns [14].

The advancement in technology has led to a sharp increase in the number of online purchase of goods and services in our country. The change is more prevalent in the digital wallet market and the usage has increased to about 33% in the year 2017 after the demonetization of November 2016[2]. A digital wallet is a smart phone installed application that allows the users to transfer amount from their bank accounts to the electronic wallet for making cashless transactions. They can also be referred to as a substitute for the usual leather wallets since it allows the user to store money digitally.

A substantial part of the population of Kerala has adopted paperless mode of payment by making use of Digital Wallets [3]. The paper aims at finding out the frequent itemset of the type of transaction for which people use Digital wallets. A survey was conducted among the citizens of Kochi as a part of the study and the responses were recorded. The results are generated using RapidMiner Studio, data mining software which provides an environment for data mining, machine learning, text mining, etc.

## II. ASSOCIATION RULE MINING

Mining of frequent patterns from vast amounts of data plays a vital role in mining associations, correlations and countless relationships between data. Association rules was introduced by R. Agrawal and R. Srikant in 1994[15], for mining the regularities between item sets in a large customer transaction database.

An association rule is composed of two parts: an antecedent(X) and a precedent(Y) which is of the form  $X \rightarrow Y$ . Support and confidence are the two measures that control the Association rules. The certainty of a rule is measured by confidence and the usability of a rule is measured by support. A set of items is called an item set. An

Revised Manuscript Received on 30 March 2019.

\* Correspondence Author

**Aparna H\***, Dept. of Computer Science and IT, Amrita School of Arts and Sciences, Amrita Vishwa Vidyapeetham, Kochi, India.

**Karthika S**, Dept. of Computer Science and IT, Amrita School of Arts and Sciences, Amrita Vishwa Vidyapeetham, Kochi, India.

**Rajalakshmi V R**, Dept. of Computer Science and IT, Amrita School of Arts and Sciences, Amrita Vishwa Vidyapeetham, Kochi, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

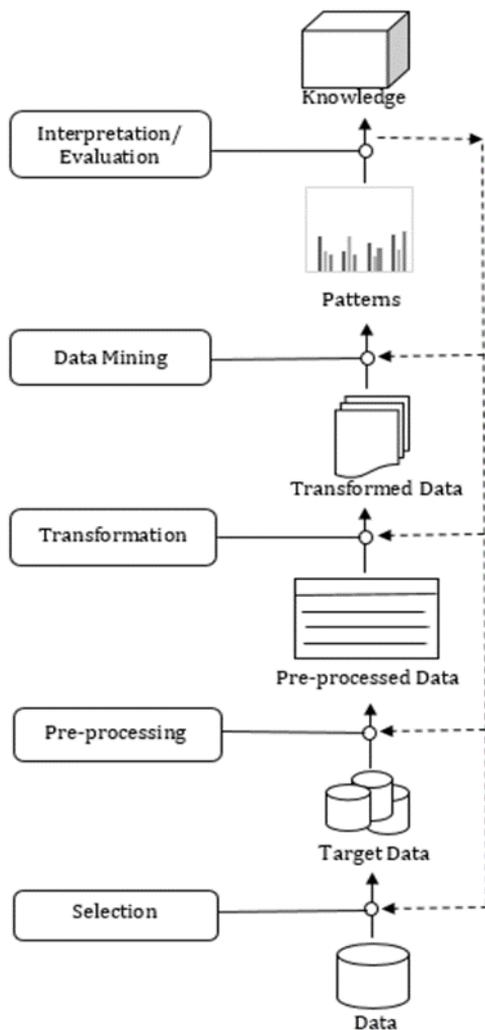


Fig. 1 KDD Process in Data Mining

association rule is actually considered more exciting if it could ensure a minimum support and confidence threshold set by the experts. If it does, then the itemset can be called as a frequent itemset. Consider a database  $D$  consisting of several transaction. If  $X \rightarrow Y$  is an association rule generated, Support signifies the probability that can be calculated by using the formula,

$$\text{Support}(X \rightarrow Y) = P(X \cup Y)$$

And Confidence represents the conditional probability, found using the formula

$$\text{Confidence}(X \rightarrow Y) = P(Y|X)$$

For example from the rule  $\{\text{onions, potatoes}\} \Rightarrow \{\text{tomatoes}\} [\text{support}=5\%, \text{confidence}=70\%]$ , it can be inferred that if a customer purchases onion and potatoes together, they would also purchase tomatoes. A support of 5% means that of all the transactions under analysis shows that onions, potatoes and tomatoes are purchased together and confidence of 70% means that 70% of customers who bought onions and potatoes also bought tomatoes.

The Association rules can be extracted using a two-step process:

- 1) Finding out all the frequent itemsets

- 2) Generating strong association rules from the frequent itemset produced.

The two prominent frequent itemset mining methods are Apriori and FP-Growth.

Apriori algorithm uses previous knowledge regarding properties of frequent itemset [13]. The algorithm was proposed by R. Agrawal and R. Srikant (1994) and the key idea was to find large itemsets that make multiple passes over the database. The major disadvantage with this algorithm is that as the number of candidate itemset increase, more memory and time is consumed [10].

### III. BACKGROUND STUDY

Association rule mining is one of the most prominent data mining tasks. Vast amount of data is getting accumulated in the databases each day and mining for frequent pattern is becoming a trend. A frequent pattern is a pattern that occurs often in a dataset. FP-growth is an efficient frequent pattern mining algorithm used these days and is one of the fastest approaches available.

The first approach for mining frequent pattern was proposed by R. Agrawal and R. Srikant[15], which was the Apriori algorithm. Various versions of the algorithm was developed to improve its performance. The main problem with this approach is the multiple database scan which cost more memory and time.

In an experimental study, conducted by M. S. Mythili and A. R. Mohamed Shanavas[10] on the Mushroom and Supermarket datasets, it was found that FP-Growth is most efficient and scalable than Apriori algorithm based on the execution time taken for both for mining patterns.

A comparison made between Apriori, FP-Growth and Tertius mining algorithms using WEKA [11] resulted in FP-Growth being the fastest algorithm among the three.

Market basket analysis is common these days for mining frequent pattern in transaction datasets. Aditya Pradana and Heru Agus Santoso[5] have shown in their paper that FP-Growth can be effectively applied in mining association rules from a transaction dataset of an organic medicine store.

Jubair. T and Yakoob. C [3] from their study has found out that there is a considerable hike in the number of digital wallet users in India, especially among the residents of Kerala. The acceptance of digital wallets are equal among people with different education level, employment status and income.

### IV. FP-GROWTH ALGORITHM

FP-Growth algorithm is a solution for mining frequent item sets from a database without generating candidate sets.

The strategy followed is divide and conquer and uses a two-step process. In the first step the entire database representing the frequent items is transformed into a compact structure called a FP-Tree or a frequent pattern tree. In step two the frequent itemsets are extracted directly from the FP-Tree. The algorithm was proposed by J. Han, H. Pei, and Y. Yin [14].

The compactness of the tree structure, which is much smaller than the whole database, saves memory and time required for database scans in the succeeding mining process and the item sets that overlap shares the same prefix path. Each path of the tree represents a frequent itemset and the nodes are sorted based on the decreasing order of support count.

**A. Algorithm**

1. FP- Tree construction

Pass 1:

1. Scan the whole database and get the support for each item.
2. Discard infrequent items.
3. Based on the values of the support, the frequent itemset must be sorted in the decreasing order.

Pass 2: Each node will now corresponds to an item and they will all have a counter.

1. The transactions will be read one at a time and is mapped to a path.
2. Predetermined order is used to overlap paths when transaction shares items (if they have the same prefix).
  - Counters are incremented in this case.
3. Pointers between nodes comprising the same item are retained, resulting in a linked list.
  - The more paths overlap, the greater the compression. FP- Tree may fit in the memory.
4. Frequent item sets is extracted from the FP- Tree.

**V. EXPERIMENT AND RESULTS**

The collected data set collected with the help of Google Forms, can be saved in a .csv format where each record represents a response. This is effective in extracting similar patterns that are concealed in the dataset. Fig. 2a and 2b represents the screenshot of the form used for the survey. The type of attributes used is given in Table I. The data collected can be represented in a binary format, where each column corresponds to a transaction type and each row corresponds to a response. Fig. 3 depict a screenshot of the binary form of the original data collected.

Name \*

Your answer

Gender \*

- Female
- Male

Are you using any Digital Wallets? \*

- Yes
- No

Name your app(s). Put nil if not applicable \*

Your answer

(a)

What type of transactions do you prefer through your app? \*

- Never used
- Fund transfer
- Mobile recharging
- Movie ticket booking
- Train/Bus/Cab/Airline ticket booking
- DTH recharge
- Donations
- Paying utility bills(electricity, telephone,etc)
- Online Shopping
- In-store payments(at grocery stores, restaurants, parking, tolls,etc with the app's QR code)
- Gift cards
- Other:

(b)

**Fig. 2(a and b) Survey Form**

Table 1

No.	Name of Transaction
1	Fund Transfer
2	Mobie Recharging
3	Movie Ticket Booking
4	Train/Bus/Cab/AirlineBooking
5	DTH
6	Donations
7	Utility Bills
8	Online Shopping
9	Instore Payments
10	Gift Cards

	A	B	C	D	E	F	G	H	I	J	K
1	fund_transfer	mob_recharge	movie_tkt_book	train/bus/cab/AL	DTH	donations	utility_bills	online_shopping	instore_paymnt	gift_cards	
2	0	1	1	0	0	0	1	1	1	0	
3	0	0	0	0	0	0	0	0	0	1	0
4	1	0	1	0	0	0	0	0	0	0	0
5	1	0	0	0	0	0	0	0	0	0	0
6	1	1	0	1	0	0	0	1	0	0	0
7	1	1	1	1	1	1	1	1	0	0	0
8	1	1	0	1	0	0	0	0	0	0	0
9	1	1	1	0	0	0	0	0	0	0	0
10	0	1	1	1	0	0	1	1	0	0	0
11	0	0	0	0	0	0	0	1	0	0	0
12	1	0	0	0	0	0	1	1	0	0	0
13	1	1	0	1	0	0	0	1	1	0	0
14	1	1	1	1	0	0	0	1	0	0	0
15	1	1	0	0	0	0	0	0	0	0	0
16	1	1	1	1	1	0	0	1	0	0	0
17	1	0	1	1	0	0	0	0	1	0	0
18	1	0	1	0	0	0	0	0	0	0	0
19	1	1	1	1	0	0	0	1	1	1	0

Fig. 3 Binary representation of data collected

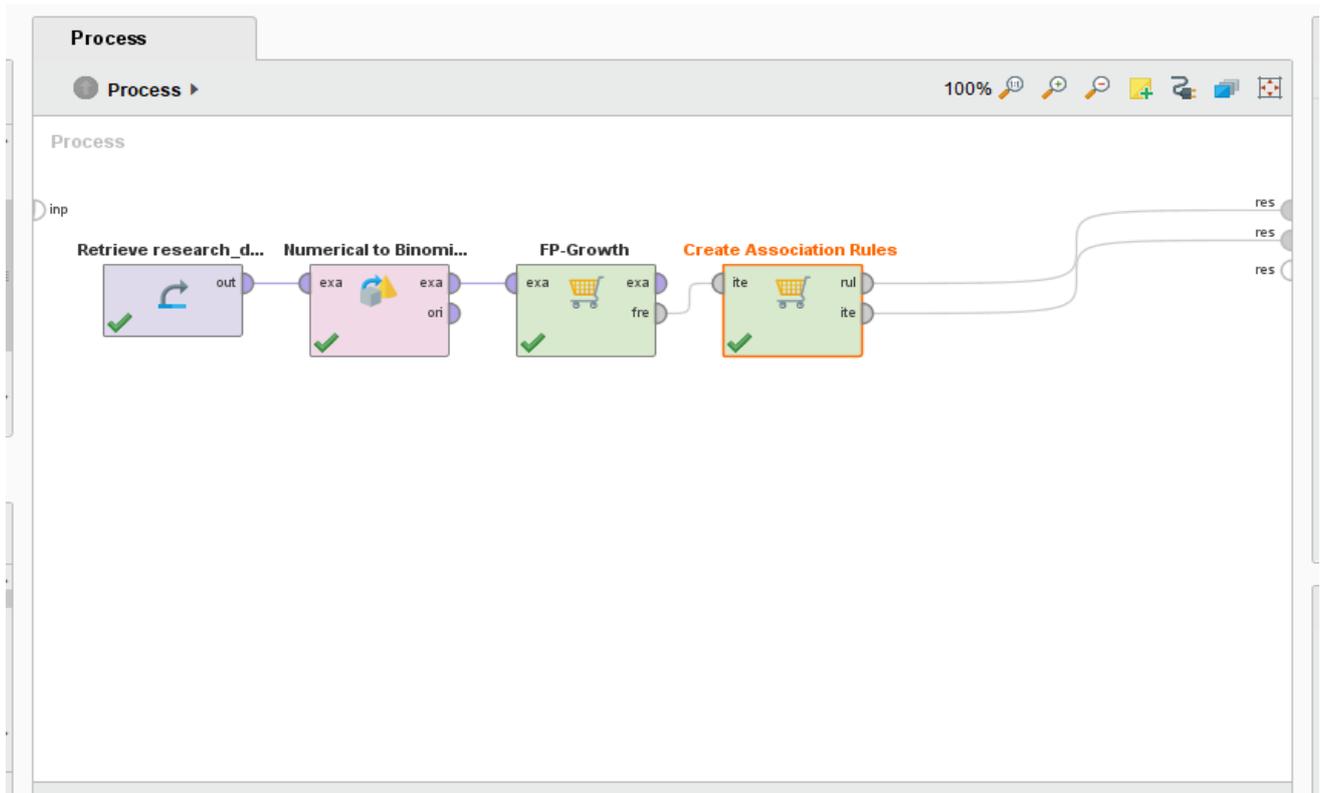


Fig. 4 Frequent itemset generation process in RapidMiner

Table II

Support

	0.5	0.6	0.7	0.8	0.9
Confidence 0.7	98	63	53	40	31
Confidence 0.8	37	25	21	13	10
Confidence 0.9	3	1	1	0	0
Confidence 1.0	0	0	0	0	0

The mining process was performed in RapidMiner Studio and Fig. 4 shows the scheme of generating association rules and frequent itemsets in the same. There were a total of 200 responses each of which can be taken as an instance and 10 attribute which corresponds to type of transactions preferred. For the study we have used various values for minimum support and minimum confidence to achieve most decisive rules.

**A. General Category**

Table II shows the minimum support and minimum confidence values used and the number of association rules produced.

Based on Table II, it can be concluded that as the value of minimum support and minimum confidence increase, less number of association rules are generated. We are going to choose the value of minimum support and minimum confidence as 0.5 and 0.9 respectively as parameters. A

support of 0.5 is chosen by taking into account the number of rules produced which are not too many and not too few. A confidence of 0.9 is taken because higher the confidence, stronger the rule. No rules are generated for minimum confidence of 1.0. Table III describes the rules generated with their corresponding minimum support and minimum confidence values. From Table III it is evident that Mobile recharging, Train/Bus/Cab/Airline booking and Movie ticket booking appears in all the 3 rules. Thus these are the type of transactions that are mostly preferred along with online shopping, utility bills and fund transfer by the Digital Wallet users of Kochi.

# A Study on the Digital Wallet Usage among Citizens of Kochi using FP-Growth Algorithm

## B. Female Category

Table IV shows the number of association rules generated for various values of support and confidence. Total number of female respondents is 105. Table V represents the association rules generated.

Train/bus/cab/airline booking, utility bills, movie ticket booking appears in most of the rules generated along with online shopping and mobile recharging. Thus these are the type of transactions preferred by female citizens of Kochi.

**Table III**

No	Rules	Support	Confidence
1	If use fund transfer, online shopping, movie ticket booking and train/bus/cab/airline booking also use mobile recharging	0.15	0.93
2	If use mobile recharging, online shopping, train/bus/cab/airline ticket booking and utility bills also use movie ticket booking	0.11	0.95
3	If use fund transfer, movie ticket booking, train/bus/cab/airline booking and utility bills also use mobile recharge	0.12	0.96

**Table IV**

Support

	0.5	0.6	0.7	0.8	0.9
0.7	98	63	53	40	31
0.8	37	25	21	13	10
0.9	3	1	1	0	0
1.0	0	0	0	0	0

**Table V**

No	Rules	Support	Confidence
1	If use train/bus/cab/airline booking and utility bills also use movie ticket booking	0.19	0.90
2	If use online shopping, fund transfer and train/bus/cab/airline booking also use mobile recharging	0.20	0.90
3	If use mobile recharging, online shopping and utility bills also use movie ticket booking	0.13	0.92
4	If use mobile recharging, train/bus/cab/airline booking and utility bills also use movie ticket booking	0.13	0.92

5	If use online shopping, train/bus/cab/airline booking and utility bills also use movie ticket booking	0.14	0.93
---	---	------	------

Table VI

C. Male Category

Table VI shows the number of association rules generated for various values of support and confidence. The total number of male respondents is 95. Table VII represents the association rules generated. Train/bus/cab/airline booking appears in all the rules and mobile recharging and movie ticket booking appears in most of the rules generated along with fund transfer, utility bills and online shopping. Thus these are the type of transactions preferred the most by the male citizens of Kochi.

		Support				
		0.5	0.6	0.7	0.8	0.9
Confidence	0.7	174	114	72	49	37
	0.8	80	48	26	17	10
	0.9	30	13	4	2	0
	1.0	0	0	0	0	0

Table VII

No	Rules	Support	Confidence
1	If use train/bus/cab/airline booking and donations also use fund transfer	0.15	0.93
2	If use fund transfer, movie ticket booking, utility bills and train/bus/cab/airline booking also use mobile recharging	0.15	0.93
3	If use mobile recharging, online shopping and train/bus/cab/airline booking use movie ticket booking	0.17	0.94
4	If use movie ticket booking, utility bills and train/bus/cab/airline booking also use mobile recharging	0.17	0.94

VI. CONCLUSION AND SCOPE FOR FUTURE

From our study, we found that citizens of Kochi use digital wallets apps the most for mobile recharging, train/bus/cab/airline booking and movie ticket booking. While women prefer Train/bus/cab/airline booking, utility bills, movie ticket booking through their Digital Wallets, men opt for Train/bus/cab/airline booking, mobile recharging and movie ticket booking. These results could be helpful for Digital wallet service providers in determining the type of transaction for which people use Digital wallets the most in order to increase their approval. Since the smartphones are becoming regular and the rates of internet has been decreasing day by day, electronic wallet is a good option for the public for making payments, recharges, etc. The algorithm we applied was FP-Growth, one of the fastest frequent pattern mining algorithms. By providing different values for support and confidence, we obtained the output in the form of association rules of the type of transactions preferred. The study is limited to Kochi region and the dataset used is relatively small. A bigger dataset and an improved algorithm can yield better results.

REFERENCES

- Mohini H. Chandwani, "Market Basket Analysis using Association Rule", International Journal of Advance Research, Ideas and Innovations in Technology, vol. 4, issue 4, 2018
- N. Vohra and K. Hazra. "The State of E-Wallets and Digital Payments in India", 2018. Available: <https://www.regalix.com/wp-content/uploads/2018/03/The-state-of-e-wallets-and-digital-payments-in-India-2018.pdf>
- Jubair. T and Yakoob. C. "Acceptance of Digital Wallets in Kerala: An Empirical Study". International Journal of Business and General Management (IJBG), vol. 6, pp. 31-38, April 2017
- D. Bhuvanewari and S. Sivakavitha, "An Intellectual Study on Preference towards the Usage of Electronic Wallets among Urban Population of Chennai City", Imperial Journal of Interdisciplinary Research (IJIR), vol. 3, issue 9, 2017
- Aditya Pradana and Heru Agus Santoso, "Market Basket Analysis Using FP-Growth Algorithm in Organic Medicine Store", 2016. Available: <http://mahasiswa.dinus.ac.id/docs/skripsi/jurnal/17928.pdf>
- Kavitha Venkatachari, "Market Basket Analysis using FP-Growth and Apriori Algorithm: A Case study of Mumbai Retail Store", BVIMSR's Journal of Management Research vol. 8, issue 1, April 2016
- Siddhrajsinh Solanki and Neha Soni, "A Survey on the Frequent Pattern Mining Methods Apriori, Eclat and FP-Growth", International Journal of Computer Techniques, vol. 10, issue 10, 2015



8. Avadh Kishor Singh, Ajeet Kumar and Ashish K Maurya, "An Empirical Analysis and Comparison of Apriori and FP-Growth Algorithm for Frequent Pattern Mining", In IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), 2014
9. Neesha Sharma and Chander Kant Verma, "Association Rule Mining: An Overview", IJCSC vol. 5, issue 1, March 2014, pp. 10-15
10. M. S. Mythili and A. R. Mohamed Shanavas, "Performance Evaluation of Apriori and FP-Growth Algorithms", International Journal of Computer Applications, vol. 79, October 2013 pp. 34-37
11. Jyoti Arora, Shelza and Sanjeev Rao, "An Efficient ARM Technique for Information Retrieval in Data Mining", International Journal of Engineering Research and Technology, vol. 2, issue 10, October 2013
12. Rakesh Kumar Soni, Neetesh Gupta and Amit Sinhal, "An FP-Growth Approach to Mining Association Rules", International Journal of Computer Science and Mobile Computing, vol. 2, issue 2, February 2013, pp. 1-5
13. J. Han, M. Kamber and J. Pei. "Data Mining Concepts and Techniques". 3rd ed. Morgan and Kaufmann, 2012, pp.1-5.
14. J. Han, H. Pei, and Y. Yin. "Mining Frequent Patterns without Candidate Generation", In Proc. Conf. on the Management of Data (SIGMOD '00, Dallas, TX). ACM Press, New York, NY, USA 2000
15. R. Agrawal, and R. Srikant, "Fast Algorithms for Mining Association Rules," In Proc. VLDB, pp. 487-499, 1994.

## AUTHORS PROFILE



**Aparna H**, PG Student, Master of Computer Application, Amrita Vishwa Vidyapeetham, Amrita School of Arts and Sciences, Brahmasthanam, Edappally North P.O. Kochi - 682 024, Kerala.



**Karthika S**, PG Student, Master of Computer Application, Amrita Vishwa Vidyapeetham, Amrita School of Arts and Sciences, Brahmasthanam, Edappally North P.O. Kochi - 682 024, Kerala.



**Rajalakshmi V R**, Assistant Professor, Dept. of CS & IT, Amrita Vishwa Vidyapeetham, Amrita School of Arts and Sciences, Edappally North P.O. Kochi-682024, Kerala. Quallification -M.sc (Information Technology) M.Tech (Information Technology). Research

papers-COMPARATIVE STUDY OF MOBILE AD HOC NETWORK ROUTING PROTOCOLS, An extension to page rank algorithm, A New Approach to Increase the Storage Efficiency of Databases Using BIT Representation, An Effective Approach to Improve Storage Efficiency Using Variable bit Representation. Area of interest-Block chains,Bigdata, Cloud Computing, Database.