

# Analyzing the Usage Pattern of University Website using Apriori Algorithm through Frequent Item set Generation

S.Muthu Mari  
M.Phil Research Scholar, Comp.Sci &Engg.  
Alagappa University  
Karaikudi, India

T.Meyyappan, PhD  
Professor, Dept of Comp.Sci & Engg.  
Alagappa University  
Karaikudi, India

## ABSTRACT

Data Mining plays important role in finding previously unknown patterns from huge volume of data. The data mining extends its branches in many areas in which Web Mining is one of the important arena. Analyzing the web usage and its content becomes very important aspect. The web log servers are maintained to estimate total number of usage of the pages in websites. It holds the details such as the Internet Protocol Address of the user, Session, Browser as well as cookies information along with the date and time of visit. To find the effective pattern analysis and collecting the interesting patterns in website (web mining), association rule mining, clustering techniques were employed. Apriori is a classic algorithm for learning association rules. In this paper, the Web server log is analyzed into two different ways. Frequent pattern and analyzed with IP address of the visitor and also with time of visit. Based on the two refinements the Apriori algorithm generates the mined patterns that occur frequently.

**Keywords**— Apriori Algorithm, Web Mining, Web Usage, Association Rules, Internet Protocol address.

## 1. INTRODUCTION

Extracting information and patterns from web data is known as web mining. Three type of web mining are in use namely, Web content mining, Web structure mining, Web usage mining. Web Content mining is the scanning and mining of text, pictures and hyperlinks of a Web page to determine the relevance of the content to the search query.

Web Content mining provides the results lists to search engines in order of highest relevance to the keywords in the query. Web structure mining is a tool used to identify the relationship between Web pages linked by information or direct link connection. Web Usage Mining is the application of data mining techniques to discover interesting patterns from Web data, in order to understand and better serve the needs of Web-based applications. The discovered patterns from the web log were represented as collections of pages, objects, or resources that are frequently accessed by users with common needs [13].

The three steps in Web usage mining are the Data preprocessing, pattern discovery and pattern analysis. Web server Log is considered as the primary source for web mining, that includes the session identification, page visit information, IP address detail along with the browsing details.. Web mining subtasks are (a) resource finding and retrieving, (b) information selection and pre-processing, (c) patterns analysis

and recognition, (d) validation, interpretation, and (e) visualization.

Web usage mining process is generally divided into two tasks: Data preparation and pattern discovery. In the task of data preparation server session file build where each session is a sequence of requests of different types made by single user during a single visit to a site. In the pattern discovery task association rules, sequential patterns, usage clusters, page clusters, user classification involve. For finding out the information that is hidden in web logs, several data mining techniques are applied on web server logs [9].

Most of the web log mining is based on the decision tree algorithms since it can be easily divided and analyzed, based on specific condition. The paper is organized as follows. Section 2 comprises of the summary of Web usage Mining techniques that uses the Apriori Algorithm. Section 3 consists of the existing work based on web log data and procedure for the proposed scheme. Section 4 includes the experimental results and their discussions. Finally section 5 concludes the paper.

## 2. LITERATURE SURVEY

This section briefly discusses the various web mining techniques and methods using the Apriori Algorithm. Various other algorithms which are the extended form of Apriori algorithm were also explained.

In this work proposed by Ankit R Kharwar et.al., [1] the Apriori is used as the influential algorithm to find out the frequent item sets from high volume of data. Once the frequent item sets are generated then, straightforwardly the association rules are incorporated to show its support and confidence between one another. All frequent item sets are found out at the first step. The frequent item set is the item set that is included at the least minimum support transactions. This algorithm is used in a way that associates the usage pattern of the clients for a particular website.

In this work proposed by Han J., Pei J., Yin Y [3] the custom built Apriori algorithm, that analyses the websites regarding educational log file to simulate various test results of educational web mining is explained. The rules or the correlations resulted from that system helps the website developer in proper decision making and can improve the usage of the site more effectively.

Web server data correspond to the user logs that are collected at Web server. Some of the typical data collected at a Web server includes IP addresses, page references, and access time of the users which are the main inputs of the present Research.

This Research work concentrates on web usage mining and especially in discovering the web usage patterns of websites from the server log files. The comparison of memory usage and time usage is compared using Apriori algorithm and Frequent Pattern Growth algorithm by the investigator. The some kind of study, though using the FP growth with Apriori algorithm was provided by Santhosh Kumar and Rukmani [11].

The proposed work by Ilayaraja M and Meyyappan T [4], the derived result presents, from the Apriori algorithm the candidate set generation is costly, especially if a large number of patterns and/or long patterns exist. The main drawback of the FP - growth algorithm is the explosive quantity of lacks a good candidate generation method.

In the work presented by Jogi.Suresh and Ramanjaneyulu T [6], a new method was developed which includes two steps. The first step is to construct the FP-tree structure, the second step is to use the Apriori algorithm to mine the FP-tree. In the second step, it is needed to add an additional node Table, named NTable. Each entry in the NTable has two fields: Item-name, and Item-support. It provides the better result that this new algorithm works much faster than Apriori.

A novel method called K-Apriori algorithm is proposed by Ashok Kumar D and Loraine charlet Annie M.C[2], to find the frequently accessed web pages from the very large binary weblog databases. Experimental results show that the proposed method has higher performance in terms of objectivity and subjectivity. The anti-monocity property makes it simple and perfect for binary databases. Initially the binary data is clustered using the multi-pass, K-means algorithm based on the linear wiener transformation and finally found it is that K-Apriori algorithm is more efficient compared to Apriori algorithm.

In the paper presented by Wang Tong,HE Pi-lian[12] the improved AprioriAll algorithm is added with the property of the User ID. While producing the candidate set the database will be scanning in every step. It decides which an item in the candidate set should be put into the Large set for produce next candidate set. At the meantime, in order to reduce the number of the database scanning, the new algorithm, by using the property of the Apriori algorithm, limits the size of the candidate set in time whenever it is produced.

## 2.1 Apriori Algorithm

Apriori is the classic algorithm used to generate frequent item sets from large transactional databases. It uses the bottom – up approach where the candidate generation is the intermediate step. The Breadth first search algorithm is the key note in Apriori technique, while searching for proper candidate key. The hash tree function technique is embedded so as to prune the unwanted data.

The problem of association rule mining is defined as follows.  $I = \{i_1, i_2, \dots, i_n\}$  be a set of n number of binary attributes called the items.  $D = \{t_1, t_2, \dots, t_n\}$  be a set of transactions called the database. Each transaction in the database has a unique transaction identification and contains a subset of the items in I. A rule is defined as an implication of the form  $X \rightarrow Y$  where  $X$ , belongs to The sets of items X and Y are called antecedent and consequent of the rule.

The Apriori algorithm works as follows. It is categorized into two sub divisions. As a first step, minimum support is applied to find all frequent item sets in a database. Secondly, these

frequent item sets and the minimum confidence constraint are used to form set of rules.

The optimal solution to find all frequent item set is a tedious process in transactions since it involves all item sets in the database. To find all item combination much space and more time is required to complete the process. Hence, the powerset can be defined to store set of all frequently occurred possible item sets. The set of possible item sets is the power set over I and has size  $2^n - 1$ . Although the size of the powerset grows exponentially in the number of items I, efficient search is possible using the downward-closure property called which guarantees that for a frequent item set, all its subsets are also frequent and thus for an infrequent item set, all its supersets must also be infrequent.

The Association rules for the Apriori algorithm contains two divisions namely the support and the confidence. For any given association rule support is the percentage of item sets that satisfy the given rule and confidence is the percentage of the population for which the antecedent is satisfied. If it goes as true then the consequent will also becomes true respectively.

By using the breadth first search method and the hash tree function apriori counts the candidate item set effectively. Candidate item set from k to k-1 is generated using the technique. Then, the pruning step is taken place which has an infrequent sub pattern.

According to the downward closure lemma, the candidate set contains all frequent k-length item sets. Then, it scans the database and determines the frequent item sets. Bottom-up subset exploration (essentially a breadth-first traversal of the subset lattice) finds any maximal subset S only after all  $2^{|S|} - 1$  of its proper subsets.

## 2.2 Procedure for Frequent Item set Generation

Following is the procedure for the proposed work:

Step 1: Read the database transactions.

Step2: Calculate the minimum support value for the transactions.

Step 3: Apply Apriori algorithm to generate the frequent item sets and all frequent item sets.

Step 4: The candidate generation for all frequent item set is done for K to K – 1 transaction.

Step 5: Calculate the confidence for each item sets based on minimum support and generate all frequently occurred patterns.

Step 6: If needed, the pruning of the all frequent item set can be taken place.

## 3. EXPERIMENTAL RESULTS AND DISCUSSIONS

The methodology for the frequent item set generation comprises following steps:

- ❖ Data Pre-processing.
- ❖ Identifying various transactions based on given dataset.
- ❖ Generating various transactions based dataset.
- ❖ Generating frequent item set for web log data collected by IPaddress and date of transactions using Apriori algorithm.

### 3.1 Data Preprocessing

An important task in any data mining application is the creation of a suitable target data set to which data mining algorithms can be applied. This is particularly important in web usage mining due to characteristics of click stream data and its relationship to other related data collected from

multiple sources and across multiple channels. Data preprocessing consists of sequence of following tasks, collection of total number of tuples from the given database. Remove the irrelevant dataset, as well as duplicate dataset. The given below is the preprocessed dataset with the fields named IP address, date and the page name session information.

IPaddress	DateFor	Pagename
172.16.19.33	24/Jan/2014:13:46:45	<a href="http://alagappauniversity.ac.in/about_au.php">http://alagappauniversity.ac.in/about_au.php</a>
172.16.19.16	24/Jan/2014:13:47:19	<a href="http://alagappauniversity.ac.in/academic/specialofficer.php">http://alagappauniversity.ac.in/academic/specialofficer.php</a>
172.16.19.01	24/Jan/2014:13:47:41	<a href="http://alagappauniversity.ac.in/examination/authorities.php">http://alagappauniversity.ac.in/examination/authorities.php</a>
172.16.19.48	24/Jan/2014:13:46:45	<a href="http://alagappauniversity.ac.in/examination/notification.php">http://alagappauniversity.ac.in/examination/notification.php</a>
172.16.19.22	24/Jan/2014:13:48:51	<a href="http://alagappauniversity.ac.in/examination/exam_results.php">http://alagappauniversity.ac.in/examination/exam_results.php</a>
172.16.19.78	24/Jan/2014:13:48:23	<a href="http://alagappauniversity.ac.in/examination/model_question_paper.php">http://alagappauniversity.ac.in/examination/model_question_paper.php</a>
172.16.19.33	24/Jan/2014:13:48:51	<a href="http://alagappauniversity.ac.in/examination/exam_results.php">http://alagappauniversity.ac.in/examination/exam_results.php</a>
172.16.19.24	24/Jan/2014:13:49:01	<a href="http://alagappauniversity.ac.in/examination/exam_results.php">http://alagappauniversity.ac.in/examination/exam_results.php</a>
.....		
.....		
172.16.19.02	24/Jan/2014:13:49:01	<a href="http://alagappauniversity.ac.in/administration/finance_committee.php">http://alagappauniversity.ac.in/administration/finance_committee.php</a>
172.16.19.16	24/Jan/2014:13:49:14	<a href="http://alagappauniversity.ac.in/examination/exam_results.php">http://alagappauniversity.ac.in/examination/exam_results.php</a>
172.16.19.33	24/Jan/2014:13:49:40	<a href="http://alagappauniversity.ac.in/examination/model_question_paper.php">http://alagappauniversity.ac.in/examination/model_question_paper.php</a>

### 3.2 IP address based Sample dataset

The following is the separated sessions with the tuples IP address and the page name of the links.

IPaddress	Pagename
172.16.19.33	<a href="http://alagappauniversity.ac.in/about_au.php">http://alagappauniversity.ac.in/about_au.php</a>
172.16.19.16	<a href="http://alagappauniversity.ac.in/academic/specialofficer.php">http://alagappauniversity.ac.in/academic/specialofficer.php</a>
172.16.19.01	<a href="http://alagappauniversity.ac.in/examination/authorities.php">http://alagappauniversity.ac.in/examination/authorities.php</a>
172.16.19.48	<a href="http://alagappauniversity.ac.in/examination/notification.php">http://alagappauniversity.ac.in/examination/notification.php</a>
172.16.19.22	<a href="http://alagappauniversity.ac.in/examination/exam_results.php">http://alagappauniversity.ac.in/examination/exam_results.php</a>
172.16.19.78	<a href="http://alagappauniversity.ac.in/examination/model_question_paper.php">http://alagappauniversity.ac.in/examination/model_question_paper.php</a>
172.16.19.33	<a href="http://alagappauniversity.ac.in/examination/exam_results.php">http://alagappauniversity.ac.in/examination/exam_results.php</a>
172.16.19.24	<a href="http://alagappauniversity.ac.in/examination/exam_results.php">http://alagappauniversity.ac.in/examination/exam_results.php</a>
172.16.19.02	<a href="http://alagappauniversity.ac.in/administration/finance_committee.php">http://alagappauniversity.ac.in/administration/finance_committee.php</a>
172.16.19.16	<a href="http://alagappauniversity.ac.in/examination/exam_results.php">http://alagappauniversity.ac.in/examination/exam_results.php</a>
.....	
.....	
172.16.19.33	<a href="http://alagappauniversity.ac.in/examination/model_question_paper.php">http://alagappauniversity.ac.in/examination/model_question_paper.php</a>

### 3.3 Date based Sample dataset

The following is the separated sessions with the tuples Date of page visit and the page name of the links.

Datefor	Pagename
24/Jan/2014:13:46:45	<a href="http://alagappauniversity.ac.in/about_au.php">http://alagappauniversity.ac.in/about_au.php</a>
24/Jan/2014:13:47:19	<a href="http://alagappauniversity.ac.in/academic/specialofficer.php">http://alagappauniversity.ac.in/academic/specialofficer.php</a>
24/Jan/2014:13:47:41	<a href="http://alagappauniversity.ac.in/examination/authorities.php">http://alagappauniversity.ac.in/examination/authorities.php</a>
24/Jan/2014:13:47:50	<a href="http://alagappauniversity.ac.in/examination/notification.php">http://alagappauniversity.ac.in/examination/notification.php</a>
24/Jan/2014:13:48:11	<a href="http://alagappauniversity.ac.in/examination/exam_results.php">http://alagappauniversity.ac.in/examination/exam_results.php</a>
24/Jan/2014:13:48:23	<a href="http://alagappauniversity.ac.in/examination/model_question_paper.php">http://alagappauniversity.ac.in/examination/model_question_paper.php</a>
24/Jan/2014:13:48:51	<a href="http://alagappauniversity.ac.in/examination/exam_results.php">http://alagappauniversity.ac.in/examination/exam_results.php</a>
24/Jan/2014:13:49:01	<a href="http://alagappauniversity.ac.in/examination/exam_results.php">http://alagappauniversity.ac.in/examination/exam_results.php</a>
24/Jan/2014:13:49:01	<a href="http://alagappauniversity.ac.in/administration/finance_committee.php">http://alagappauniversity.ac.in/administration/finance_committee.php</a>
24/Jan/2014:13:49:14	<a href="http://alagappauniversity.ac.in/examination/exam_results.php">http://alagappauniversity.ac.in/examination/exam_results.php</a>
.....	
.....	
24/Jan/2014:13:49:20	<a href="http://alagappauniversity.ac.in/examination/model_question_paper.php">http://alagappauniversity.ac.in/examination/model_question_paper.php</a>

The dataset were replaced with zero or one coding based on the collected IP Address wise and date wise . Now the dataset is convert to the ARFF (Attribute Relation File Format). Since, the processing is done at the WEKA tool, the conversion is taken place. The Apriori algorithm accepts input files in such

format. An ARFF file is an ASCII text file that describes a list of instances sharing a set of attributes. ARFF files have two distinct sections. The first section have the Header information, which is followed by the coding. The Lines that begin with a ‘ % ‘ are comments.

### 3.4 Converting Web log File to ARFF (Attribute Relation File Format)

#### 3.4.1 IP address web log file to ARFF Conversion

```
@relation University_log-Details
@attribute 'http://alagappauniversity.ac.in/about_au.php' {1,0}
@attribute 'http://alagappauniversity.ac.in/administration/vicechancellor.php' {1,0}
.....
.....
.....
@attribute 'http://alagappauniversity.ac.in/examination/exam_results.php' {1,0}
@attribute 'http://alagappauniversity.ac.in/administration/coe.php' {1,0}
@data
1,1,0,1,0,0,0,0,1,0,0,0,0,0,0,1,1,0,0,0,0,0,0,0,0,0,0,1,0,0,0,1,0,1,1
1,1,1,1,1,0,0,0,1,1,1,0,0,1,0,1,1,0,1,1,1,1,1,0,0,0,0,0,0,0,1,0,1,1
1,1,1,1,1,1,0,0,0,1,0,0,0,1,1,0,1,0,0,0,0,0,1,0,1,1,1,0,0,1,1,1,1
1,1,1,0,0,0,0,0,1,1,0,0,0,1,0,0,1,0,1,1,1,1,0,0,1,0,0,0,0,0,1,0,0,1
1,1,1,1,1,1,0,1,0,0,0,1,0,0,0,1,1,0,0,0,0,0,0,0,1,1,1,1,0,0,1,1,1,1
1,1,1,0,0,0,0,0,0,0,0,1,1,0,1,1,0,1,0,0,0,1,0,0,0,0,1,0,0,0,1,0,1,1
1,1,0,1,0,0,0,0,0,0,0,0,1,1,1,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,1,0,1,1
1,1,1,1,1,0,1,0,0,0,1,0,0,0,1,1,0,0,0,0,0,0,0,0,1,1,1,1,0,0,1,1,1,1
1,1,1,0,1,0,0,0,0,0,0,0,0,1,1,0,1,0,0,0,0,0,0,0,1,0,0,0,1,0,1,1
```

#### 3.4.2 Date based web log file to ARFF Conversion

```
@relation UniversityDetails
@attribute 'http://alagappauniversity.ac.in/about_au.php' {1,0}
@attribute 'http://alagappauniversity.ac.in/administration/vicechancellor.php' {1,0}
@attribute 'http://alagappauniversity.ac.in/administration/registrar.php' {1,0}
@attribute 'http://alagappauniversity.ac.in/administration/coe.php' {1,0}
@attribute 'http://alagappauniversity.ac.in/administration/financeofficer.php' {1,0}
@attribute 'http://alagappauniversity.ac.in/administration/stand_committee.php' {1,0}
@attribute 'http://alagappauniversity.ac.in/administration/finance_committee.php' {1,0}
@attribute 'http://alagappauniversity.ac.in/administration/IQAC.php' {1,0}
.....
.....
.....
@attribute 'http://alagappauniversity.ac.in/dde/faculty_members.php' {1,0}
@attribute 'http://alagappauniversity.ac.in/dde/grievance.php' {1,0}
@attribute 'http://alagappauniversity.ac.in/dde/contactus.php' {1,0}
@attribute 'http://alagappauniversity.ac.in/dde/app_forms.php' {1,0}
@attribute 'http://alagappauniversity.ac.in/dde/syllabus.php' {1,0}
@ data
1,0,0,1,0,0,1,0,1,0,1,0,0,0,0,1,1,1,0,0,0,0,1,0,0,1,0,1,0,1,0,1,0,1,1
1,1,1,1,1,0,1,0,1,1,1,0,0,1,0,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,1,0,1,1
1,1,0,1,0,1,0,1,0,1,0,1,1,0,0,1,1,0,1,0,1,0,1,0,1,0,1,1,1,0,0,1,1,1,1
1,1,1,0,0,0,0,0,1,1,0,0,0,1,0,0,1,0,1,0,1,0,1,0,0,1,0,0,1,0,1,1,0,0,1
1,0,1,1,0,1,0,1,0,1,0,1,0,1,0,1,1,0,1,0,1,0,1,0,1,0,1,0,0,1,0,1,0,1
0,1,1,0,1,0,1,0,1,0,1,0,0,1,0,0,1,0,1,0,1,0,1,1,0,1,0,1,0,1,0,1,1,0,1
1,0,1,1,0,1,0,1,0,1,0,1,0,0,0,1,0,1,1,0,1,0,1,0,1,0,1,0,0,1,0,0,1,0,1,0
0,1,0,1,1,0,1,0,1,0,1,0,1,0,0,1,1,0,1,0,1,0,1,0,1,0,1,1,0,1,0,1,0,0,1
0,1,0,1,0,0,1,0,1,0,0,1,0,1,0,1,1,0,1,0,1,0,0,1,0,1,0,1,0,1,0,0,1
1,0,1,0,1,0,1,0,1,0,1,1,0,1,0,0,1,0,1,0,1,1,0,1,0,1,0,0,0,0,1,0,1,1
1,1,1,1,1,1,0,0,1,0,0,1,0,0,0,1,1,0,1,0,0,0,0,0,1,0,1,0,1,0,0,1,1,1,1
1,1,0,1,0,0,0,0,0,0,0,0,0,0,1,0,1,1,0,1,0,0,0,0,0,0,1,0,0,0,1,0,1,1
```

1,1,1,1,0,0,1,0,1,1,0,0,0,1,0,0,1,1,1,0,1,0,1,0,0,0,0,0,0,1,1,1,0,1,1  
 1,1,1,0,0,0,0,0,0,0,0,0,0,1,0,1,1,1,1,0,1,1,1,0,0,0,0,1,0,0,0,1,0,1,1  
 1,1,0,1,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,0,1,1,0,0,0,0,0,1,0,1,0,1,0,1,1  
 1,1,1,1,1,1,0,1,0,0,0,1,1,0,0,0,1,1,1,0,0,0,0,0,1,0,1,0,0,1,1,1,1  
 1,0,0,1,0,0,0,0,0,0,0,0,0,1,1,1,0,0,0,1,0,0,0,0,0,0,0,0,0,0,1,0,1,1  
 1,1,1,0,1,0,0,0,1,0,0,0,0,0,0,1,1,1,1,1,1,1,1,0,1,0,0,0,0,0,1,0,1,1

The analysis has been conducted on the web log data that consists of the web pages visited for given month. The first process in the web log mining is to clean the irrelevant data from the database and automatic web requests prior to running association rule discovery algorithm, as well as to

group web requests into visitor sessions. The output frequent Itemsets generated for each of the dataset is given below. The graph shows the instance of the frequent Itemsets generated in both type of transactions.

**IPAddress based, Frequent Item set Generated.**

Minimum support: 0.95 (29 instances)

**Frequent Item set:**

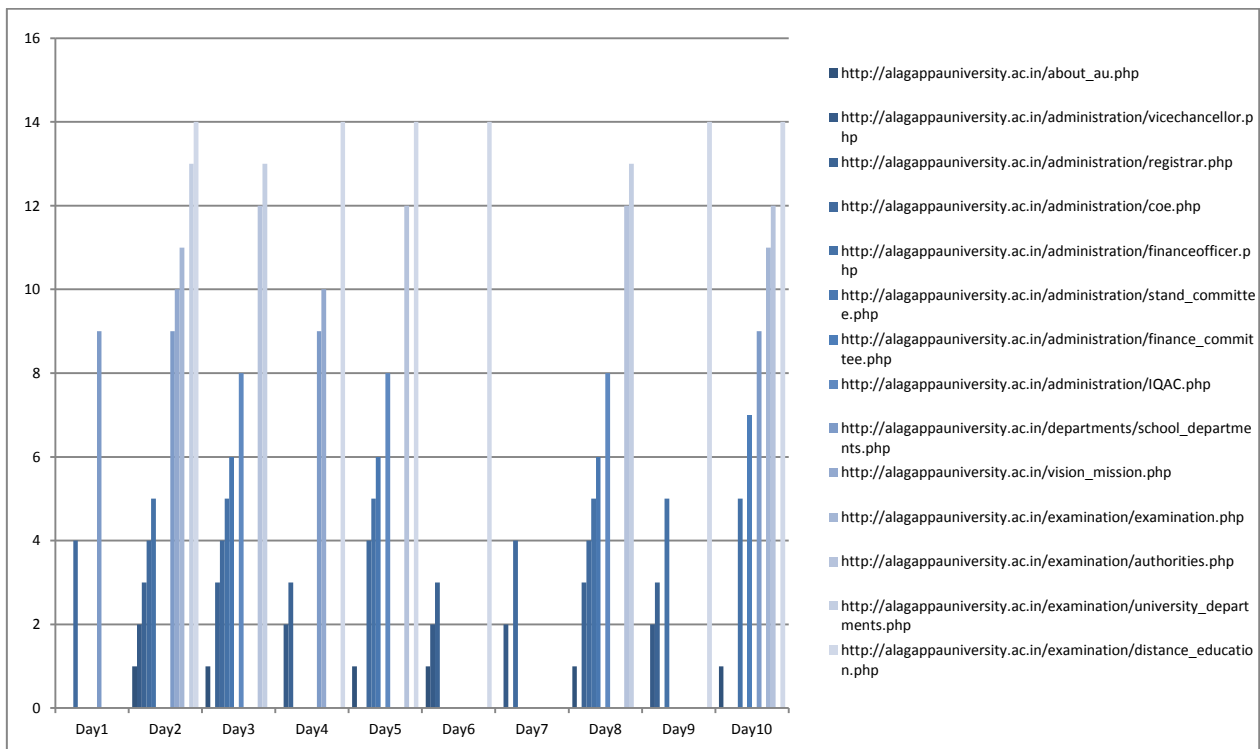
- [http://alagappauniversity.ac.in/examination/model\\_question\\_paper.php](http://alagappauniversity.ac.in/examination/model_question_paper.php)
- <http://alagappauniversity.ac.in/dde/grievance.php>
- <http://alagappauniversity.ac.in/dde/syllabus.php>

**Date based Frequent Item set Generation:**

Minimum support: 0.95 (29 instances)

**Frequent Itemsets:**

- [http://alagappauniversity.ac.in/about\\_au.php](http://alagappauniversity.ac.in/about_au.php)
- <http://alagappauniversity.ac.in/administration/vicechancellor.php>
- [http://alagappauniversity.ac.in/examination/model\\_question\\_paper.php](http://alagappauniversity.ac.in/examination/model_question_paper.php)
- <http://alagappauniversity.ac.in/dde/grievance.php>
- <http://alagappauniversity.ac.in/dde/syllabus.php>



**Fig .1 Graphical Representation of Number of Visitors Generated for Date based.**

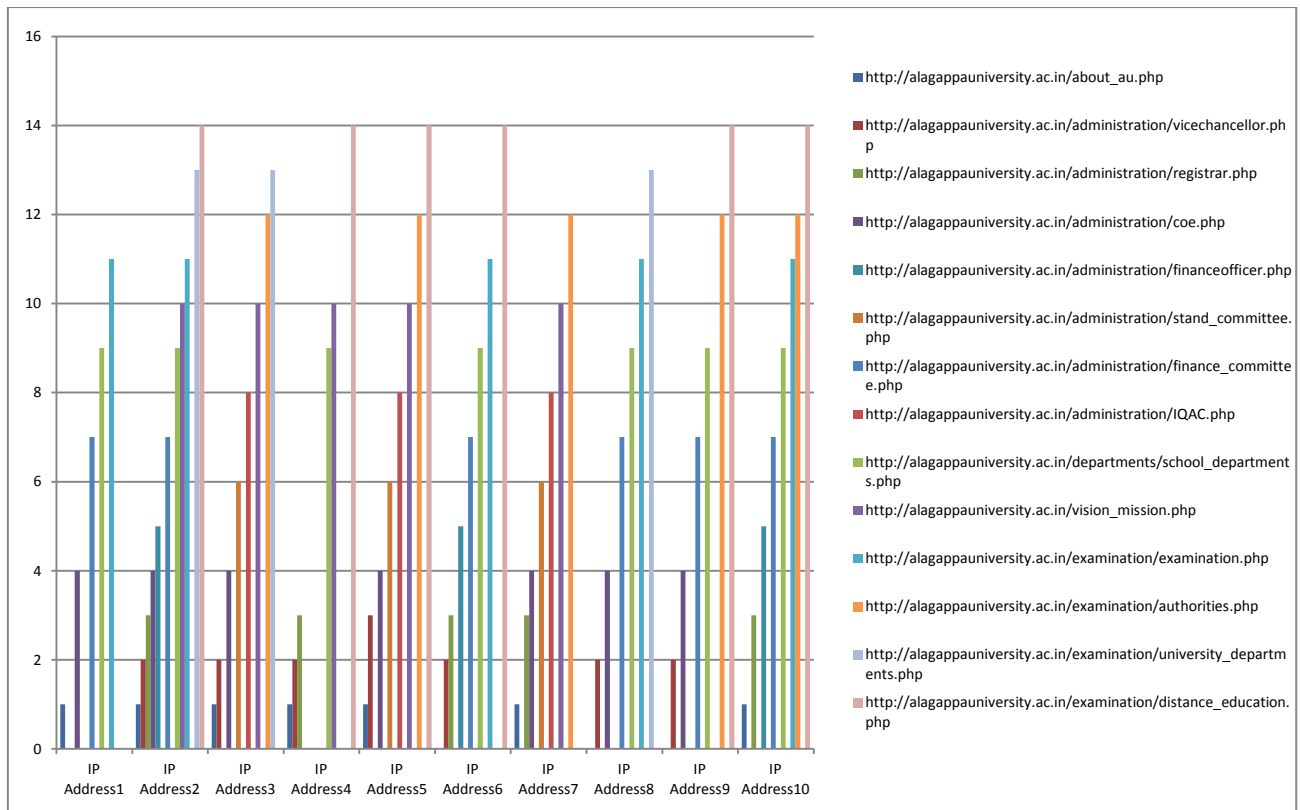


Fig. 2 Graphical Representation of Number of Visitors Generated for IP Address based.

## 4. CONCLUSION

The proposed methodology works with the objective of creating frequent Itemsets in the large transactional databases, to knowledge discovery with the help of data mining techniques. The preprocessing step removes the irrelevant data from the database. The transactions were categorized based on the IP address as well as the Date stamping. It creates a new view of web log data for frequent item set generation. From the obtained result, nearly for 30 days of web log data, 5 - 10 pages were frequently visited by the user for the considered university dataset. The frequent item set is generated with minimum support and confidence instance using the Apriori algorithm. In this proposed work, rather than existing methods the same dataset can be viewed at multidimensional and same dataset at multiple axes are used.

## 5. REFERENCES

- [1] Ankit R Kharwar, Viral Kapadia, Nilesh Prajapati, Premal Patel, "Implementing APRIORI Algorithm on Web server log", National Conference on Recent Trends in Engineering & Technology, B.V.M. Engineering College, V.V. Nagar, Gujarat, India, 13-14 May 2011.
- [2] Ashok Kumar.D and Loraine charlet Annie M.C, "Web Log mining Using K - Priori Algorithm", in International Journal of Computer Applications, Vol 41, No 11, March 2012, pp: 16-20.
- [3] Han J., Pei J., Yin Y, "Mining frequent patterns without candidate generation: A frequent-pattern tree approach "Data Mining and Knowledge, 2003.
- [4] Ilayaraja M and Meyyappan T, "Mining medical data to identify frequent diseases using Apriori algorithm", Pattern Recognition, Informatics and Mobile Engineering (PRIME), 2013 International Conference, ISBN:978-1-4673-5843-9, 21-22 Feb. 2013.
- [5] Jiawei Han, Micheline Kamber. Data Mining Concepts and Techniques, Beijing: China Machine Press, 2001, p290-297
- [6] Jogi.Suresh and T.Ramanjaneyulu, "Mining Frequent Itemsets Using Apriori Algorithm", International Journal of Computer Trends and Technology (IJCTT) - volume4 Issue4 - April 2013
- [7] Kirti S. Patil, Sandip S. Patil, "Sequential Pattern Mining Using Apriori Algorithm & Frequent Pattern Tree Algorithm", IOSR Journal of Engineering (IOSRJEN)-ISSN: 2250-3021, p-ISSN: 2278-8719 Vol. 3, Issue 1 (Jan. 2013), ||V4|| PP 26-30.
- [8] Latheefa.V and Rohini.V, "Web Mining Patterns Discovery and Analysis Using Custom-Built Apriori Algorithm", International Journal of Engineering Inventions, e-ISSN: 2278-7461, p-ISSN: 2319-6491, Volume 2, Issue 5 (March 2013) PP: 16-21.
- [9] Pooja Sharma and Rupali Bhartiya, "An Efficient Algorithm for Improved Web Usage Mining", in International Journal of Computer Technology and Applications, Volume 3, Issue:002, pp.766-769.

- [10] Sandeep Singh Rawat, Lakshmi Rajamani, "*Discovering Potential User Browsing Behaviors Using Custom-Built Apriori Algorithm*", International journal of computer science & information Technology (IJCSIT) Vol.2, No.4, August 2010.
- [11] B.Santhosh Kumar, K.V.Rukmani, "*Implementation of Web Usage Mining Using APRIORI and FP Growth Algorithms*", Int. J. of Advanced Networking and Applications Volume:01, Issue:06, Pages: 400-404 (2010).
- [12] Wang Tong, HE Pi-lian, "*Web Log Mining by an Improved AprioriAll Algorithm*", World Academy of Science, Engineering and Technology 4 2007.
- [13] Quingyu Zhang and Richard s.Segall, 2008, "*Web Mining: A Survey of current research, Techniques and Software*", in International Journal of Information Technology and Decision Making, Volume.07, Issue: 04, pp-683-720.