

**“FINDING USER PREFERENCE AND ANOMALY FROM WI-FI LOGS
USING HIERARCHICAL DIRICHLET PROCESS”**¹M.Subhashini, ²Dr.V.Kathiresan¹M.Phil Scholar, Dept. of Computer Science, Dr.SNS College of Arts and Science, Coimbatore, Tamil Nadu, India²Head Of the Dept., Dept. of Computer Application(PG), Dr.SNS College of Arts and Science, Coimbatore, Tamil Nadu, India

ABSTRACT - In recent trends mobile devices are very popular and the growth of those mobile devices is tremendous in nature. The proposed system helps to detect anomaly and user preferences by using Wi-Fi logs from a mobile device. There is a massive challenge to perform the elimination of auxiliary information's and performing user summary from Wi-Fi logs.

Keywords: Web Mining, Weblog Mining, Wi-Fi Log Mining

I. INTRODUCTION

1.1 INTRODUCTION TO WEB MINING

With more than two billion pages being created by millions of Web page authors and organizations, the World Wide Web (WWW) is growing tremendously as a rich knowledge base. The main demand in extracting knowledge from web mining is the size of the Web and its unstructured and unstable content with its multilingual nature [1].

1.2 SCOPE OF THE RESEARCH:

In web mining process especially in web usage mining, many existing system proposed with the analysis of web log without the location access. So, analyzing user profiles from social, Wi-Fi and web log will lead the research in an effective manner.

1.3 PROBLEM DEFINITION:

There is several previous works for finding user preference from web log. But there is an only one paper, which handles Wi-Fi logs to detect user preferences. Analyzing and preprocessing the Wi-Fi logs is very challenging task.

- Dimensionality problem : arise when analyzing and organizing data in high-dimensional spaces
- Maximum coverage problem: computational complexity and this type of process are more suitable in the theoretical way and it needs more resource when implementing practically.

2. LITERATURE REVIEW

2.1 REVIEWS ON DATA PREPROCESSING IN WEBLOG MINING

Author Castellano *et al.* [10] proposed a new approach for data preprocessing of web logs. Here the author proves the pre-processing is the first stage of web usage mining. In particular, the author presents LODAP (LOGData Preprocessor).

Author Ankit *et al* [11] has introduces many data pre-processing techniques helps to identify unique users and user data session; this can be used to improve the performance features in web usage mining.

Reddy, K. Sudheer, M. Kantha Reddy, and V. Sitaramulu [12] analyzed and summarized about various details about data preprocessing activities that are necessary to perform Web Usage Mining (WUM).

2.2 REVIEWS ON FEATURE EXTRACTION IN WEB LOG MINING

There are several feature extraction techniques available for web log mining process. The techniques such as association rule mining algorithms like Apriori, FPgrowth and EClat.

Wang, Hengshan, Cheng Yang, and Hua Zeng., [14] introduced two prevalent data mining algorithms - FPgrowth and PrefixSpan into web usage mining. Authors enhance the FP growth into Maximum Forward Path (MFP)..

Sandeep Singh Rawat *et al.*, [15] proposed a custom-built apriori algorithm which is based on the existing Apriori algorithm, to find the effective patterns from the web log.

3. PROPOSED SYSTEM

3.1 INTRODUCTION

In the current scenario, Wi-Fi based web access is useful for almost every activity. So there is a rapid development of World Wide Web (WWW) in its volume of traffic and the size and complexity of web sites.

3.1.1 Contributions of the proposed system:

A Wi-Fi server usually registers a log entry for every access of a mobile user across the network. First, raw Wi-Fi log data needs to be cleaned, condensed and transformed in order to retrieve and analyze significant and useful information. Second, pattern mining can be performed on log records to find association patterns, sequential patterns and trends of web accessing. The followings are the contributions of the proposed system.

- The **Session based Visiting Pattern (SVP)** analysis Algorithm is proposed to identify user and session are very important for identifying behavioral patterns.
- **FLAME (Fuzzy clustering by Local Approximation of MEmbership)** clustering algorithm is proposed to help in identifying very relevant similar groups
- The proposed **Hierarchical Dirichlet Process** strategy is used for finding and pruning unwanted SSID and log details. And it also helps to detect every pattern from SVP.

3.1.2 Features of the proposed system:

1. The proposed system able find individual user behavior and preference from WiFi logs
2. From the results of the proposed system, the suggestion and restriction of access can be made within the campus/Wi-Fi area.
3. The proposed system can effectively detect Anomaly from the user log and summarizes the abnormal activities

This chapter focuses on providing techniques for better Wi-Fi data handling and feature extraction from the Wi-Fi and web log. Wi-Fi Log data is usually noisy and ambiguous and preprocessing is an important process for efficient mining procedure.

Fig 3.1 Overview of the proposed system

3.2 METHODOLOGIES

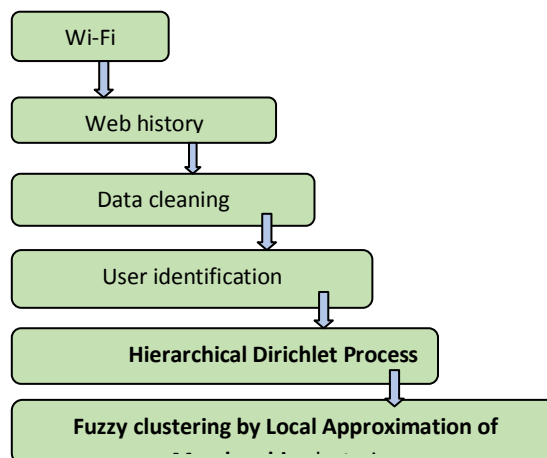
The chapter discusses about the methodologies included in the proposed system. The system flow and the desired algorithms explain the proposed method in detail.

3.2.1 Introduction:

The next primary goal is to learn the user's location patterns and their use of web resources in web log mining. After the process of data cleaning extract features from the accessible transactions in the log. Then the last stage in the preprocessing technique is the feature extraction process in which the feature sets are extracted using the methods

- SSID extraction
- SSID service details extraction

Following figure 3.2 describes the main techniques information of unique



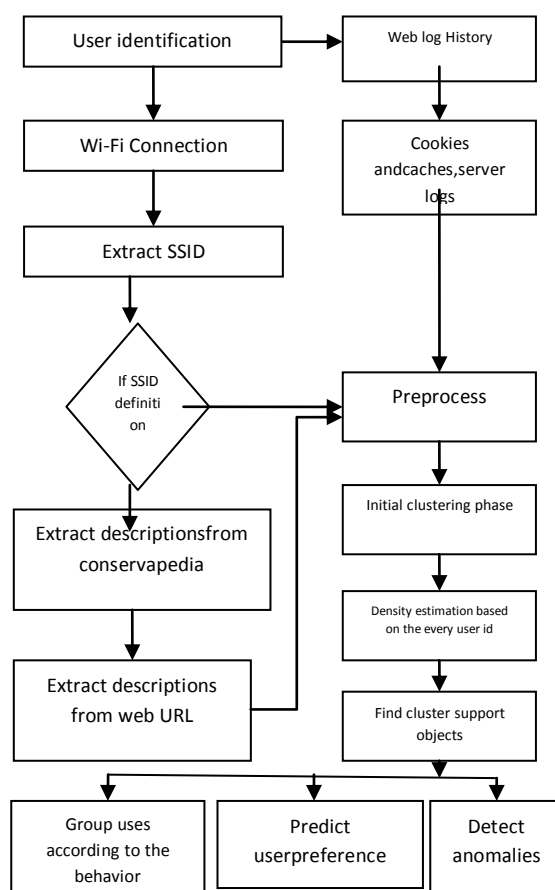


Fig 3.2 detailed process of the proposed work

To model user location based behavior, the initial thing is an identification of unique user from raw log file. Unique user identification is very challenging and complex task because of issues like:

- (1) One user uses multiple mobiles and systems to access the internet
- (2) Some Wi-Fi hotspots are not properly described.
- (3) Proxy server may hide details of unique users

3.2.1.1 User Identification process:

The user detection and their preference prediction need a set of user history and Wi-Fi logs. The followings are the tools and software used to crawl the Wi-Fi and Web logs.

a) WifiHistoryView:

The WifiHistoryView software is used to crawl the Wi-Fi connection history from hotspot, which connected the specific hotspot on the system. For every event that the computer connected to a wireless network or disconnected from it, the following information is displayed:

- The date/time - that the event occurred,
- network name (SSID)
- profile name
- network adapter name
- BSSID of the router/Access Point, and more.

b) Cookies and caches:

The another technique used to extract the web log is cookies and caches method. The basic information present in the log file is as follows:

- **User Id:** This identifies who visited the web sites

- **Link:** The path taken by the user while searching and viewing the website.
- **Set of links visited:** This identifies the path taken by the user within the web site using the various links.
Ex:<https://articles.forensicfocus.com/2014/10/28/extracting-data.aspx>
- **Date time:** This is identified as the session. Session is the time variable.
- **URL:** The resource accessed by the user. It may be an HTML page, a CGI program, or a script.
- **Request Mode:** The method used for information transfer is noted. The methods like GET, POST.

3.2.1.2 Preprocessing phase

Data pre-processing is an important step to filter and organize the appropriate information before applying any algorithm. Pre-processing consist of following steps

- Field extraction
- Data cleansing
- User identification
- Session identification
- Semantic SSID detection

a) Field extraction:

Each connection and web log entry is represented as a single line of the log file. The log entry contains many fields as discussed in the earlier section 4.1.1 The server used different characters which work as separators. The most used separator character is '/', 'tab' and “character.

b) Delimiter based Field Extraction algorithm:

Input : Wi-Fi connection and web Log Files, delimiter L

Output: preprocessed data file (PDF)

Steps:

1. Open Log File
2. Read all fields contain in Server Log Files
3. Separate out the Attribute using the delimiter L
4. Extract all fields and Add into the PDF
5. Close and end process

An example of PDF is shown in fig 4.2.. The detailed description of the letters mentioned in the table is given below the table.

Wi-Fi Connection Log					
Time	ssid	wlevel	bssid	cid	
12/14/2015	Edman	-72	88c7c879cbbd2	1	
12/14/2015	Dominos	-72	5c6b86c99849a	2	
12/14/2015	McDonalds	-74	6466b34e6b30	3	
12/14/2015	Dominos	-76	744d03b886d1e	4	
12/14/2015	wireless	-94	0013071b0c863	5	
12/14/2015	Edman	-90	88d11175645a	6	

Wi-Fi Web Log						
no	cid	mainlink	sublinks	ipadd	datetime	uname
1	1	https://in.yahoo.com/	https://in.yahoo.com/	192.168.1.86	11/13/2015 6:22..	hahai
33	2	https://www.codproject.com/	/	192.168.1.82	11/14/2015 1:49..	hena
71	9	https://in.yahoo.com/	http://moviesadrv.com/bollywood..	192.168.1.86	11/13/2015 6:22..	hena
129	5	https://in.yahoo.com/	https://in.moviesadrv.com/post..	192.168.1.87	11/13/2015 6:24..	hena
182	27	https://in.yahoo.com/	https://in.yahoo.com/	192.168.1.2	11/13/2015 6:25..	hena
241	14	https://in.yahoo.com/	http://portal.kiplinger.com/shibeh..	192.168.1.7	11/13/2015 6:26..	hena
243	2	https://www.codproject.com/	https://searchapi?doe=hs0&id=502	192.168.1.23	11/14/2015 1:49..	hena
442	8	https://www.codproject.com/	https://searchapi?doe=hs0&id=101	192.168.1.40	11/14/2015 1:54..	hena
543	16	https://in.yahoo.com/?p=us	https://cricket.yahoo.com/news/..	192.168.1.82	11/14/2015 7:17..	hena

In Data cleansing process, following steps are performed

- Remove unwanted attributes, which are not necessary for the work
- Remove log entry nodes contains extension like jpg, gif means remove request such as multimedia files, image, page style file

Algorithm for Data Cleansing at the time of SSID and URL extraction

Steps:

1. Read Entries in PDF (Pre-processed Data File)
 - a. For each Entry in PDF
 - b. Read fields (ssid)
2. If ssid != ' ' and mainlink != ' ' Then
 - a. Get IP_address and URL_link
 - b. If suffix.URL_Link= {*.gif,*.jpg,*.css} Then
 - c. Remove suffix.URL_link
 - d. SSID,Ip address and URL_Link
 - e. End if
3. Else
 - a. Next EntryEnd if

CLUSTERING MODELS

The similar user behaviors are grouped and categorized their SSID based preference this phenomenon is known as clustering process. It is an unsupervised learning technique, which does not need any labeled training samples. The successive clustering methods can be found by density based algorithms using neighbor details. These algorithms are commonly follows the neighbor details and calculate the distance to perform the clustering.. The proposed system used improved I-FLAME (**Fuzzy clustering by Local Approximation of Memberships**) clustering.

The FLAME algorithm is mainly divided into three steps:

1. Extraction of the structure information from the weblog dataset:
 1. Construct a neighborhood graph to connect each object to its K-Nearest Neighbors (KNN) in the given datasets D;
 - a. Based on the IP, the neighborhood graph will be constructed.
 2. Estimate a density for each object based on its neighbors to its KNN; number of neighbors for each user based on IP and query.
 3. Users and Objects are classified into 3 types:
 - a. Cluster Supporting Object (CSO): object with density higher than all its neighbors; Based on the weblog, the system finds highest frequency link.
 - b. Cluster Outliers: object with density lower than all its neighbors, and lower than a predefined threshold;
 - c. the rest are grouped as the other category
2. Local/Neighborhood approximation of fuzzy memberships:
 1. Initialization of fuzzy membership:
 - i. Each CSO is assigned with fixed and full membership to itself to represent one cluster;
 - ii. All outliers are assigned with fixed and full membership to the outlier group;
 - iii. The rest are assigned with equal memberships to all clusters and the outlier group;
 2. Then the fuzzy memberships of all type 3 objects are updated by a converging iterative procedure called *Local/Neighborhood Approximation of Fuzzy Memberships*, in which the fuzzy membership of each object is updated by a linear combination of the fuzzy memberships of its nearest neighbors.
3. Cluster construction from fuzzy memberships in two possible ways:
 1. One-to-one object-cluster assignment, to assign each object to the cluster in which it has the highest membership;

- One-to-multiple object-clusters assignment, to assign each object to the cluster in which it has a membership higher than a threshold.

4. EXPERIMENTS AND RESULTS

4.1 EXPERIMENTS

4.1.1 DATA SETS:

In the implementation dynamic synthetic datasets have been used to evaluate the proposed system.

Time	ssid	wlevel	ssid	sid
12/14/2...	Edimax	-94	20:cf:30:87:dd:3b	27
12/14/2...	Edimax	-96	0c:82:68:34:90:22	28
12/14/2...	TINASONIC	-79	00:1f:c6:27:e9:ce	29
12/14/2...	Edimax	-87	00:24:a5:34:0f:96	30
12/14/2...	Dominos	-73	74:d0:2b:88:6d:1c	4
12/14/2...	wenshan	-94	00:13:d7:1b:c8:63	5

Fig 4.1 Wi-Fi Connection dataset sample

eno	keywords	mainlink	sublinks	ipadd	datetime	uname
18	dotnet	http://localhost:5942/FCLAM/articles/	http://localhost:5942/FCLAM/articles/Dotnet/	192.168.1.34	10/16/2015 6:20:48 PM	r
19	data	http://localhost:5942/FCLAM/articles/	http://localhost:5942/FCLAM/articles/	192.168.1.31	10/16/2015 6:21:02 PM	aa
20	data	http://localhost:5942/FCLAM/articles/	http://localhost:5942/FCLAM/articles/DataMining/	192.168.1.31	10/16/2015 6:21:02 PM	Admin
21	university	http://localhost:5942/FCLAM/berlin/	http://localhost:5942/FCLAM/study/berlin/	192.168.1.34	10/16/2015 6:21:44 PM	corvitz1
22	university	http://localhost:5942/FCLAM/berlin/	http://localhost:5942/FCLAM/study/berlin/	192.168.1.56	10/16/2015 6:21:53 PM	ss
23	university	http://localhost:5942/FCLAM/berlin/	http://localhost:5942/FCLAM/study/berlin/	192.168.1.74	10/16/2015 6:21:57 PM	sample
24	university	http://localhost:5942/FCLAM/berlin/	http://localhost:5942/FCLAM/study/berlin/	192.168.1.7	10/16/2015 6:21:57 PM	sd
25	distributed	http://localhost:5942/FCLAM/articles/	http://localhost:5942/FCLAM/articles/	192.168.1.15	10/16/2015 6:22:03 PM	Admin
26	distributed	http://localhost:5942/FCLAM/articles/	http://localhost:5942/FCLAM/articles/Distributed/	192.168.1.15	10/16/2015 6:22:03 PM	corvitz1
27	berlin	http://localhost:5942/FCLAM/berlin/	http://localhost:5942/FCLAM/study/berlin/	192.168.1.79	10/16/2015 6:22:24 PM	Admin
28	berlin	http://localhost:5942/FCLAM/berlin/	http://localhost:5942/FCLAM/study/berlin/	192.168.1.79	10/16/2015 6:22:24 PM	corvitz1
29	berlin	http://localhost:5942/FCLAM/berlin/	http://localhost:5942/FCLAM/study/berlin/	192.168.1.20	10/16/2015 6:27:55 PM	kumar
30	berlin	http://localhost:5942/FCLAM/berlin/	http://localhost:5942/FCLAM/study/berlin/	192.168.1.62	10/16/2015 6:28:03 PM	ss

Fig 4.2 web log sample

4.1.2 EXPERIMENTAL SETUP

All the experiments were run on a Intel dual core 2 duo processor machine with 2GB RAM. The machine was running windows 8 operating system and the clustering algorithm was implemented using the .Net framework development environment.

SOFTWARE REQUIREMENTS

Operating System : Windows 8,
Front End : Visual Studio.NET 2010
Coding language : C#.NET
Back End : SQL Server 2008

Time	ssid	wlevel	ssid	sid
12/14/2...	Edimax	-94	20:cf:30:87:dd:3b	27
12/14/2...	Edimax	-96	0c:82:68:34:90:22	28
12/14/2...	TINASONIC	-79	00:1f:c6:27:e9:ce	29
12/14/2...	Edimax	-87	00:24:a5:34:0f:96	30
12/14/2...	Dominos	-73	74:d0:2b:88:6d:1c	4
12/14/2...	wenshan	-94	00:13:d7:1b:c8:63	5

Fig 4.1 Wi-Fi Connection dataset sample

4.2 PERFORMANCE ANALYSIS:

The following information's are gathered from the proposed system experiments.

Top 5 frequently connected hotspots	
Connected hotspot	frequency
PSG	34
McDonalds	23
NIIT	12
pizzahut	32
hifi	44

Table 4.1 connected Wi-Fi hotspots and its frequency

The performance analysis collected different hotspot connection, which includes the different types of Wi-fi location types. The connected hotspot samples are given in the above table 4.1 .for the above table respective chart is created and displayed in fig 4.3.

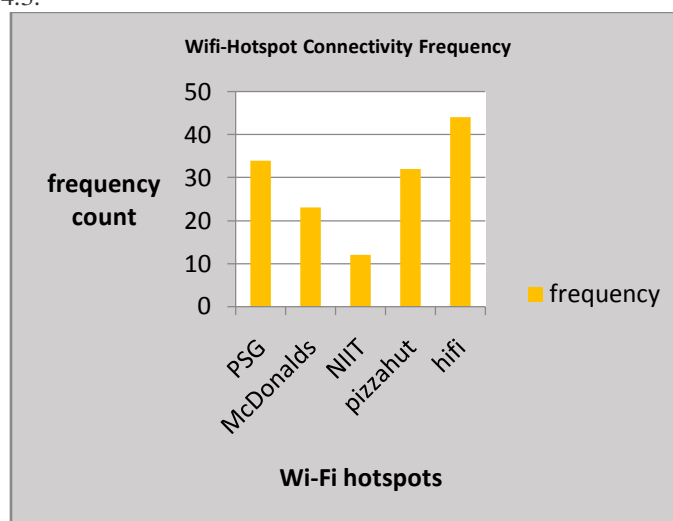


Fig 4.3 connected Wi-Fi hotspots and its frequency chart

Top 5 frequently visited pages and no of visitors	
Visit page	Visitors
games	498
twitter	487
yahoo	390
java	250
facebook	171

Table 4.2 connected Wi-Fi hotspots web pages and its frequency

Another set of data is collected web page access profile from different hotspot connection, which includes the different types of web pages and the total visiting patterns. The connected hotspot samples are given in the above table 4.2 .for the above table respective chart is created and displayed in fig 4.4

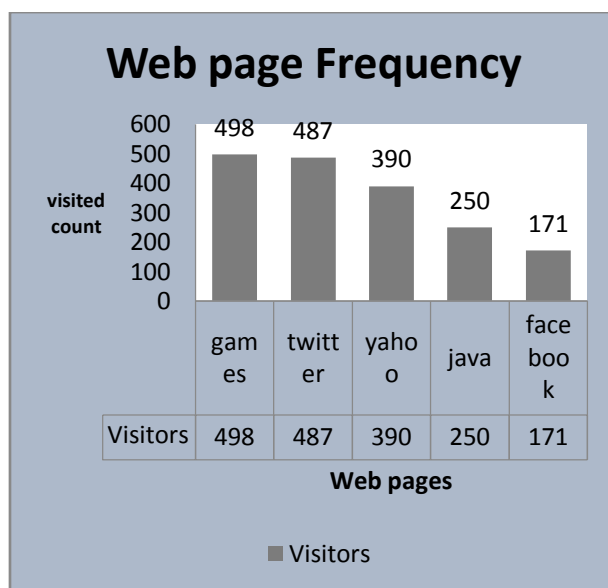
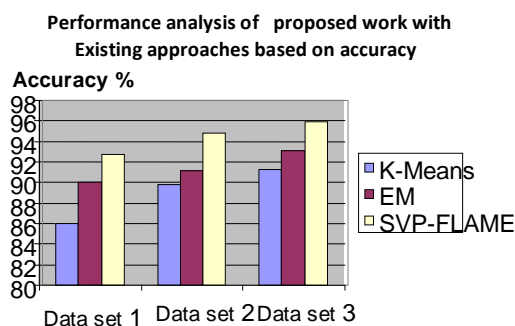


Fig 4.4 connected Wi-Fi hotspots web pages and its frequency

Table 4.5 Performance analysis in terms of accuracy

Techniques	Data set 1	Data set 2	Data set 3
K-MEANS	86.03	89.75	91.26
EM	90.08	91.14	93.09
SVP-FLAME	92.74	94.81	95.96



4.5 Performance analysis in terms of accuracy

The table 4.2 and figure 4.4 shows that the proposed SVP-FLAME performs better for all the three datasets with the highest accuracy of 92.74%, 94.81% and 95.96% than the K-MEANS and EM techniques.

Table 4.4 Performance analysis in terms of time taken

Techniques	Data set 1	Data set 2	Data set 3
K-MEANS	20	18	45
EM	22	20	52
SVP-FLAME	19	18	40

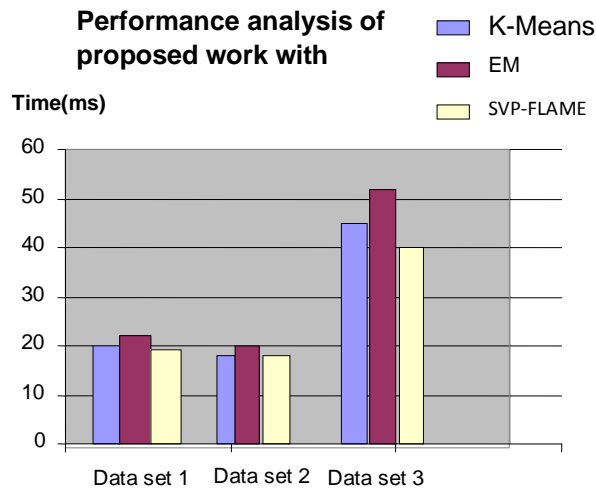


Figure 4.6 Performance analysis in terms of time taken

From the table 4.3 and figure 4.5, it is shown that the time taken by the proposed work is very less while comparing the other two existing techniques. The EM takes highest computation time.

Table 4.6: Performance analysis in terms of F-measure

Techniques	Data set 1	Data set 2	Data set 3
K-MEANS	0.63	0.69	0.75
EM	0.75	0.75	0.78
SVP-FLAME	0.82	0.85	0.87

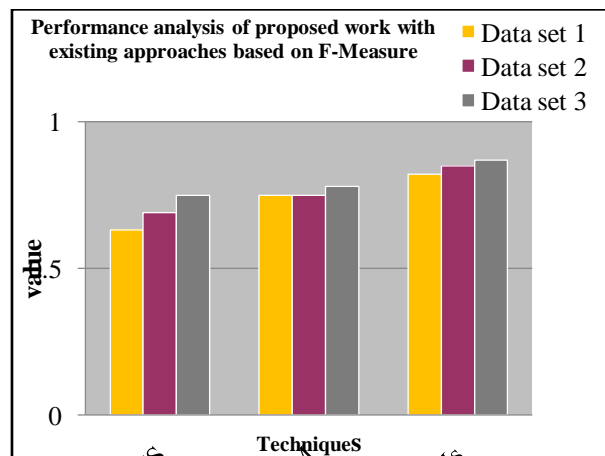


Figure 4.8 Performance analyses in terms of F-measure

From the table 4.6 and figure 4.8, it shows that the proposed work has highest F-measure rate 0.82, 0.85 and 0.87 for dataset 1, dataset 2 and dataset 3, respectively.

4. CONCLUSION

Internet and Mobile based services plays an important role in modern era. Several e-commerce websites and Business related services gathering web user access behavior and their profiles to improve the services. This thesis provides the techniques and methods used to detect user access behaviors from Wi-Fi connection log and web log.

FUTURE ENHANCEMENT

The proposed research work has tried to cover the basics of a user preference analysis process in this work, but there is plenty that can be done with this data. However this is the first work involved in this area and it is a time consuming process and ineffective when there is no descriptions for SSID.

REFERENCES:

- [1]. Shrivastava, Aditi, and Nitin Shukla. "Extracting Knowledge from User Access Logs." *International Journal of Scientific and Research Publications* 2.4 (2012): 1.
- [2]. Eirinaki, Magdalini, and Michalis Vazirgiannis. "Web mining for web personalization." *ACM Transactions on Internet Technology (TOIT)* 3.1 (2003): 1-27.
- [3]. Fu, Yongjian, Kanwalpreet Sandhu, and Ming-Yi Shih. "Clustering of web users based on access patterns." *Proceedings of the 1999 KDD Workshop on Web Mining*. San Diego, CA. Springer-Verlag, 1999.
- [4]. Cooley, Robert, Bamshad Mobasher, and Jaideep Srivastava. "Web mining: Information and pattern discovery on the world wide web." *Tools with Artificial Intelligence, 1997. Proceedings., Ninth IEEE International Conference on*. IEEE, 1997.
- [5]. Mobasher, Bamshad, et al. "Using sequential and non-sequential patterns in predictive web usage mining tasks." *Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on*. IEEE, 2002.
- [6]. Zaiane, Osmar R., and J. Luo. "Web usage mining for a better web-based learning environment." *Proceedings of conference on advanced technology for education*. 2001.
- [7]. Agosti, Maristella, Franco Crivellari, and Giorgio Maria Di Nunzio. "Web log analysis: a review of a decade of studies about information acquisition, inspection and interpretation of user interaction." *Data Mining and Knowledge Discovery* 24.3 (2012): 663-696.
- [8]. Tanasa, Doru, and Brigitte Trousse. "Advanced data preprocessing for intersites web usage mining." *IEEE Intelligent Systems* 19.2 (2004): 59-65.
- [9]. Thakare, Sanjay Bapu, and Sangram Z. Gawali. "A effective and complete preprocessing for Web Usage Mining." *International Journal on Computer Science and Engineering* 2.03 (2010): 848-851.
- [10]. Kaushik, Shailandra. "An overview of technical aspect for WiFi networks technology." *International Journal of Electronics and Computer Science Engineering (IJECSSE, ISSN: 2277-1956)* 1.01 (2012): 28-34.
- [11]. Castellano, G., A. M. Fanelli, and M. A. Torsello. "Log data preparation for mining web usage patterns." *IADIS International Conference Applied Computing*. No. 10000. 2007.
- [12]. Ankit R. K., A. N. Chandni and K. D. Niyanta, 2013. A Complete Pre Processing Method for Web Usage Mining, *International Journal of Emerging Technology and Advanced Engineering* , 3(10):638-640.
- [13]. Reddy, K. Sudheer, M. Kantha Reddy, and V. Sitaramulu. "An effective data preprocessing method for Web Usage Mining." *Information Communication and Embedded Systems (ICICES), 2013 International Conference on*. IEEE, 2013.
- [14]. Kumar, B. Santhosh, and K. V. Rukmani. "Implementation of web usage mining using APRIORI and FP growth algorithms." *Int. J. of Advanced Networking and Applications* 1.06 (2010): 400-404.
- [15]. Wang, Hengshan, Cheng Yang, and Hua Zeng. "Design and implementation of a web usage mining model based on upgrowth and prefixspan." *Communications of the IIMA* 6.2 (2015): 10.
- [16]. Sandeep, S. R. and R. Lakshmi, 2010, Discovering Potential User Browsing Behaviors Using Custom-Built Apriori Algorithm, *International journal of computer science and information Technology*, 2 (4): 28-37.
- [17]. Y. Zheng, X. Xie, and W.-Y. Ma, "Geolife: A collaborative social networking service among user, location and trajectory," *IEEE Data Eng. Bull.*, vol. 33, no. 2, pp. 32–39, Apr. 2010.
- [18]. Fan, Yao-Chung, et al. "A Framework for Enabling User Preference Profiling through Wi-Fi Logs." *IEEE Transactions on Knowledge and Data Engineering* 28.3 (2016): 592-603.
- [19]. Buchner, A. G. and M. D. Mulvenna, 1998. Discovering Internet Marketing Intelligence through Online Analytical Web Usage Mining, *ACM SIGMOD*, 27(4):54-61.
- [20]. Buchner, A. G., M. Baumgarten, S. S. Anand, M. D. Mulvenna, and J. G. Hughes, 1999, Navigation Pattern Discovery from Internet Data, In *WEBKDD*, San Diego. 1(1):74-91.

- [21]. Castellano G., A. M. Fanelli and M. A. Torsello, 2007. Log Data Preparation for Mining Web Usage Patterns, IADIS International Conference Applied Computing.
- [22]. Chang-bin, J. and L. Chen, 2010. Web Log Data Preprocessing Based On Collaborative Filtering. IEEE 2nd International Workshop on Education Technology and Computer Science: 118-121.
- [23]. Cheng, Y., and Fu, K, 1985. Conceptual clustering in knowledge organization. IEEE Transactions on Pattern Analysis and Machine Intelligence, 7(1):592-598.
- [24]. Chi E. H., A. Rosien and J. Heer, 2002. LumberJack: Intelligent Discovery and Analysis of Web User Traffic Composition. In Proceedings of ACM-SIGKDD Workshop on Web Mining for Usage Patterns and User Profiles, 2703(1): 1-16.
- [25]. Chidansh A. B. and S. K. Mohan, 2010. Multimedia Data Mining: State Of The Art and Challenges, Published Online: 16 November 2010, Springer Science Business Media, 51(1):35-76.
- [26]. Chitraa, V. and S. D. Antony, 2010. A Survey on Preprocessing Methods for Web Usage Data, (IJCSIS) International Journal of Computer Science and Information Security, 7(3): 1-12.
- [27]. Cooley R., B. Mobasher and J. Srivastava, 1999. Data preparation for mining World Wide Web browsing patterns. Knowledge Information Systems, 1(1): 5-32.
- [28]. Cooley R., B. Mobasher and J. Srivastava, 2000. Web Usage Mining: Discovery and Application of Interesting Patterns from Web Data, Ph.D Thesis, University of Minnesota.
- [29]. Cooley R., B. Mobasher and J. Srivastava, 1997. Web mining: Information and Pattern Discovery on the World Wide Web. In International conference on Tools with Artificial Intelligence, Newport Beach: 558-567.