

A KNOWLEDGE BASED METHODOLOGY TO UNDERSTAND THE USER BROWSING BEHAVIOR FOR QUALITY MEASUREMENT OF THE WEBSITES USING WEB USAGE MINING (A Comparative Study of Public Sector and Private Sector Banks)

Prof. Poonam Sawant¹ Dr. R.V. Kulkarni²

¹ Asst. Professor, Computer Department , Department of Computer , Singhgad Institute of Management and Computer Application, Pune, Maharashtra, India
poonam25m@gmail.com

² Professor and Head, Department of Computer , chh. Shahu Central Institute of Business Education and Research, Kolhapur, Maharashtra, India.
drvvkulkarni@siberindia.co.in

Abstract

World Wide Web is a huge repository of web pages and links. It provides huge information for the Internet users. The growth of web is tremendous as approximately one million pages are added daily. Users' accesses are recorded in web logs. Because of the tremendous usage of web, the web log files are growing at a faster rate and the size is becoming huge. Web data mining is the application of data mining techniques. Extraction of user behavior is an important work in Web Mining. Web Usage Mining applies mining techniques in log data to extract the behavior of users. Today most of the banks are entered in the web world. They are launching various online facilities day to day for customers. This paper presents the browsing behavior of the user in banks websites and the quality measurement factors of the website. In this study we have analyzed log server of three major public sector banks i.e. State Bank of India, Bank of Maharashtra and Bank of Baroda and three major private sector banks i.e. ICICI bank, HDFC bank and Federal bank. The obtained results can be used in different applications like, site modifications, business intelligence, system improvement, and web personalization and useful to tell, how to provide customer with more exactly needed and accurate on line information to improve banking activities and achieve business goal.

Keywords: *Data mining, Web Mining, Web Mining Taxonomy, preprocessing, data cleaning, Clustering*

Introduction:

Customer relationship management is one of the major task in banking sector and it is an important

applications of Web mining. As the various banks are using wide range of e-services a website should be designed to entice the customers. Due to increasing

competition between suppliers of goods and services, a strong relation between a provider and each one of its customers is nowadays more important than ever before. On the other hand, we are faced with tendencies that hinder direct customer contact, such as e.g. worldwide globalization, fusions of enterprises, and an increasing number of services ordering remote transactions, e.g. internet and phone banking. The challenge, banks are faced with, consists in finding ways to understand customer behavior, their preferences and desires in order to provide each customer excellent and personalized services at low costs. Web Mining analyses visitor's behavior and makes predictions on their future interaction. This can be exploited to improve website performance and to recommend policies or links based on user's behavior. Visitors entering the site exhibits different behavior. They might just surf through or the process might end up in a purchase. For understanding customer behavior and thus improve the performance of your web site, certain standards should be used. Web metrics provide a method to evaluate the performance.

This publication presents a methodology for web usage mining and its application for understanding user behavior in a private and public sector banks to measure quality and performance of websites. This paper provides an overview of the web mining and in particular web usage mining. In Section 3 we proposed a methodology for web usage mining with special emphasis on applications in banking areas. Section 4 shows the application of our methodology in a particular case, the identification of user behavior, measurement of quality and performance of

websites of private and public sector banks. Section 5 and 6 presents some recommendations and methodology to test recommendations respectively. Section 7 presents techniques used for analysis and Section 8 concludes this paper.

1. Literature Review:

Recently, many research projects are dealing with Web usage mining and Web personalization areas. Most of the efforts focus on extracting useful patterns and rules using data mining techniques in order to understand the users' navigational behavior, so that decisions concerning site restructuring or modification can then be made by humans. In several cases, a recommendation engine helps the user navigate through a site. Some of the more advanced systems provide much more functionality, introducing the notion of adaptive Web sites and providing means of dynamically changing a site's structure. All research efforts combine more than one of the aforementioned methods in Web personalization, namely, user profiling, Web usage mining techniques, content management and publishing mechanisms.

Masseglia et al. [1999] apply data mining techniques such as association rules and sequential pattern discovery on Web log files and then use them to customize the server hypertext organization dynamically. They regard Web usage mining as a two-phase process, consisting of the preprocessing phase where all irrelevant data are removed and log file entries are clustered based on time considerations, and the Web mining phase where data mining techniques are applied. The

prototype system, WebTool, also provides a visual query language in order to improve the mining process. A generator of dynamic links uses the rules generated from sequential patterns or association rules, and each time the navigation pattern of a visitor matches a rule, the hypertext organization is dynamically modified.

According to Usha P.M(2010) now almost all the banks have ventured into this area. Enormous amount of data gets stored through banking transactions. Success factor is the amount of valuable knowledge that is extracted from this data store. Customer profile can be generated. This helps the bank executives in identifying the appropriate customer for certain category of products and the risk in allotting loan facilities. Credit card usage patterns can be identified and special offers can be provided. Defaulters of payment can be identified easily. Banks like ICICI bank and HSBC bank identify the customers for certain offers like home equity loan using web mining.

2. Overview on Web Mining :

World Wide Web (WWW) is very popular and interactive. It has become an important source of information and services. In this internet era web sites on the internet are useful source of information in day to day activities. So there is a rapid development of World Wide Web in its volume of traffic and the size and complexity of web sites. The web is huge, diverse and dynamic. Extraction of interesting information from Web data has become more popular and as a result of that web mining has attracted lot of attention in

recent time. Web mining is an application of data mining to large web data repositories. It can be divided in to three categories namely web structure mining, web content mining and web usage mining. Web usage mining is the automatic discovery of user access patterns from web servers. Web usage mining is an important technology for understanding user's behaviors on the web and is one of the favorite areas of many researchers in the recent time. Web servers accumulate data about user's interactions in log files whenever requests for resources are received. Log files records information such as client IP address, URL requested etc., in different formats such as Common Log format, Extended Common Log format which is issued by Apache and IIS. Most of the servers use ECLF format. Obtained user access patterns can be used in variety of applications, for example, one can keep track of previously accessed pages of a user. These pages can be used to identify the typical behavior of the user and to make prediction about desired pages. Thus personalization for a user can be achieved through web usage mining.

Many commercial web log analyzer tools are available in the market that analyzes the web server log data to produce different kinds of statistics. In this study, web log expert program and Google web site analyzer tool have been used to analyze server log data of a website. Program generated different types of reports on server log data that can be useful from the point of view of the system administrator or web designer to increase effectiveness of the site.

3. A Proposed Methodology

Web users are facing the problems of information overload and drowning due to the significant and rapid growth in the amount of information and the number of users. As a result, how to provide Web users with more exactly needed information is becoming a critical issue in web-based information retrieval and Web applications. Though the banks provide Internet Banking Facilities for convenience, and to save the time of the customers, only 10% customers are using online banking facility due to some problems like finding relevant information, producing new knowledge, poor performance of the web pages etc. Based on suggestions from literature we developed a methodology shown in figure 1, which can be used for general applications of web usage mining. It is a behavioral adaptation process to the analysis of web data. Web data are a real source to analyze the user behavior in the web. An important step is cleaning and preprocessing of the web data.

The application of web mining techniques allows finding unknown patterns. These patterns must be validated or rejected by an expert in the business under investigation. Personalization of web sites is a very challenging field of both, current research as well as applications that have as goals e.g. individualized marketing for e-banking or dynamic recommendations to a web visitor based on his/her profile and usage behavior. Analyzing web data can also be used for system improvements providing the key to understanding web traffic behavior. Advanced load balancing, data distribution or policies for web caching as well as higher security standards are potential benefits of such improvements. Similar analyses could be used for modification of web sites. Understanding visitors' behavior in a web site provides hints for adequate design and update decisions. Business intelligence covers the application of intelligent techniques in order to improve certain businesses, mainly in marketing.

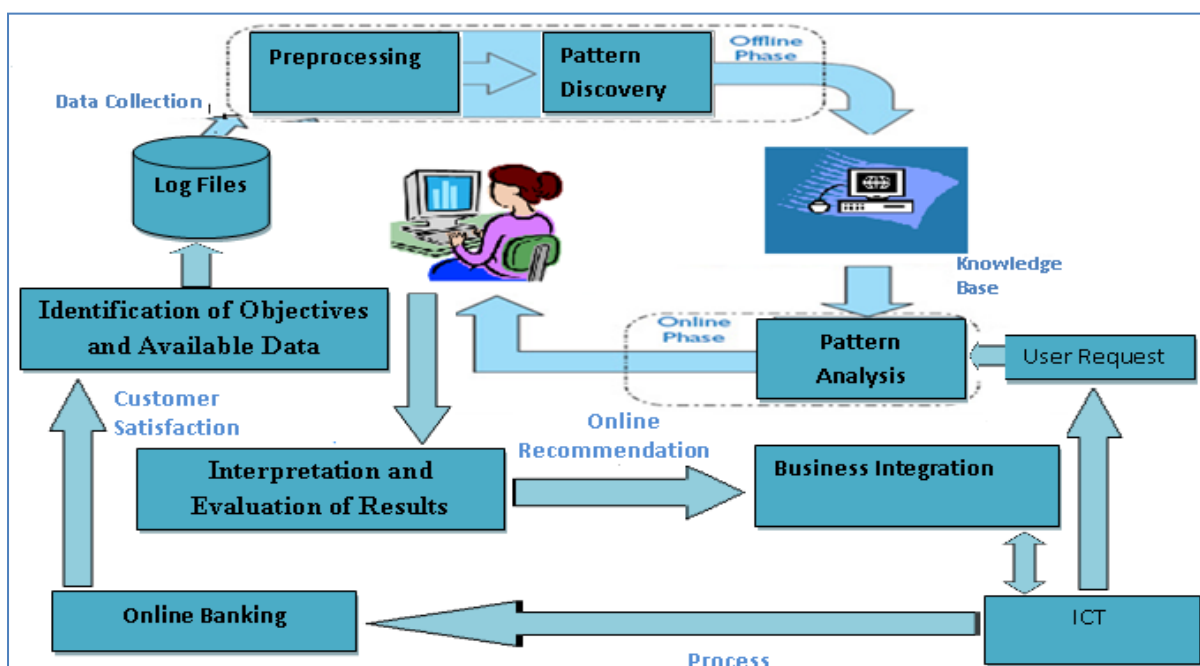


Fig1: Proposed Methodology

Our methodology consists of the following steps that build an iterative process as shown in figure

- 1.
1. Identification of objectives
2. Data Collection
3. Data preprocessing,
4. Pattern Discovery
5. Pattern analysis,
6. Interpretation and evaluation of results,
7. Online Recommendation and Business integration.

4.1 Identification of objectives and available data:

Both, objectives and available data play a fundamental role in the subsequent steps of the proposed methodology. Some possible objectives of web usage mining are:

- 1 To understand the user browsing behavior in websites of public sector and private sector bank.
- 2 To examine the quality and effectiveness of the websites.
- 3 To improve the relationship between user and websites .

In practice, often we are faced with some vague ideas a decision-maker has in mind and have to “translate” them into one or several of the objectives mentioned above. This is also the case of the application we are presenting below.

4.2 Data collection:

Data collection is the first step of web usage mining, the data authenticity and integrity will directly affect the following works smoothly

carrying on and the final recommendation of characteristic service’s quality. Therefore it must use scientific, reasonable and advanced technology to gather various data. At present, towards web usage mining technology, the main data origin has three kinds: server data, client data and middle data (agent server data and package detecting).

4.3 Data preprocessing:

Some databases are insufficient, inconsistent and including noise. The data pretreatment is to carry on a unification transformation to those databases. The result is that the database will to become integrate and consistent, thus establish the database which may mine. In the data pretreatment work, mainly include data cleaning, user identification, session identification and path completion.

4.4 Pattern Discovery:

Use statistical method to carry on the analysis and mine the pretreated data. We may discover the user or the user community's interests then construct interest model. At present the usually used machine learning methods mainly have clustering, classifying, the relation discovery and the order model discovery. Each method has its own excellence and shortcomings, but the quite effective method mainly is classifying and clustering at the present.

4.5 Pattern analysis:

Challenges of Pattern Analysis are to filter uninteresting information and to visualize and interpret the interesting patterns to the user. First delete the less significance rules or models from the interested model storehouse; Next use technology of OLAP and so on to carry on the comprehensive mining and analysis; Once more, let discovered data or knowledge be visible; Finally, provide the characteristic service to the electronic commerce website.

4.6 Interpretation and Evaluation of data

Depending on the previously defined objectives statistical analysis, web traffic analysis, and different data mining techniques can be applied. Since real-world applications often are guided by vaguely described objectives we generally have to apply a combination of data mining approaches.

4.7 Online Recommendation and Business Integration

The results from the web mining steps have to be integrated into the respective business. Applications of web usage mining for e-banking show some particularities, such as the possibility to perform web site structure and content modification, online marketing, place advertisements in a web site dynamically, among others.

5 Application of the proposed methodology

In this section we present the application of the proposed methodology to understand browsing behavior of the user and to find out quality and

effectiveness of bank websites to target the customers. A customer is a person that uses at least one service of the bank. Among all customers we want to analyse web usage behavior of the customers who are using bank's web site. Since last 15 years private and public sector banks are providing internet banking. But only 15 to 20 percent customers are using this facility. Online transactions cause typically less than 10% of traditional transaction costs. Most of the customers, however, are still not registered for online banking.

Objective of the presented work is to increase the use of internet banking by transforming the visitors into customers and retain the existing ones. The next task in this research was to identify the adequate web mining tasks in order to reach the mentioned goal.

Following are some questions that helped us to "translate" this business objective into web mining tasks.

- How do my customers should behave?
- How can we improve our web site?

In order to reach the above-mentioned objective we defined the following web mining tasks:

- Continuous improvement of the web site structure and content.
- Personalization of the relationship between the user and the web site.
- Understanding the user behavior in the web site.

5.1 Data Collection

In this study we have analyzed log server of three major public sector banks i.e. State Bank of India, Bank of Maharashtra and Bank of Baroda and

three major private sector banks i.e. ICICI bank, HDFC bank and Federal bank. From the log files we selected the visits to the web site from Dec 2011 to Feb 2012.

5.2 Preprocessing

We used Web Log Expert program and Google Analytic tool. In this duration log files have stored 20 MB data and we have got 2 MB data after preprocessing. We have determined,

- Total Percentage of visitors through referrers.
- Percentage of reach, page views per page, page view per user, bounce rate and search visit.
- Estimated time, Page load time, time spent in every visit and time spent in every page.
- Website performance factors like trustworthiness, vendor reliability, privacy and relevancy.
- Query Reputation and Impact Factor.
- Unique number of visitors etc.

5.3 Pattern Discovery

The above-mentioned business objective had been “translated” into two tasks which we reach in the following two steps:

- i) We have analyzed and compared user browsing behavior in both the sector’s websites by applying some statistical and web mining techniques.
- ii) We have compared the quality and performance of both the sector’s websites.

User browsing behavior can be examined on the basis of the search traffic of the web pages. A model for the visitor browsing behavior must consider,

- The visited pages, pages sequence, time spent and the.
- Percentage of reach, bounce rate and search visit.
- Percentage of page views per page and page views per user.
- Percentage of visitors through referrers.
- Query popularity, impact factor of queries and keyword density

Quality of the Website has been examined on the basis of keyword relevancy, trustworthiness, vendor reliability, privacy and unique visitors of the web site. Performance of the website is based on average number of urls, page load time, average time spends in every visit and time spends on each page. For both tasks, feature selection is necessary. From all attributes in the data warehouse we have to select the relevant features for segmentation and classification. In particular, we run the algorithm k-mean to determine the visitor behavior cluster. We run the STIRR clustering to determine search traffic in websites. We have applied Kolmogorov Smirnov Test to determine the percentage of visitors through referrers, Query Popularity and Query Impact Factor in the web site. We have applied Mann Whitney U Test and Huffman Coding Algorithm to determine quality and performance of the websites.

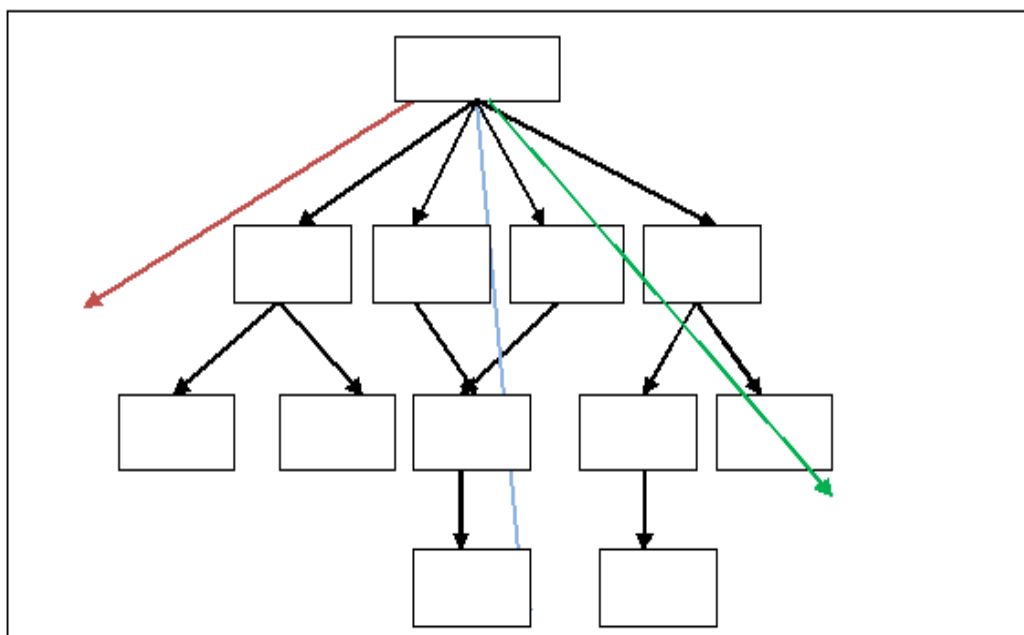


Fig.2: User Browsing Behavior

Table 1 shows the result of the STIRR cluster analysis for search traffic. Table 2 shows the result of Kolmogorov Smirnov Test to show the behavior of the customers through referrers. Table 3 and 4 shows the result of k-Mean clustering to analyzing the keywords on the web page. Table 5 and 6 also show the result of Kolmogorov Smirnov Test to determine the query popularity and query impact factor within the websites respectively. Table 7 shows the result of Mann Whitney U Test to determine quality of the website. Table 8 shows the result of Huffman Coding Algorithm to determine the performance

of the websites. Due to the competition between the banks we are not allowed to present the real number of transactions.

Test table 1 and fig 3 shows the percentage of search traffic in private and public sector banks' websites. Bounce Rate in private sector bank's websites is lower than public sector bank's websites. As compared to Private sector banks' websites percentage of search visit in public sector banks' websites is more but the percentage of reach is less than it. According to the statistics the browsing behavior of private sector bank's visitors is quiet good than public sector.

Test Table1 1 Search Traffic

Cluster	Traffic(%)	Private Sector	Public Sector
1	Bounce Rate	22.83	51.5
2	Search Visit	9	25.8
3	Page Views per Page	0.005	0.001
4	Page Views Per User	3.5	2.3
5	Reach	0.14	0.01

(Source:Analysis Work)

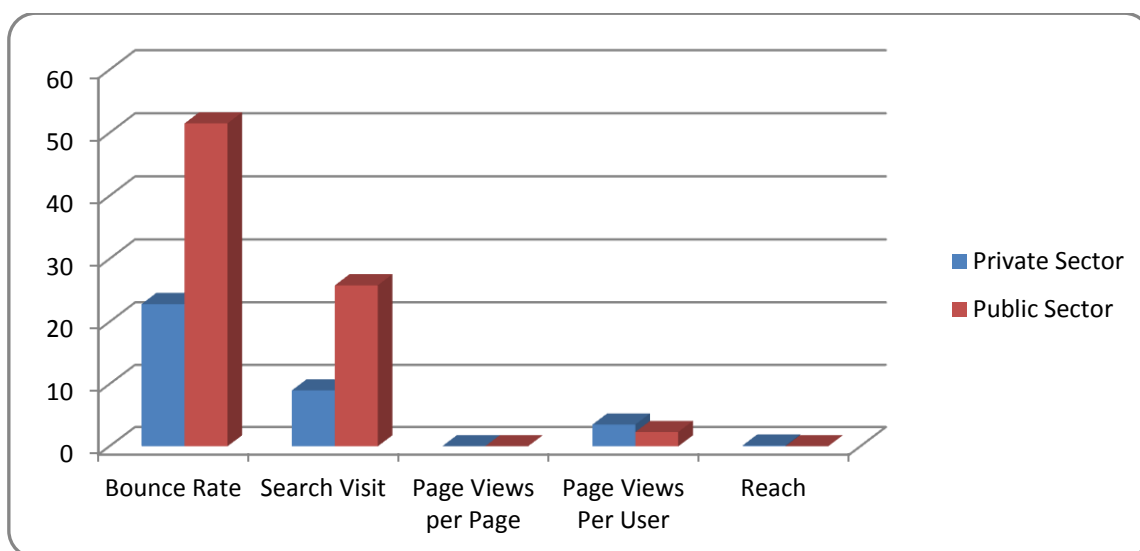


Fig 3. Search Traffic

Test Table 2 and fig 4 shows the percentage of the visitors through referrers. The visitors of private sector banks uses referrers more efficiently than

visitors of private sector banks for accessing the information from the websites of particular banks.

Test Table 2: Percentage of Visitors through referrers

Private Sector Banks					Public Sector Banks				
Code	f	cf	E	D cf-E	Code	f	cf	E	D cf-E
A	144	0.16	0.125	0.035	A	66	0.48	0.125	0.35
B	189	0.27	0.125	0.14	B	113	0.63	0.125	0.50
C	203	0.32	0.125	0.19	C	133	0.67	0.125	0.54
D	216	0.36	0.125	0.23	D	145	0.72	0.125	0.59
E	226	0.38	0.125	0.25	E	157	0.75	0.125	0.62
F	233	0.40	0.125	0.27	F	166	0.77	0.125	0.64
G	248	0.44	0.125	0.31	G	183	0.85	0.125	0.75
H	300	1	0.125	0.87	H	411	1	0.125	0.87
	n2=300					n1=411			
Z=0.078					Z=0.067				

(Source: Analysis Work)

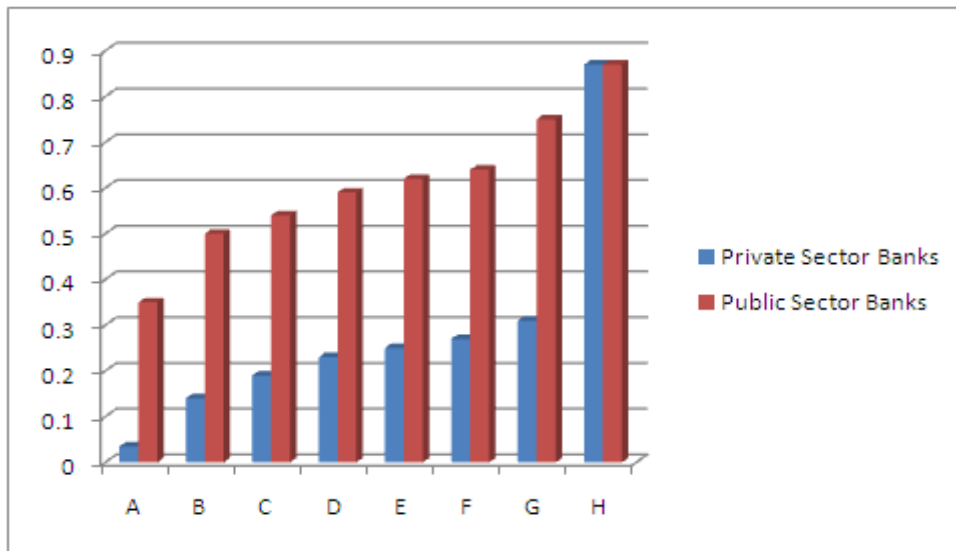


Fig 4: Percentage of Visitors through referrers

Table No.3 Cluster Analysis using K-Mean Clustering

Cluster	Keywords	Density
Private Sector		
1	w1,w2,w3,w5,w6,w7	28.93,6,4.76,3.19,2.93,2.16
2	w4,w8,w9,w11,w12	4.25,1.85,0.92,0.74,0.72
3	w14,w15,w16,w18	0.38,0.36,0.32,0.30
Public Sector		
1	w1,w2,w3,w4,w9	33,4.7,7.22,4.6,1.27
2	w5,w6,w7,w8,w12,w13,w14	2.34,1.93,1.53,1.27,0.76,0.63,0.53
3	w15,w16,w18	0.48,0.46,0.42

(Source: Analysis Work)

- The cluster must content only pages related by main topic.
- The business experts define the main topic per page.
- Reviews which words in each page cluster are the same.

Table No.4 Some Identified Keywords

No.	Keywords	No.	Keywords
1	Name	8	Online Banking
2	Login	9	Payment
3	Credit card	10	Car loan
4	Internet Banking	11	Customer Care

5	Net Banking	12	Bank.com
6	Personal loan	13	Money to India
7	Home loan	14	Billdesk

(Source : Analysis Work)

Test Table 5: Impact Factor of Query

Private Sector Banks					Public Sector Banks				
Code	f	cf	E	D cf-E	Code	f	cf	E	D cf-E
A	24	0.48	0.052	0.42	A	33	0.44	0.052	0.38
B	27	0.54	0.052	0.48	B	48	0.64	0.052	0.58
C	30	0.60	0.052	0.54	C	52	0.69	0.052	0.63
D	32	0.64	0.052	0.58	D	57	0.76	0.052	0.70
E	34	0.68	0.052	0.62	E	59	0.78	0.052	0.72
F	36	0.72	0.052	0.66	F	61	0.81	0.052	0.75
G	38	0.76	0.052	0.70	G	63	0.84	0.052	0.78
H	40	0.80	0.052	0.74	H	65	0.86	0.052	0.80
I	41	0.82	0.052	0.76	I	66	0.88	0.052	0.82
J	42	0.84	0.052	0.78	J	67	0.89	0.052	0.83
K	43	0.86	0.052	0.80	K	68	0.90	0.052	0.84
L	44	0.88	0.052	0.82	L	69	0.92	0.052	0.86
M	45	0.90	0.052	0.84	M	70	0.93	0.052	0.87
N	46	0.92	0.052	0.86	N	71	0.94	0.052	0.88
O	47	0.94	0.052	0.88	O	72	0.96	0.052	0.90
P	48	0.96	0.052	0.90	P	73	0.97	0.052	0.91
Q	49	0.98	0.052	0.92	Q	74	0.98	0.052	0.92
R	49	0.99	0.052	0.93	R	74.5	0.99	0.052	0.93
S	50	1	0.052	0.94	S	75	1	0.052	0.94
	n1=50					n2=75			
Z=0.19					Z=0.15				

(Source: Analysis Work)

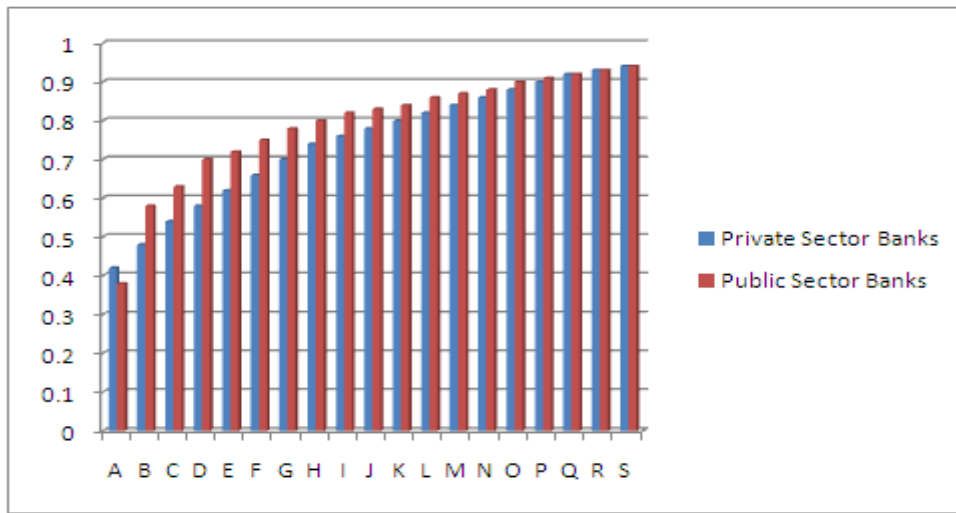


Fig 5: Percentage of Query Impact Factor

As the Query impact factor shown in the table 5 and fig 5 it is highest in private sector banks' websites than the public sector banks' website.

Test Table 6: Query Popularity

Private Sector Banks					Public Sector Banks				
Code	f	cf	E	D cf-E	Code	f	cf	E	D cf-E
A	47	0.08	0.052	0.02	A	40	0.07	0.052	0.01
B	87	0.15	0.052	0.09	B	73	0.13	0.052	0.07
C	126	0.22	0.052	0.16	C	114	0.21	0.052	0.15
D	141	0.25	0.052	0.19	D	154	0.28	0.052	0.22
E	174	0.31	0.052	0.25	E	180	0.33	0.052	0.27
F	201	0.36	0.052	0.30	F	208	0.38	0.052	0.32
G	232	0.41	0.052	0.35	G	251	0.46	0.052	0.40
H	268	0.48	0.052	0.42	H	268	0.49	0.052	0.43
I	293	0.52	0.052	0.46	I	294	0.54	0.052	0.48
J	324	0.58	0.052	0.52	J	324	0.60	0.052	0.54
K	355	0.63	0.052	0.57	K	355	0.66	0.052	0.60
L	378	0.68	0.052	0.62	L	378	0.70	0.052	0.64
M	396	0.71	0.052	0.65	M	402	0.74	0.052	0.68
N	423	0.76	0.052	0.70	N	425	0.79	0.052	0.73
O	448	0.80	0.052	0.74	O	457	0.85	0.052	0.79

P	475	0.85	0.052	0.79	P	481	0.89	0.052	0.83
Q	504	0.90	0.052	0.84	Q	499	0.92	0.052	0.86
R	533	0.96	0.052	0.92	R	512	0.95	0.052	0.89
S	555	1	0.052	0.94	S	537	1	0.052	0.94
	n1=555					n2=537			
$Z=0.057$					$Z=0.070$				

(Source: Analysis Work)

Table 6 and fig 6 gives the frequency of query popularity among the banks' websites. According to the statistics the Query Popularity of private

sector banks' websites is less than public sector banks website's for accessing information about leading banks is same.

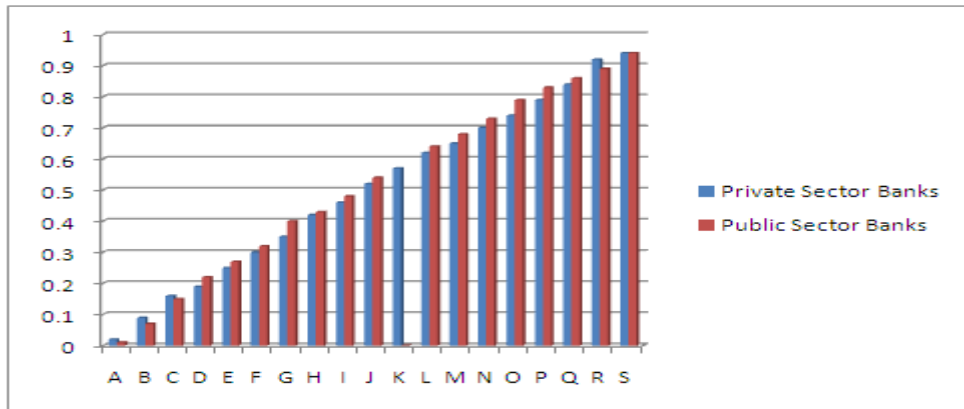


Fig 6 Query Popularity

Test Table No.7: Quality of Websites

Statistics Summary	Mean(μ)	S.D.(σ)	Statistic (U)	Z(U- μ)
Keywords Relevancy	4.5	3.240370	6.5	0.61
Page Load time(Sec)	4.5	3.240370	5.0	0.50
Trustworthiness	4.5	3.240370	6.5	0.61
Vendor reliability	4.5	3.240370	5.0	0.15
Privacy	4.5	3.240370	5.5	0.30
Unique Visitors	4.5	3.240370	6.0	0.46

(Source: Analysis Work)

The critical value of z at 0.05 level of significance is 1.96. Since the calculated values of z for relevancy, page load time, trustworthiness, vendor reliability, privacy and unique visitor are

less than 1.96. That is there is no significance difference in the quality of public sector banks websites and private sector banks website.

According to the test table 1 the calculated value of z for page load time is 0.50 which is less than 1.96 i.e. there is no much significance difference in page loading time of public sector banks and

private sector banks i.e. there is no much difference in the performance of the both sectors banks.

Test Table 8 Performance of Websites

	Private Sector Banks	Public Sector Banks
Average Time Spend in every Visit(min)	13	10
Average no. of urls	141	152

(Source: Analysis Work)

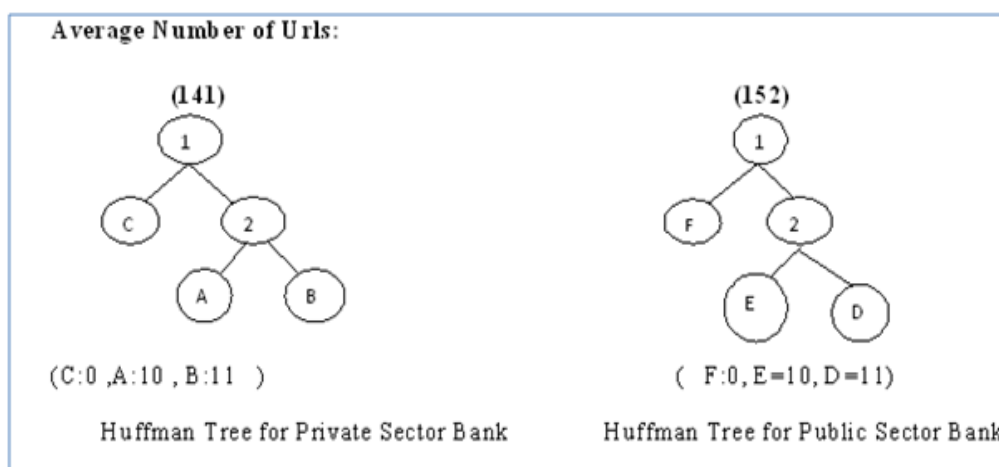


Fig 7 Average no. of urls

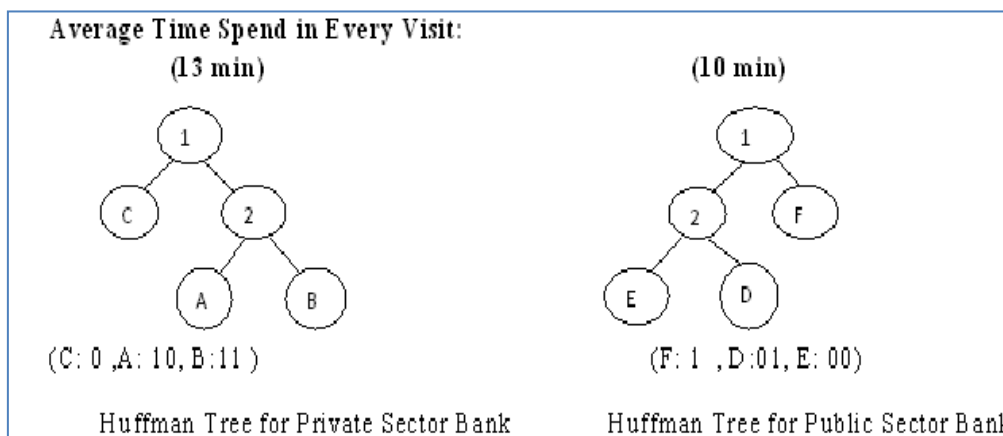


Fig 8 Average Time Spend in every Visit(min)
Average numbers of urls of the private sector bank’s websites are 141 and public sector bank’s websites are 152. The performance and effectiveness increases if the no of urls are

less. As compared to Private sector banks the average numbers of urls of public sector bank’s websites are more. Spending more time proves the effectiveness of that website.

Average time spend in every visit by the visitors of private sector banks is more than Average time spend in every visit by the visitors of public sector banks, i.e. the private sector bank's websites are quiet effective than public sector bank's websites.

6 Recommendations

1. Web pages are classified into some clusters found.
2. Keywords are classified into some clusters found.
3. The online navigation recommendation is created as a set of links to pages belonging to the current web site.
4. The user can select some links or not.
5. Make sure that when a visitor clicks your website's logo, they are taken back to the homepage.
6. Fix or remove broken links (external and internal).
7. Provide the same basic navigation menu on every page of your website.

8. Add your analytics code to every page of your website.
9. Get rid of any pages that automatically resize visitors' browser window.
10. Find what fits. Navigate around these sites to determine which best fit for your business.
11. Show a confirmation any time a visitor completes a significant action on your website.
12. Create one new piece of quality content for your website every week.
13. Default case: No recommendation.
14. It is too risky to apply the recommendations in the real web site. However it is possible to make a simulation.

7 Converting methodology into knowledge base to test recommendations

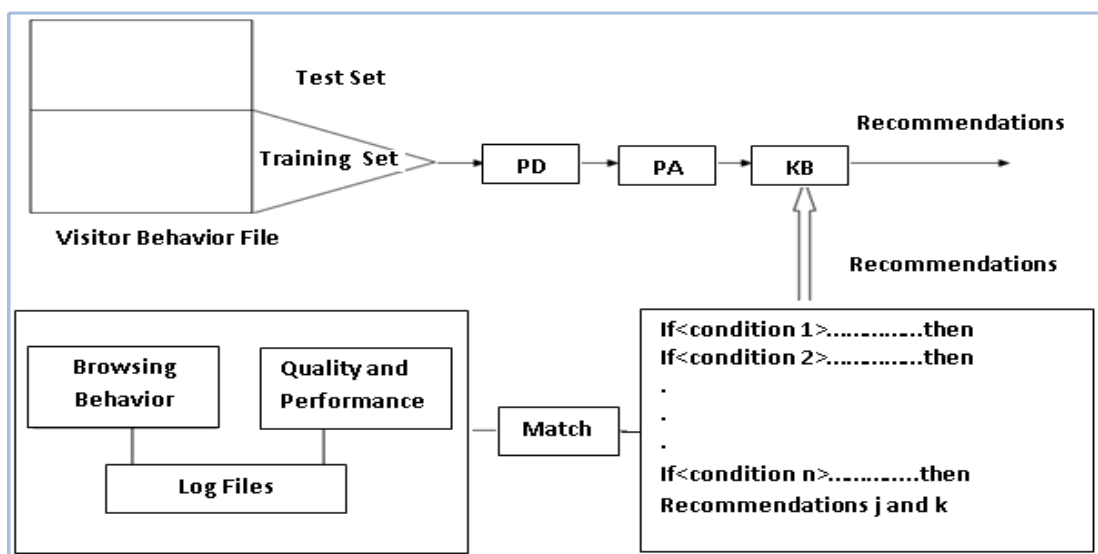


Fig 9: Proposed Knowledge Based Methodology to Test Recommendations

7 Conclusion

We presented a methodology for web usage mining and applied it for understanding user browsing behavior in a bank. Web data are a real source to analyze the user behavior in the web. An important step is cleaning and preprocessing of the web data. The application of web mining techniques allows finding unknown patterns. These patterns must be validated or rejected by an expert in the business under investigation. In this case, special emphasis was laid on promoting the use of Internet banking. The goal of increasing has been reached. This particular application used web usage mining to measure quality and performance of the websites to improve customer relationship management in financial services. As per the results Quality and Performance of websites of both sectors banks' is quite same but user behavior is different. Private Sector banks are giving good services to customers than the public sector. Depends on the results we have suggested some recommendations and a knowledge based methodology to test the suggested recommendations to improve the websites to attract the users.

8 References

[1] Mining web data: Techniques for understanding the user behavior in the Web Juan D. Velásquez Silva PhD Information Engineering, University of Tokyo, JapanK.
[2] Analysis of Server Log by Web Usage Mining for Website Improvement Navin Kumar Tyagi¹, A. K. Solanki² and Manoj Wadhwa³ IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 4, No 8, July 2010 17 ISSN

[3] Web Usage Mining: A Research Area in Web Mining Rajni Pamnani, Pramila Chawan *Department of computer technology, VJTI University, Mumbai*
[4] An Efficient Path Completion Technique for web log mining 2010 IEEE International Conference on Computational Intelligence and Computing Research
[5] R. Cooley, B. Mobasher, J. Srivastava, Data preparation for mining world wide web browsing patterns, J. Knowledge Inform. Systems 1 (1999) 5–32.
[6] F. Crespo, R. Weber, A methodology for dynamic data mining based on fuzzy clustering, Fuzzy Sets and Systems, in press.
[7] A. Famili, W.-M. Shen, R. Weber, E. Simoudis, Data preprocessing and intelligent data analysis, *Intell. Data Anal.* 1 (1) (1997) 3–23.
[8] Web Usage Mining Using Self Organized Maps Paola Britos, Damián Martinelli, Hernán Merlino, Ramón García-Martínez IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.6, June 2007
[9] J. Han, M. Kamber, *Data Mining—Concepts and Techniques*, Morgan Kaufmann Publishers, San Francisco, 2001.
[10] H. Hruschka, Market definition and segmentation using fuzzy clustering methods, *Internat. J. Res. Marketing* 3 (1986) 117–134.
[11] S.K. Pal, V. Talwar, P. Mitra, Web mining in soft computing framework: relevance, state of the art and future directions, *IEEE Trans. Neural Networks* 13 (1.5) (2002) 1163–1177.
[12] J. Peppard, Customer relationship management (CRM) in financial services, *European Management J.* 18 (3) (2000)