User Behaviour Analysis Using Ensemble Learning Algorithm In Web Mining

M. NITHYA[#1] *Dr. P. KANMANI, M.C.A., M.Phil., Ph.D.,* [#2]
M.Phil Research Scholar , *Assistant Professor,*
P.G. and Research Department of Computer Science,
Thiruvalluvar Government Arts College,
Rasipuram - 637 401.

ABSTRACT:

Information on Internet, especially on Web sites increasing rapidly day by day. Web sites play an important role where a lot of Web users are always upload, download and brows a lot of contents based on their needs. With the fast growth of the data and information in Web environment made a necessity to use sophisticated techniques that have never used in other domains to extract knowledge and significant Web patterns. The work entitled "User Behaviour Analysis Using Ensemble Learning Algorithm in Web Mining" to predict the buying intentions of the user based on his behaviour within and outsized e-commerce website. It uses traditional machine learning techniques with the most advanced Ensemble learning approaches and it analyzes the effectiveness of various machine learning classification models for predicting personalized procedure. It utilizes individual's phone log data. The classifier based on Ensemble learning is examined by conducting a range of experiments on the real datasets collected from individual users. The general investigational results and discussions can help both the researches and application developers to design and build intelligent Machine learning techniques for users.This approach uses Adaboost, Stacking and Bagging Methods of Ensemble learning algorithm. It helps to find out the frequently searched keyword of a user. By using this keyword, this Methodology can predict the user requirements. Adaboost support the user search and boost up the searching. Stacking methodology of Ensemble learning algorithm in support for personalized dataset analyses the database. Bagging process is use to filter the user behaviour level and final output. The quality and outcome of the chosen Ensemble learning algorithms came out the best with classification accuracy of 93.8%.

***Keyword*: web Mining, ensemble learning algorithm, , Stacking and Bagging Methods.**

I. INTRODUCTION

World Wide Web serves as a huge, frequently distributed, international information service inside for news, advertisements, consumer information, financial management, education, e-commerce etc. Information is arranged in proper hierarchy in the form of websites. The web also contains a rich and dynamic group of hyperlink information. Group of web pages named websites are accessed via hyperlinks. Nowadays internet plays a dynamic role for proving that information to all kinds of users to find their needs. Day - by- day the usage and availability of internet is increasing extremely. Websites are very useful for providing any kind of information to any kind of users at any time. A web server regularly registers a weblog entry for all access of webpage.

Whenever the user interacts with the web site, the communication details are routinely documented in web server in the method of web logs. The web performs two huge for active data warehousing and mining. Also the density of web pages is far larger than that of any old text documents. Only a small part of the information on the web is truly relevant. It is possible to get lots of data on user access patterns and also possible to mine interesting nuggets of information. It is one of the most important applications of data mining, artificial intelligence and so on to the web data and forecast the user's visiting behaviours and obtains their interests by investigating the samples. Web usage mining involves the analysis and discovery of user access patterns from Web servers logs in order to better serve the user's needs. In web usage mining or web log mining, user's behaviour or interests are revealed by applying data mining techniques on web log file.

The ability to know the patterns of user's habits and interests helps the operational strategies of enterprises. It is highly important for the website analyst to know and understand about the user level of interest and behaviour for variety of reasons. In web usage mining web logs plays an important role to know about user behaviour. The algorithms in web usage Mining have done the pre-processing activities for reducing the size of the log file and to identify the number of unique users and sessions. According to intelligent system web usage pre-processor categorize human and search engine accesses before applying the pre-processing techniques. Web usage mining is the one type in web mining.

This type of web mining allows for the collection of web method in information for web pages. The usage data that is gathered provides the companies with the facility to produce results more effective to their businesses and increasing of sales. Usage data can also be helpful for developing marketing skills that will out-sell the competitors and promote the company's services or product on a higher level. The pre-processing of web log data for finding frequent patterns using weighted association rule mining technique can be done to other industrial and social organizations. In recent days, the web usage mining has great possible and frequently employed for the tasks similar to web personalization, web pages perfecting and website reorganization etc. So, it is required to know the users behaviour when interaction is made with the web.

II. LITERATURE REVIEW

As like net, mobile usage is likewise developing incredibly and it creates a rich knowledge base. Even from the precise characteristics of net like link structure and its content and languages the information can be extracted. Analysis of those traits frequently exposes thrilling patterns and new expertise. The extracted expertise can be used to decorate users' efficiency and value in searching for statistics on the Web and also for using them in programs unrelated to the Web, like aid for decision making or enterprise control. The major demand in extracting information from web mining is the scale of the Web and its unstructured and volatile content material with its multilingual nature. To upload further, the Web generates a big amount of facts in other formats that incorporate worthy records. To consider as an instance, a Web server logs' statistics approximately user get entry to samples may be used for statistics personalization or improving internet content and it also facilitates to customize the net pages. In information mining, ML (Machine gaining knowledge of) techniques characterize one viable method to address the problem. In data mining, gadget studying strategies were employed in various critical programs that is associated with the web access primarily based behaviour analysis.

Tasawar et al., [1] proposed a hierarchical cluster primarily based preprocessing technique for Web Usage Mining. In Web Usage Mining (WUM), internet session clustering plays a important function to categorize web users in keeping with the consumer click history and similarity measure. Web consultation clustering in keeping with Swarm assists in numerous manner for the purpose of coping with the internet sources correctly like internet personalization, schema amendment, internet site alteration and web server performance. The author provides a framework for internet consultation clustering within the preprocessing degree of internet usage mining. The framework will envelop the data preprocessing section to exercise the net log information and trade the specific internet log facts into numerical information. A consultation vector is determined, in order that suitable similarity and swarm optimization will be used to cluster the web log data. The hierarchical cluster based totally technique will improve the traditional net consultation technique for extra based records approximately the consumer sessions. Yaxiu et al., [2] put forth internet utilization mining primarily based on fuzzy clustering.

Anand N. Describes a web usage details and affords them with the tools to recognize the online behavior in their teenage kids. Singh A.P. And Jain R.C. [14] Different kinds of net usage mining techniques with their basic fashions and ideas are provided. In addition to that, for discovering the hidden patterns from internet get admission to log files a brand new version primarily based on visual clustering is likewise counseled. The analysis of different techniques of web utilization mining.

Mishra R. And Choubey R. [15] describe the FP growth algorithm is acquiring a maximum regularly get right of entry to paths and pages from the net log records and providing precious facts to person behavior. Zubi Z. S. And Riani M.S.E. [16] discusses the usage of net mining strategies is used to categorise the net page's type in step with consumer visits. This category is allows to understand the internet consumer conduct. The category and association rule techniques for discovering the exciting facts from browsing patterns. Avneet Saluja et al. [17] in their work is person destiny request prediction using web log facts and consumer statistics. The cause of the effort is to provide a benchmark for comparing a numerous techniques used inside the beyond, a gift and which can be used in a future to decrease the hunt time of a person on the network.

Singh A.P. And Jain R.C. [6] Different forms of web utilization mining strategies with their primary fashions and concepts are supplied. In addition to that, for coming across the hidden styles from web get entry to log files a new version based totally on visible clustering is likewise suggested. The evaluation of different strategies of web usage mining. Mishra R. And Choubey R. [7] describe the FP-increase set of rules is acquiring a most frequently get entry to paths and pages from the internet log statistics and providing valuable statistics to person behavior. Zubi Z. S. And Riani M.S.E. [8] discusses using net mining strategies is used to classify the internet page's type in keeping with consumer visits. This classification is helps to understand the web user conduct. The type and affiliation rule strategies for coming across the thrilling records from browsing patterns. Avneet Saluja et al. [9] of their paintings is consumer destiny request prediction using net log facts and consumer records. The purpose of the effort is to provide a benchmark for evaluating a diverse methods used in the past, a present and which can be used in a destiny to decrease the quest time of a person at the network

.

III. METHODOLOGY

Click circulation information can be used to quantify search behaviour using machine studying techniques, in general targeted on purchase records. While buying suggests consumer's last preferences in the same category, search is additionally an imperative element to measure intentionality in the direction of a particular category. We will use a probabilistic generative manner to mannequin person exploratory and purchase history, in which the latent context variable is added to capture the simultaneous influence from each time and location. By identifying the search patterns of the consumers, we can predict their click decisions in unique contexts and advise the right products.

Modern search engines use laptop mastering tactics to predict user recreation inside web content. Popular fashions include logistic regression (LR) and boosted selection trees. Neural Networks have the gain more LR because they are in a location to capture non-linear connection between the input facets and the deeper" architecture has essentially enhanced modeling strength. On the different hand decision timbers - albeit popular in this domain - face additional challenges with high-dimensional and sparse records.

A. Proposed Web Mining Based Ensemble Learning Algorithm

Each coin has two faces every face have its personal property and features. It's time to uncover the faces of ML. A very great tool that holds the possible to transform the way effects work.

Ensemble methods are techniques that generate multiple models and then combine them to produce improved results. Ensemble methods frequently produce more exact solutions than a single model would. This has been the container in a amount of machine learning competitions, where the winning solutions used ensemble methods.

Ensemble Learning and process web usage data, extract interesting behaviour patterns from the formulate data, express interactive visualizations to better analyze the extracted patterns and allow individuals to compare themselves over time.

To begin with, qualitative and quantitative web usage data features are identified such as dwell time, number of hits, category, idle time, and time of occurrence. A web browser add-on logs these data features on the trigger of different browser events such as creating of the tab/window, updating the tab/window, closing tab/window, status of window changes etc.

B. Advantages Of Proposed Method

Easily identifies trends and patterns

Machine Learning can evaluate large volumes of data and determine specific trends and patterns that would not be obvious to humans.

No human intervention needed (automation)

With ML, you don’t need to look after your project every step of the way. Since it means giving machines the ability to learn, it lets them create predictions and also improve the algorithms on their own.

Continuous Improvement

As ML algorithms achieve experience, they maintain improving in accuracy and efficiency.

Handling multi-dimensional and multi-variety data

Machine Learning algorithms are good at managing data that are multi-dimensional and multi-variety, and they can do this in dynamic or doubtful environments.

Wide Applications

You could be an e-tailer or a healthcare supplier and make ML work for you.

C. Ensemble Algorithm

```
Input: Data set D={(x1, y1),(x2, y2),···,(xm, ym)};
Base learning algorithm L;
Number of learning rounds T.
Process: D1(i) = 1/m.   // Initialize the weight distribution $D_t$
fort=1,···,T:
    ht = L(D, $D_t$); // Train a base learner $h_t$ from D using distribution $D_t$
$\alpha_t = ½ \ln 1- \text{€}_t / \text{€}_t$   // Determine the weight of $h_t$
Dt+1(i)  =Dt(i) / Zt X $\exp(-\alpha_t)$ if $h_t(x_i) =y_i$
$\exp(\alpha_t)$ if $h_t(x_i)6=y_i$
$D_t(i)\exp(-\alpha_t y_i h_t(x_i))/ z_t$ // Update the distribution, where $Z_t$ is a normalization
// factor which enables $D_{t+1}$ to be a distribution
end.
Output: $H(x) = \text{sign}(f(x)) = \text{sign} \sum^T{}_{t=1} \alpha_t h_t(x)$
```

D. Ensemble Algorithm Learning Processes

Step1:

Collecting the data

This stage involves the collection of all relevant data from various sources.

These are collecting the data’s in various data base location and database servers.

Frame work architecture to support the data collecting processes. The frame work process is intergraded technology, this technology to integrate the various database servers. User easy to search the various keywords. Quickly analysis the user behaviours. Server analysis various dataset from various places. This are support the ensemble leaning algorithm.

Three type of learning method has been applied ensemble learning

Supervised Learning: It is the one where you contain input variables (x) and an output variable (Y) and you utilize an algorithm to learn the mapping function from the input to the output.

Unsupervised Learning: Sometimes the given data is unstructured and unlabeled. So it becomes difficult to organize that data in different categories. Unsupervised learning helps to solve this problem. This learning is used to cluster the input data in lessons on the beginning of their statistical properties.

Reinforcement Learning: It is all about taking appropriate action in order to maximize the reward in a particular situation.

Input X,

User define X value, these value may be numbers, keyword etc.

Step2:

Then assign the formula:

Y=f(x)

Here, Y is the output result; f(x) is used to define the data. These data based on user behaviour.

Data wrangling

It is the process of cleaning and converting “Raw Data” into a format that allows appropriate consumption

- Discover
- Structure
- Clean and enrich

Step3:

Analyze Data

Data is analyzed to select and filter the data required to prepare the model.

Step4:

Train algorithm

The algorithm is trained on the training dataset, through which the algorithm understands the pattern and the regulations which manage the data

We want to predict the probability. We predict the user analysis: supervised and unsupervised on different parameters. We need to design an algorithm to predict the possibility. Let us first initialize a few demotions.

X is individual predictors.

f(x) is the weight given to the x the predictor

Initial f(x) for x in [1, 11] are all 1.

We created a reproduction with 1000 predictions done by each of the 11 predictors.

Step5:

Test algorithm

The testing dataset determine the accuracy of our model.

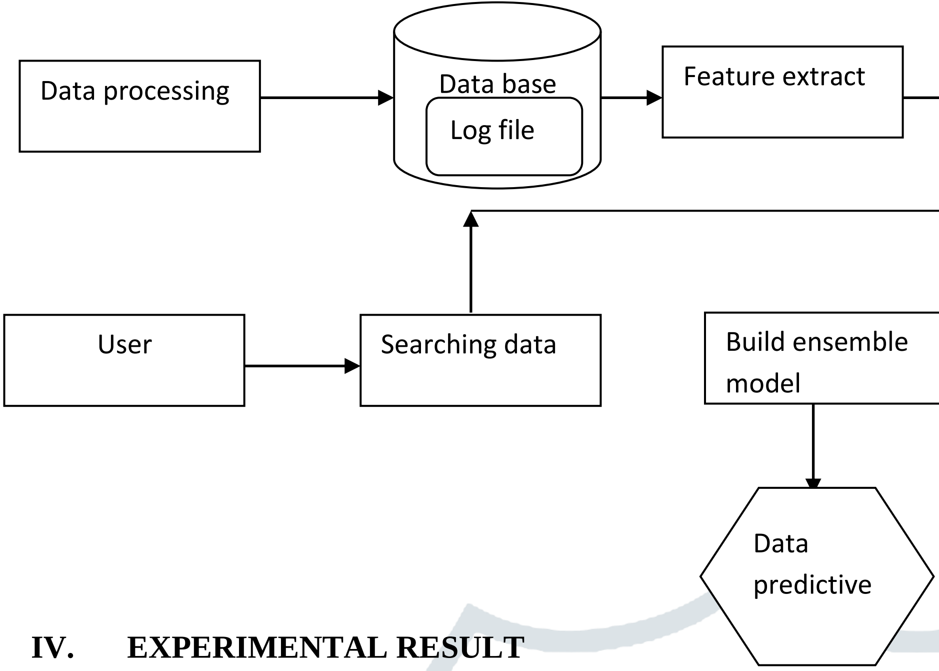
Sum (f(x) for all supervised) > Sum (f(x) for all unsupervised prediction)

Step6:

Deployment

If the speed and accuracy of the model are acceptable, then that model should be deployed in the real system. After the model is deployed based upon its presentation the model is updated and enhanced if there is a difference in performance.

E. *System Flow Diagram*



IV. EXPERIMENTAL RESULT

The experimental evaluation is carried out in ML .NET Frame Work. The parameters considered are Ensemble. Analysis for the user behaviours has been taken for experimental analysis.

A. *Performance Evaluation*

This chapter discuss about experimental evaluations and discussions of the proposed approaches. The performances of the proposed method haven compared with the standard existing approaches such as deep learning method. It is observed from the previous chapters that the presented approaches outperformed the existing approach.

	Adaboosting	Bagging	Stacking	Ensemble
Dataset-2	86.1	89.4	89.1	93.2

The blending of resolution boundaries achieved by the stacking classifier. Summarize also shows that stacking achieves upper accuracy than individual classifiers and based on learning curves, it shows no signs of over fitting.

Stacking is a commonly used technique for winning the Kaggle data science struggle. For example, the first position for the Otto Group creation group challenge was won by a stacking ensemble of over 30 models whose output was used as characteristics for three meta-classifiers:

Adaboost, Bagging, and Stacking.

Table 2: Result Analysis for user dataset -2

Procedure To Obtain Comparison Using Ensemble Algorithm

Step1: This stage involves the collection of all relevant data from various sources.

Three type of learning method

Supervised Learning: It is the one where you contain input variables (x) and an output variable (Y) and you utilize an algorithm to learn the mapping function from the input to the output.

Unsupervised Learning: Sometimes the given data is unstructured and unlabeled. Input X,

User define X value, these value may be numbers, keyword etc.

Step2: Then assign the formula: Y=f(x)

Here, Y is the output result; f(x) is used to define the data. These data based on user behaviour. It is the process of cleaning and converting "Raw Data" into a format that allows convenient utilization.

Step3: Data is analyzed to select and filter the data required to prepare the model.

Step4: The algorithm is trained on the training dataset, through which the algorithm understands the pattern and the rules which govern the data. X is individual predictors.

f(x) is the weight given to the x the predictor

Initial f(x) for x in [1,11] are all 1.

We created a recreation with 1000 predictions done by each of the 11 predictors.

Step5: The testing dataset determines the accuracy of our model.

Sum (f(x) for all supervised) > Sum (f(x) for all unsupervised prediction)

Step6: If the speed and accuracy of the model are acceptable, then that model should be deployed in the real system.

	Adaboosting	Bagging	Stacking	Ensemble
Dataset-1	85.1	88.2	89	92.2

Table 1: Result Analysis for user dataset -1

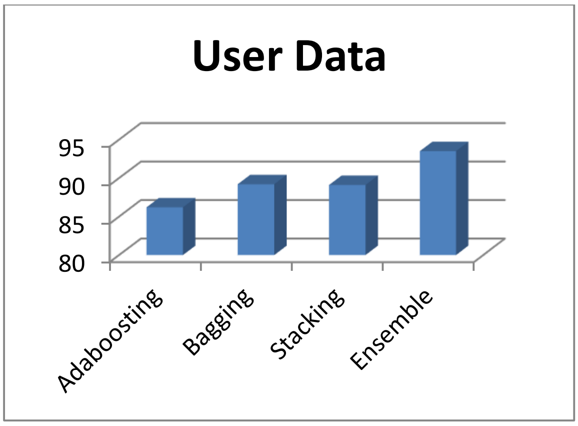


Figure 1: User Dataset- 1

Here, first level of performace ratio, adaboost is better than baggging, seceond level of performance ratio, bagging is better than stacking. Third level of performance ratio, stacking is better than ensemble. dataset –1 comparison with adaboost, bagging and stacking algorithm better the ratio performance increase the ensemble leaning 92.2%. The results are illustrated in table 5.1

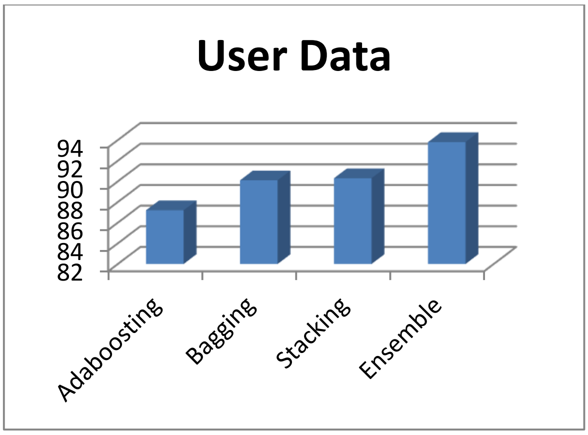


Figure 2: User Dataset- 2

Here, first level of performace ratio, adaboost is better than baggging, seceond level of performance ratio, bagging is better than stacking. Third level of performance ratio, stacking is better than ensemble.dataset –2 comparison with adaboost, bagging and stacking algorithm better the ratio performance increase the ensemble leaning 93.2%. The results are illustrated in table 5.2

	Adaboosting	Bagging	Stacking	Ensemble
Datatset-3	87.2	90.1	90.3	93.8

Table 3:

Result Analysis for user dataset -3

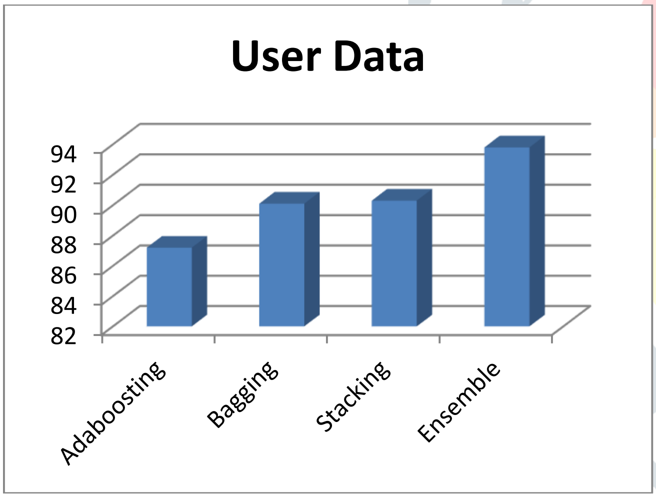


Figure 3: User Dataset- 3

Here, first level of performace ratio, adaboost is better than baggging, seceond level of performance ratio, bagging is better than stacking. Third level of performance ratio, stacking is better than ensemble.dataset –3 comparison with adaboost, bagging and stacking algorithm better the ratio performance increase the ensemble leaning 93.8%. The results are illustrated in table 5.3

B. Performance Output

ML.NET Framework Ensemble algorithm increase the performance of User behaviour Analysis comparing various graphs.

The Figure of 5.1 to 5.3 is demonstrate the user behaviour analysis Ensemble learning algorithm for better performance of output.

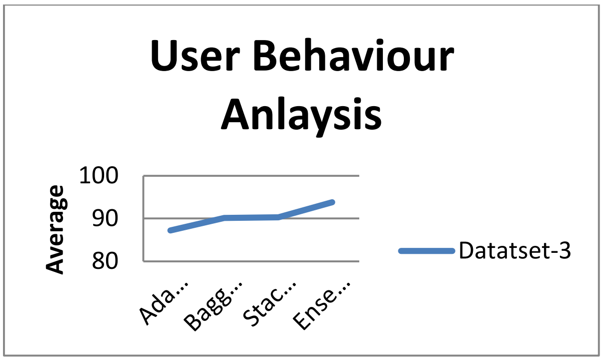


Figure 2: User Behaviour Analysis

The qualified analysis shows that multilayer perception performed most excellent with classification accuracy of 86.5% of adaboosting , 90.1% of bagging , 92.1% of stacking, 93.2% of ensemble algorithm.

V. CONCLUSION

It is common to use ensembles in deep learning by training various and accurate classifiers. Diversity can be achieved by altering architectures, hyper-parameter settings, and training techniques. Ensemble methods have been very successful in setting record presentation on challenging datasets. Web Usage Mining procedure forever depend upon the trouble of attractive application. Data Mining is the development to mine the interesting knowledge from the huge amount of data.

Various methods of data mining technique and describe the apply data mining techniques for user behaviour analysis for machine leaning. To compare the existing algorithm adaboosting, bagging and stacking are better than ensemble learning algorithm have outstanding performance. The work is to develop a ensemble learning to predict user behaviour based on web server log files.

Future Work

- Additional research can include testing on real-time data, and see the presentation effects on a real time.
- More work would require to be done on improving time effectiveness in terms of scalability.
- It is helpful to initialize the neural network and train it with unsupervised contrastive divergence on a huge volume of dataset.
- Better results can be achieved by train several networks in parallel and consider all of their outputs separately instead of considering single result.

REFERENCES

[1] S. Jagan, and S.P. Rajagopalan, “A survey on web personalization of web usage mining”, IRJET- International Research Journal of Engineering and Technology, 2015.

[2] A. Ladekar, P. Pawar, D. Raikar and J. Chaudhari, “Web Log Based Analysis of User’s Browsing Behavior”, IJCSIT - International Journal of Computer Science and Information Technologies, Vol. 6 (2), 2015.

[3] S. Parvatikar and B. Joshi, “Analysis of User Behavior through Web Usage Mining”, ICAST - International Conference on Advances in Science and Technology, 2014.

[4] A. Deepa, and P. Raajan, “An efficient preprocessing methodology of log file for Web

usage mining”, NCRIIAMI - National Conference on Research Issues in Image Analysis and Mining Intelligence, 2015.

[5] N. Anand, “Effective prediction of kid’s behavior based on internet use”, International Journal of Information and Computation Technology, 2014.

[6] A.P. Singh, R. C. Jain, “A Survey on Different Phases of Web Usage Mining for Anomaly User Behavior Investigation”, IJETTCS - International Journal of Emerging Trends & Technology in Computer Science, Vol 3, 2014.

[7] R. Mishra, A. Choubey, “Discovery of Frequent Patterns from Web Log Data by using FP-Growth algorithm for Web Usage Mining”, International Journal of Advanced Research in Computer Science and Software Engineering, Vol 2, 2012.

[8] Z.S. Zubi, M.S. Riani, “Applying web mining application for userbehavior understanding”, Recent Advances in Image, Audio andSignal Processing.

[9] A. Saluja, B. Gour, and L. Singh., “Web Usage Mining Approachesfor User’s Request Prediction: A Survey”, IJCSIT-InternationalJournal of Computer Science and Information Technologies, Vol. 6(3), 2015.

[10] L. Tamrakar, M. Ghosh., “Identification of Frequen NavigationPattern Using Web Usage Mining”, IJARCST-International Journalof Advanced Research in Computer Science & Technolog, Vol. 2,2014.

[11] S.P. Ajeetkumar, P.K Anagha, “Review on Exploring User’s SurfingBehavior for Recommended Based System”, IJETTCS -International Journal of Emerging Trends & Technology inComputer Science, Vol 3, 2014.

[12] A. D. Kasliwal, and G. S. Katkar, “Web Usage mining for PredictingUser Access Behavior”, IJCSIT-International Journal of ComputerScience and Information Technologies, Vol. 6 (1), 2015.

[13] V. Neha, Patil et al, “Prediction of Web Users BrowsingBehavior: A Review”, International Journal of Computer Science andMobile Computing, Vol.4,209-212, 2015.

[14] S. Khan, Y. Singh and A. Kumar. Sachan, “Web Mining Approach inAnalysing User Behavior and Interest for Website Modification”,International Journal of Advanced Research in Computer Scienceand Software Engineering, Vol 5, 2015.

[15] A. Vishwakarma, and K.N. Singh, “A survey on web log miningpattern discovery”, IJCSIT - International Journal of ComputerScience and Information Technologies, pp: 7022-7031, 2014.

[16] S. Parvatikar, B. Joshi, “Analysis of user behavior through webusage mining”, ICAST - International Conference on Advances inScience and Technology,pp: 27-31, 2014.

[17] Y.C. Liu, Y.C. Hsu, “Predicting Adolescent Deviant Behaviorsthrough Data Mining Technologies”, Educational Technology &Society, 2013.

[18] J. Vellingiri, S. C. Pandian., “A Survey on Web Usage Mining”,Global Journal of Computer Science and Technology, Volume 11Issue 4, 2011.