International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 04 | Apr 2019 www.irjet.net p-ISSN: 2395-0072

WEB TRAFFIC ANALYSIS THROUGH DATA ANALYSIS AND MACHINE LEARNING

K. Muragesh[1], B. Masthan baba[2]

[1]Reg.no:1116126, MCA VI semester, Department of MCA, Sree Vidyanikethan Institute of Management, S V University.
[2]Assistant Professor, Department of MCA, Sree Vidyanikethan Institute of Management, A.Rangampeta, Tirupati.

Abstract:- The web is a collection of different networks with relative information. The information will process throughout the globe with the help of web. Now days the size of netizens are increased and most of the servers are not responding to the user's request. So it became necessary to analyze the traffic on the web and suggest the improvement of the web services. Data Analysis is one of the best tools used for analysis purpose. By implementing the Data analysis on various servers there is a possibility to analyze the usage, traffic and response time of the server. The rendezvous point may be, identified and the variations of the traffic may be monitored and analyzed. The analysis report presents a set of suggestions to the website owners or servers to increase or decrease the server capacity. So the web access synchronizes with the traffic. The analysis of web helps to estimate the traffic according to the changes of access in the website. The ratio of visitors and views are estimated according to the traffic.

Key Words: rendezvous point, traffic analysis, data metrics, sessions, views

Introduction:

The need for the web traffic analysis is increasing day to day. Because the usage of the net and searching for data, communication and data transfer has been increased day to day. So most of the servers are getting busy and the services are providing in proper manner to the users [1]. The users are scaring to use web based services due to an improper access. Online financial transactions are creating very complexity. The web traffic analysis can give a solution for all these types of problems. The present paper analyzes the web traffic from time to time and provides the analysis report. Based on the analysis report one can take necessary actions to reduce the web traffic and provide maximum access to the users [2].

Basic concepts of the web Analytics:

Data collection: The primary approach for web analytics is data collection. The raw data is available in various formats. The data may be varied and purely depends on the usage of the net. The data usage maintains a count of things to analyze the web traffic [3]. The web usage generates huge volumes of data. So it is necessary to gather and procure the huge volumes of data.

Data Process: The data collection provides a raw data, to implement the data analysis we need to process the data. The processed data may be implemented into the machine learning for the web analysis. The required data metrics we can decide in this stage [4].

The web Traffic counting:

Web traffic is depends on the basis of the web usage. The web usage is depends on various factors. The factors are

- Number of users
- Number of New Users
- Number of sessions
- Number of sessions per user
- Page views
- Session Allocations

Various factors of the web analysis as shown in figure 1.0

Figure 1.0: The factors of the web analysis

Identifying the important performance indicators:

The process of identifying the key indicators: The count values are converted into ratios. The predefined business functionalities are implemented on the count data values [5]. The conversion aspects are referred as key performance Indicators also referred as KPI.

Implementation of online Strategies:

The required goals, standards, and objectives are defined in this section. Most of the required goals we can set. The related strategies are developed.

The web Analysis Methods:

The web analysis methods are broadly categorized into 2 types. It is purely based on accessing the web.

- Online Traffic
- Offline Traffic

The online traffic is the most required traffic analysis method. It is essential to analyze the traffic before hosting any the web page or the web site [6]. The organization maintains a log file to record the online data traffic. The required data may be extracted from the log file to analyze the traffic.

The offline traffic refers the page tagging, the page tagging records the data values when the data was downloaded from the web. The above online and offline log files data will be extracted to analyze the traffic.

Methods for The web traffic analysis:

Tick analytics:

This is one of the most analytics methods for the analysis of the web traffic, most of the users are using tick with the help of the mouse. The users are using the tick analytics to perform the analysis of the web traffic. The web traffic is purely impressive of the new users in this society [7]. The tick analysis plays a major role in the real world to perform the web of their choice. The performance of the web users purely depends on the web traffic, and the performance measure is one of the relative factors to measure the web traffic.

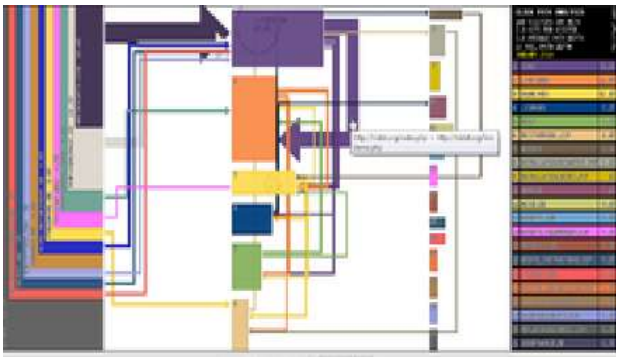

Figure 2: Representing the traffic on the web

The graph that represents the traffic on the web is defined in the figure 3.0

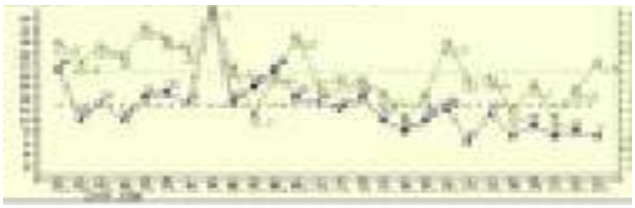

Figure 3: Traffic on the web

Identifying the location of users:

To analyze the web traffic location plays a major role. Based on the location we have to decide the server capacity. This will give the Internet Protocol usage with the help of address defines by the location of the user. The connection type and the ISP (Internet Service Provider) provide the usage of net according to the location. The location is defined as target. The users may also be divided into segments. The segmentation allows companies to target the promotion [8]. The location id helps to reduce the crime and identifying the fraud, supports local search and also help to distribute the contents.

Customer Lock in:

Customers try to go with their routine taste. But the web site is not properly access then they will try to change their options. So we need to provide better service. Every individual user contains a separate data point.

Implementing Machine Learning for The web analysis:

Machine learning is one of the best tolls of the web analysis. The Machine Learning accepts the bulk data with different formats. The data can be formatted and converted into machine readable form. Various parameters of the web traffic will be accepted as input and process the data [9]. The data validations and cleaning will be taken part. The data cleaning always refers to eliminate the unnecessary data from the dataset. The dataset converts into a clean dataset and implement on the web traffic [10]. The web traffic was analyzed on each the website and other locations. The traffic analysis defined as shown in the table 1.0.

Table1.0 Web Traffic analysis with source and users

S.No	Source	No of Users	New Users	The web Analysis
1	The web Page	Limited	Limited	Good Performance

2	The web Site	Limited	Medium	User Required Performance
3	Region wide	Medium	High	Average
4	Segment wide	High	High	Average
5	Locality wide	High	Very High	less

CONCLUSION

Internet usage was increased and people are searching the data on the net. The search process and the data downloading creating a lot of problem for web traffic, so it is essential to monitor the web traffic and providing required capacity for the netizens became very important. In our study we have concentrated more to analyze the web traffic. The machine learning allows inputting huge records and monitoring the web traffic based on the requirement.

REFERENCES

[1] T. Bemers-Lee, R. Cailliau, A. Luotonen, H. Nielsen, and A. Secret, —Tbe World-Wide Web,‖ Colnm!rt. CM vol. 37, no, 8. pp, 76-82, Aug. 1993.

[2] V. Paxson, —Growth trends in wide area TCP conneclions,‖ IEEE : Network, vol. S. pp, 8-17, July/Aug. 1994.

[3] Navin kr Tyagi ,A.K. Solanki, Manoj Wadhwa : Analysis of Server Log by Web Usage Mining for Website Improvement, IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 4, No 8,pp. 17-21,2010.

[4] L.K. Joshila Grace, V.Maheswari, Dhinaharan Nagamalai: Analysis of Web Logs and Web User In Web Mining, International Journal of Network Security & Its Applications (IJNSA), Vol.3, No.1, January 2011.

[5] Theint Aye: Web Log Cleaning for Mining of Web Usage Patterns, IEEE, 2011.

[6] Brad Reese. (2007 November 11) Cisco invention NetFlow appears missing in action as Cisco invests into the network behavior analysis business [Online]. Available: http://www.networkworld.com/community/node/22284.

[7] Cisco. (2007, October) Cisco IOS Netflow, Introduction to Cisco IOS Netflow – A Technical Overview [Online]. Available: http://www.cisco.com/en/US/prod/collateral/iosswrel/ps6537/ps6555/ps6601/prod_white_paper0900aecd80406232.html

[8]. Martin Arjovsky, Amar Shah, and Yoshua Bengio. Unitary evolution recurrent neural networks. In International Conference on Machine Learning, pp. 1120–1128, 2016.

[9] Y. Bengio, P. Simard, and P. Frasconi. Learning long-term dependencies with gradient descent is difficult. Trans. Neur. Netw., 5(2):157–166, March 1994. ISSN 1045-9227. doi: 10.1109/72. 279181. URL http://dx.doi.org/10.1109/72.279181.

[10]. Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555, 2014.