A Comprehensive Review of Web Analytics

[1]Deekshith, [2]Prof. Suma B

[1]Student, [2]Assistant Professor,
[1]Computer Science and Engineering,
[1]RV College of Engineering, Bangalore, India.

Abstract : Understanding user engagement in websites and understanding the performance of websites plays a very important role in any application. Web analytics answers questions like what is the user's behavior on the website, how many users visit the website, what is the session length of the users etc. Web analytics not only answers these questions but also provides a foundation for understanding user patterns and thus improving website performance and improved features to users. This paper provides an overview of how web analytics can be used for understanding user behaviors, metrics that can be used to understand user patterns, and the application of web analytics for business growth.

IndexTerms **- Web analytics, HTTP, log file, page tagging, JavaScript.**

I. INTRODUCTION

Web analytics started with simple HTTP function traffic logging and currently, it's more than a comprehensive suite of usage data tracking, analysis, and reporting. It is mainly used for measuring insights of many aspects of direct interactions with application like number of visits, time spent on the site, and click path, etc. Understanding visitor's behavior is one of the goals of web analytics. To use web analytics, understanding what kind of information can be collected from a website is very important and also how that information will be useful for understanding business goals is also important. The emphasis is on creating measures based on organizational and user goals, and then using web analytics data to assess if the goals were met or not, as well as to drive strategy and improve the user experience.

Fundamental and useful information which can be recorded are the pages which are viewed by a user when he or she visits the website. Also from collected data, a web analytic tool gives insights into how users are moving in the entire application. Also, web analytics captures how a user comes to the application and how long they retain in pages. The right way of data collection and the right tool for web analytics helps greatly in analyzing user engagement and behavior patterns.

II. BACKGROUND

One of the main ways for web analytics after the World Wide Web was log files. The first experiment of website analytics using data which is collected from web server logs was done by Web Trends, a Portland a Oregon based company and it was one of the pioneers of weblog analysis. There is a limit exists for the type of data collected in the web server logs. Though weblogs were useful, weblogs do not account for important information such as screen size, interaction on-page by the user, user's clickstream events, etc. Page tagging can overcome the limitations of mentioned problems. Web analytics are majorly used for

1. Improving user experience: For an application user interface is very important and it's the direct measure of how a user feels when he/she is interacting with the application. It mainly provides a positive experience to the users which helps the user to trust the brand or product and also retains users in the application. Additionally, most of the business success is mainly from a meaningful user experience of the application. User experience is different for everyone. By identifying user interest/attention areas, web analytics provides insights about how improved features could be provided to users which could result in a very good user experience.

2. Improving application/website Design: Understanding where users are struggling in the application could be one of the important metrics which gives insights about pain points in the application. Web analytics can be used to get insights about such bad experiences in an application and helps in further improvement which can be done. For example, Website's active areas like users clicks on the entire website can be analyzed through heatmap. This technique can be used by developers or product managers to enhance existing features or add new features to the application.

3. Understanding User behavior about new commercial campaigns or new program initiatives in applications. Web analytics should be able to answer questions like "where did the visitors are coming to the application". Web analytics mainly helps in understanding various types of traffic sources to the application, various marketing channels like email, public announcements, etc, various visitor types. These traffic sources are directly correlated with the marketing campaign of the application and thus it helps in understanding the effectiveness of the marketing campaign and its return on investments.

4. Identifying issues present in the application and improving the performance of the application. Performance issues of application/website can also be addressed by web analytics. Average page loading time based on geographical location can be identified and improved through web analytics. The real-time and historical analysis could be used for proactive detection, investigation, and diagnosis of performance issues. To understand and measure performance one can use page loading metrics such as average load time of the browser and geographic location.

5. Understanding customer retention: Customer retention mainly is the process of engaging existing customers of the application to continue buying products or services from your business. Customer retention is very important part of growing and sustainable business. A study from Harvard Business School says that increasing customer retention rates by

only 5% increases a company's profits by 25-95%. Customer retention is very important for any customer-driven business model. Metrics such as the number of active daily users, weekly active users could be analyzed and also analysis of the inactive users can be done using web analytics tools. This will help in making retention strategies for inactive users of the applications.

III. METRICS

Understanding web analytics metrics is necessary for the benefits of website analysis. Analysis of web traffic and improvement upon website traffic can be done by analyzing the metrics. Matrices can be broadly divided into one of four categories: site usage, site content analysis, performance analysis, referrers details (or how visitors arrived at your site), and quality assurance. Below is the example for metrics:

3.1 Visitor type

In the initial stage of web analytics, web page view count was mainly used to understand website traffic. In later stage analytics tools started using the number of unique users instead of web page view count. Two types of users are present for any website/application, those who used the website for the first time and those who have used it already. The system should be able to identify each user who uses the website. One unique user is an individual user for a website. Since all modern websites support login using multiple devices, a web analytics system should also be able to identify such users. Most of the modern analytics tools have features that address the mentioned scenario. Analytics tool uses cookies for storing unique user details, so unless the user deletes or disables cookies it is considered as a unique user. If a user disables or clears cookies then the user is considered as a new user. Also instead of cookies few analytical tools use sessions instead of cookies and each session is considered as unique users and these session data are not based on cookies. Also, the system should be able to differentiate multiple users from the same device(share computer).

3.2 Location and system details

Location/demographic metrics physical location of the device which is used to make page requests. These kinds of metrics are useful for services that provide region-specific content. For example, suppose an application is only specific to one region and wants to expand in a new region, analysis of how many requests are coming from a new region can be done using web analytics, this will provide insights about how many users have already tried to access the application. Web analytics helps in understanding information about the system users are using. For example, what types of operating systems, what browsers are mostly used by the users, what are the common screen resolutions users are preferring etc... These kinds of web analytics metrics help in developing a well-designed application which is compatible with all types of the operating system, browsers, etc.

3.3 Referrers

Page visited by the user immediately before coming to the website is called a referral page. In simple terms, it is the webpage that directs traffic to the current website. Example If a user searches in google.com and he/she was redirected to an application then google.com is a referral. These metrics can be effectively used for understanding the effect of marketing campaigns for the application and also this metric shows search engine popularity. As advertising is a very crucial part of any business, this metric plays a very important role in analyzing the effectiveness of the advertising campaign. Also, the unexpected referral can be investigated using the metric. Reports which are generated based on referral helps to get insights about public relation reach of marketing and also helps in identifying spam referrer or referrer.

3.4 Errors

Errors are the main reason which leads users to exit an application and thus it leads to a decrease in the retention rate of users in an application. Error tracking plays an important role in any software development. One of the advantages of using web analytics is error tracking. In an application error tracking is one of the difficult tasks, and the error users are facing is mainly obtained through communication from users through support channels. Web analytics can be used to track all failures/errors in an application, which helps developers to identify and resolve them before any negative impact on business. This metric provides what errors users are facing in applications and few metadata about the errors and also it provides insights about frequently occurring errors and also these error information can be segmented using various regions, system specifications, etc.

3.5 Conversion Rates:

Conversion rate is mainly the percentage of users who have done the desired action on the website. For example, if 100 users have visited an e-commerce website and 60 users have purchased a product then the conversion rate of purchasing a product is 60%. Conversion rates are often used in e-commerce websites.

3.6 Average visit Duration of users

The average amount of time a user spends on the website is called the average visit time. This metric is useful for performing group analytics like finding which group of users are having more average visit duration, which group of users having less average visit duration. And also the behavior of a group of users who have less average time can be analyzed using web analytics.

3.7 Bounce Rate

The bounce rate of a page or a group of pages is the number of users who entered a website and left the website without visiting any actions like clicks, navigation etc.. in the website, divided by the total number of users who entered the website on that page. It considers user counts who just come to the website and doesn't perform any events. Bounce rate mainly depends on the type of application we are performing in web analytics. The bounce rate will be high for an application that is designed such that it shows the information to the user and then navigates to a different page without any user action. An application with many user-friendly features expected to have less bounce rate.

IV. WHO USES WEB ANALYTICS

Several stakeholders will get benefits using web analytics, few groups of stakeholders are administrators, web designers, product managers, marketing team, etc. The administrator mainly uses web Analytics to monitor the availability of the application and also to understand web traffic to the application and the growth and performance of the application in the market. Product managers mainly use web analytics for analyzing insights into user engagement and behavioural patterns. Analyzing users behaviour patterns is very important as it helps in building applications with more features and also eliminating few poorly designed features. For example, product managers can assess the user experience of the application and also site usability by different groups of users analysing where users spend most of the time on the application and where users are facing difficulty in using the application and leaving the application. Developers mainly use web analytics for error tracking purposes. Most of the applications are prone to error when the number of users increases or web traffic increases. Error tracking is very important for building a resilience system, web analytics could help in tracking errors faced by users or any unexpected or failure in the application. Also, system properties which are collected from web analytics can be used to understand system information of the user like what operating system is used more, what's the most type of screen resolution used by the user, mostly used browsers etc... This helps applications to be designed in such a way that it is compatible for all types of devices. Also, web analytics helps in understanding what features are frequently used by users and what features are not, this helps in the enhancement of the feature in the application. Different advertising campaigns can be effectively analyzed and tracking of the same can be done using web analytics. Web analytics truly helps in understanding the effectiveness of the marketing campaign.

V. DATA COLLECTION

There are many ways for collecting behavioural data for web analytics. Few of the important methods are web logging, JavaScript tagging, web beacons and packet sniffing.

5.1 Web logging:

Server-side data collection involves storing all the transactions on the server in a text file called log files. When a flag is turned on the server, the server logs all information about requests which are coming from the client, along with metadata in a file called "logfile". Each line in log files contains information like IP address, request timestamps, items requested, the user agent of the request referrer of the request, etc... and each line is called a "hit". Many hits to the server mainly correspond to a single page and normally one hit doesn't account for a single page view count. In the late 1990s, because of increasing search engine spiders on the web, web proxies, advanced techniques in Internet services (example-Dynamic IP) there was a need for a new approach. Web hosting companies use these log files mainly to manage storage and bandwidth issues. Also, advanced software can be used to parse these log files and analyze them to get insightful information that would help in business growth.

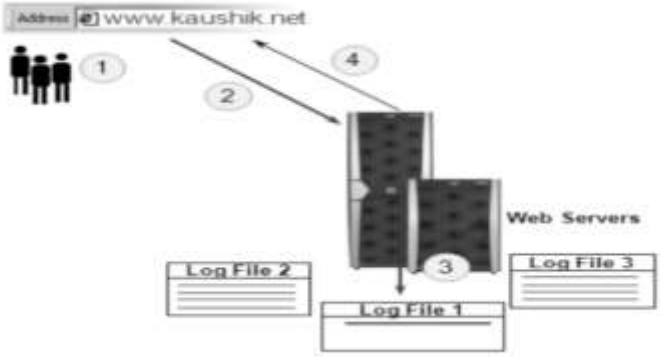

Figure 1: Log File Data Collection Visualization. Source: [2]

5.2 Page Tagging

Weblog analysis was not an efficient solution as it had many drawbacks like inaccurate results etc... These were some of the early examples of web traffic analysis on the client side. This technology eventually evolved into what it is now: script-based data collecting, which issues a cookie to each user, analyses their website behaviour, and then processes the data remotely. In many circumstances, delegating the task to a remote service provider facilitates configuration and reduces overhead. Both of these tracking approaches have benefits and drawbacks, and when used separately, they may not provide a complete picture of a website's performance. JavaScript tagging is done by adding small code snippets to every page. So when a user visits every page JavaScript code is activated and user information and other event details are stored in a separate file. One of the advantages is it counts page visits by the user whenever the page is loaded[1].

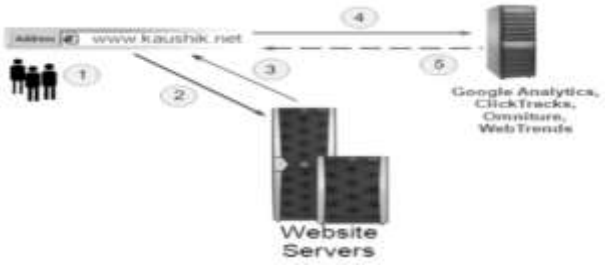

Figure 2: JavaScript Tagging Data Collection Visualization. Source: [3]

JavaScript tagging is shown in figure 2. When a user types a URL in a web browser web server receives a request from the user agent and along with the response, it sends a JavaScript code snippet. When the page loads, it executes and details about the session and other information are captured, and data will be sent to the analytics server.

5.3 Web Beacons

Here web beacons are used to collect analytical data from web pages. In this method, a "tracking pixel" which helps in measuring banner impressions and click-throughs [1]. For online advertising, especially when using banner ads this method is used. Usually, banner ads are images and one pixel in banner ads is used as a tracking pixel[1].

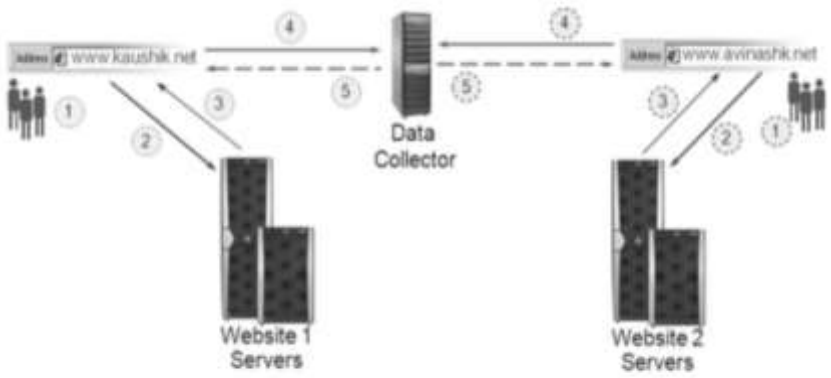

Figure 3: Web Beacon Data Collection Visualization, notice tracking across 2 different websites. Source [3]

Working of web beacon is shown in figure 3. When a user types a web URL in a web browser, a request is received by web servers. A page will be sent back as a response and also a GET request to a third-party server for 1x1 pixel image is also included. It executes a call for a 1x1 pixel image when the page loads up and information will be sent to the third-party server. Then the third-party server sends the image back to the browser along with a code that can read cookies and capture anonymous visitor data.

5.4 Packet Sniffing

The technique of gathering, collecting, and logging some or all packets that transit through a computer network, regardless of how the packet is addressed, is known as packet sniffing. These packets can be utilised to do further analysis. The first component of a packet sniffer is a network adapter that connects it to the network, and the second component is software that allows you to evaluate the data obtained from the devices. Package sniffer is present between a web server and user. A packet sniffer is software or hardware that resides between the user's device and the Web Server. All packets which go between a server and user are passed through a packet sniffer and the packet sniffer reads it. The main advantage of packet sniffing is there is no need to add tags because all packets will pass through the packet sniffer. A multiple packet sniffer is required for capturing data on the entire network. A collector may not be able to see traffic that exists on the other side of the router or switches but it can only see traffic which is received by the network adapters.

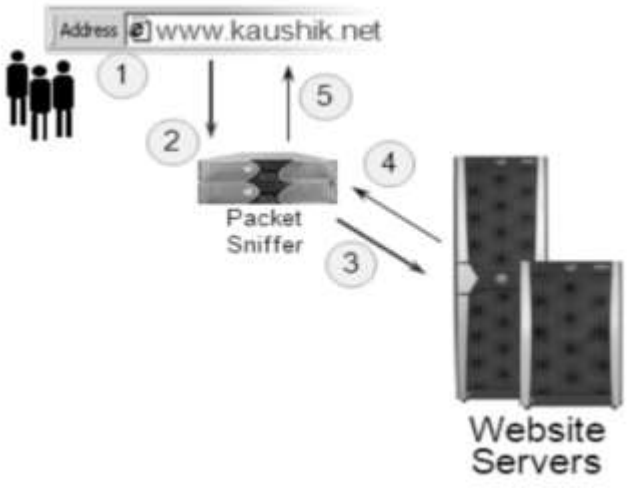

Figure 4: Packet Sniffing Data Collection Visualization. Source [1]

When a user types a URL in a web browser, the request is sent from the user agent and request is received by web servers. Before reaching the web server it passes through a packet sniffer which collects the attributes of the request. After collecting attributes from the packet it sends requests to the web server. When the server sends the response to the server, these packets also go through a packet sniffer. Information about the pages which are going back to the user is captured by a packet sniffer, it stores that data and sends the page to the user's browser [2]. Some packet sniffers append a JavaScript tag that can send

more data about the visitor back to the packet sniffer [2]. Every packet can be collected and can be analysed further using a packet sniffer. For example, collected data can be used for analysis and monitoring of bandwidth and traffic.

VI. DATA SOURCES FOR WEB ANALYTICS

The main goal of web analytics is to collect and analyse data from web traffic in order to get more insight into user behaviour patterns in an application, as well as to evaluate web traffic to the application. A few of the primary source for web analytics are:

1. Direct HTTP request data: This includes data that comes directly from the HTTP header from the browser.
2. Network-level and server-generated information: Much information about the user can be obtained from headers like user browser, version of browser, the operating system user is using, also use IP address etc... And also using user IP many pieces of information like location, region, city etc can be obtained which is very helpful for web analytics.
3. Application-level data associated with the request: Request also contains application-level data such as sessions, referrals etc. and this information can also be used as the data source for web analytics.
4. External Data: Mainly external data is obtained from information that is obtained from the above resources. For example, IP addresses can be obtained from network level information. IP addresses can be used to obtain the latitude and longitude of the user and thus what region, city, country a user is trying to access the application etc... This kind of information can help get insights about how users are present geographically and also how the web traffic from various locations looks like. It also helps in analyzing user behaviour patterns and providing customised contents based on the analysis.

A web client (for example, a web browser) sends an HTTP request to the web server to request a resource. IETF RFC 2616 specifies the format of the HTTP 1.1 request-id. Figure 5 depicts an HTTP request message.

Figure 5: HTTP request header

A request command and HTTP headers make up an HTTP request. The URI, which typically includes the hosts' domain IP address and a directory path, is included in the request command. The host header must be included if the host information is not included as part of the URI. A set of headers that can be included is defined in the HTTP 1.1 specification. The header mainly contains client information. The following are some of the tracking header fields:

- User-agent: The User-Agent is a request header that contains a string that allows network protocol peers to identify the web server's Operating System and Browser. Every request a browser makes to a website includes the user agent information in the header. Example for user agent field- "Mozilla/5.0 (Windows NT 6.1; Win64; x64; rv:47.0) Gecko/20100101 Firefox/47.0". Depending on the browser header's data may vary.

- Referrer: The domain or website address of the page which is making the request is stored in the Referer field of the HTTP request header. When a website redirects to another page or another domain, then the referrer is the page that contains the address or link of another domain or page. This header information helps the server in understanding how users are coming to a website and these details can be used for analyzing caching, optimizing, and more.

- Accept-Language: The accept-language header specifies which locale variant is preferred and which language the client can understand. The language browser provides appropriate values in the header for the Accept-Language field based on the user interface, and even if the user changes it, the value will change. Content of the Accept-Language changes according to the geolocation of the user(like when the user is traveling) and usually, the value of the Accept-Language is often not in control of the user. Few times even the user also wants content to be in the local language of his geographic location.

- Cookie: HTTP Cookies are frequently used to track, customise, and save a website user's session. "session" refers to the time a user spends on the website. cookies are stored as key-value pairs. Cookies contain useful information to the web server like authentication tokens, session details, etc. The cookie field in the request is optional and can be omitted. for example, when a user's privacy setting of a website blocks the cookie of the website.

VII. PRIVACY AND ACCURACY

Privacy is a major concern for all users who use the internet.In web analytics, privacy and accuracy are two major problems. Since the evolution of web analytics which involves tracking of the user, action leads to concern more about user privacy. And also User's privacy settings do affect the accuracy of the data collected through web analytics. Web analytics mainly uses cookies for the collection of user activity information and sends it to the analytical server. Most of the users don't want to share cookies mainly because cookies usually contain private information. When a user blocks cookies for a website the metrics like the number of page visits will be affected and thus inaccuracy in the information will be present. The major browser provides a way to avoid third-party cookies and do not track options for the users.

IIX. ETHICS IN WEB ANALYTICS

User's data is the foundation for web analytics, mainly how the user uses a website/application is mainly tracked in web analytics. Sometimes user's data might be stored in third-party servers and this information is stored without the user's knowledge. Also, there is a lot of concern over what if the stored data is sold for advertising purposes. So there is a lot of concern over ethical aspects of web analytics. According to [10] there are few minimum duties for an organization that collects data for web analytics, they are as follow:

- The customer is the owner of information about them.
- The details about information an organization is collecting from the user and for what purpose should be informed to the user.
- An opt-out option should be given to the customer.
- Customer's information should be bound to the organization and be protected from any third party access.
- Appropriate rules related to management and destruction of users data should be established.

8.1 Privacy-enhancing technologies

Few countermeasures help individuals protect themselves from being identified and benign tracked. But these measures don't guarantee you that all users are not being tracked. Hiding IP addresses is one privacy enhancement technique that helps in hiding user IP addresses and exposing proxy addresses to the internet and user addresses are only exposed to proxy servers. All the traffic from a user will flow through the proxy and only the proxy's address is visible to the destination web server. Also using onion routing mechanism which involves encrypted connections between multiple chained relays. Applications don't directly contact servers instead they send data through tor proxy.

In addition, there exists browser-based blocking and plugins to prevent users from being tracked such extensions are Tracking Protection Lists, Torbutton, Adblock Plus, NoScript, and RequestPolicy. Also, private browsing mode can be used which doesn't cache any information and does not store any user's history. Also HTTP 'do not track' the header which is set by the browser which tells the server that the client doesn't want to be tracked.

8.2 Non-tracking Web Analytics

Keeping user privacy and also collecting web analytics data is a difficult task and there is a lot of research going on in this field where the user has ownership of his data. An example for non-tracking is provided in the paper [8].

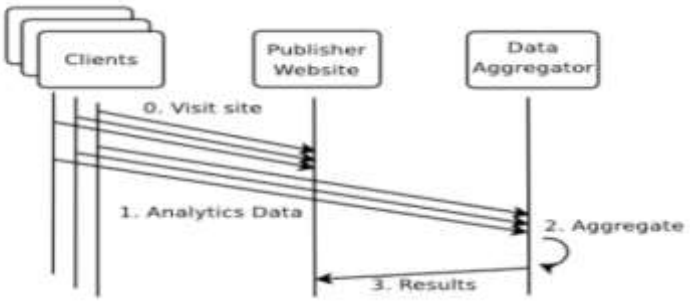

figure 5(a):Today's web analytics system Source: [8]

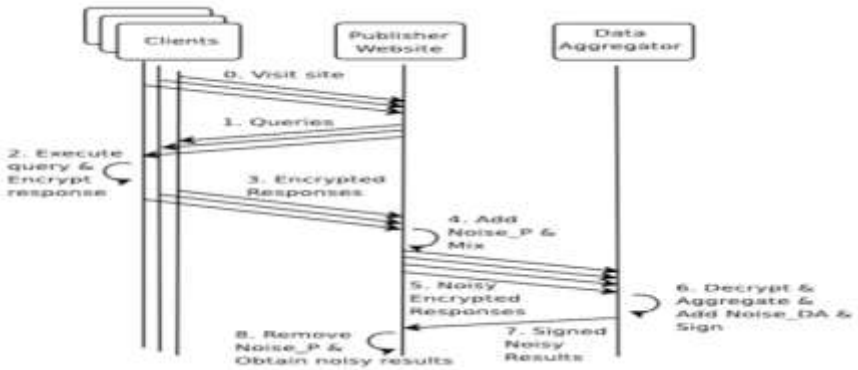

figure 5(b) : Query workflow of [4]'s system. Source:[8]

Web Analytics In addition to having privacy-enhancing techniques on the users' side, there is some research from both academia and the industry to create non-tracking Web Analytics systems to get the important web analytics from users while maintaining users' privacy and ownership of their data. In non-tracking system client's data is present in the client's database and publishers, who want to conduct the web analytics, query the client's database using structured query language. Via publishers client sends data to the data aggregators via publisher. And for security reasons data is encrypted with a public key at the aggregator for protection, in addition, and for more security and privacy, noise is added to the information at the publisher and the aggregator so that data is hidden from others. [8] also, it assumes the publisher and the aggregator are trustworthy. Figure 5(a) shows how traditional Web Analytics with 3rd party servers currently work and Figure 5(b) shows the system presented in [8].

VI. WEB ANALYTICS IN THE FUTURE

9.1 Analytics in mobile application:

Web analytics can also be used for mobile applications which helps in understanding users from mobile applications also. Most of the web analytics tools have been moved to mobile applications also. Analytics in mobile applications is the same as web analytics, in mobile applications one can measure events like how many users have installed applications, demographic aspects of the users who use the application, how often a user uses, what information or menus users are frequently looking for, etc... Also similar to web analytics, user's actions can be tracked.

9.2 Cross-device measurement

Web analytics uses cookies for identifying the users. If a user switches to a new browser or new device in that case it identifies it as a new user. In such a case even though the same user is accessing the application, the user count will be increased. Many web analytics tools like mixpanel provide a way to handle such scenarios very efficiently by merging all these profiles into one profile.

9.3 Exporting data to other data sources

Connecting web data collected from web analytics is easier to integrate with other data sources which helps in the easier analysis of data. Most of the companies collect data of their customers like customer relationship management(CRM) etc.. The main goal is to better understand the cost-effectiveness marketing will drive the growth of integration.

X. CONCLUSION

Web analytics helps in understanding how users are using the application and the behavior of users in the application. It also provides various matrices which help ineffective analysis of the user's behavior in the application. From the product perspective, it helps to improvise existing features and adding a new feature, also helps in understanding and analyzing errors users are facing in the application. As the Internet is growing the number of fraudulent activities is also increasing. Preventing fraudulent activities is one of the important tasks and web analytics helps in understanding the behavior patterns of fraudulent users also. Non-tracking Web Analytics is also growing since it gives importance to the user's privacy rather than user's data and users' are the owner of their data. Web analytics becoming essential for any business which is customer-centric and its demand for web analytics is growing in the market.

REFERENCES

[1] Omar Shaya, Georg-August-Universität Göttingen Computer Science. Web Analytics. Researchgate July 2015.
[2] Michael Beasley UX Designer, ITHAKA Ypsilanti, Michigan, USA. Practical Web Analytics for User Experience How Analytics Can Help You Understand Your Users. Pages: 227-220, 2013.
[3] Daniel Waisberg and Avinash Kaushik. Web Analytics 2.0: Empowering Customer Centricity. In SEMJ.org, volume2, issue 1, 2009.
[4] U. Padma Jyothi, Sridevi Bonthu , B. V. Prasanthi, A Study on Raise of Web Analytics and its Benefits, International Journal of Computer Sciences and Engineering, Volume-5, Issue-10, 2017.
[5] Danielle Booth,USA Bernard J. Jansen Pennsylvania State University, A Review of Methodologies for Analyzing Websites. Chapter VIII, 2019.
[6] World Wide Web. https://en.wikipedia.org/wiki/Internet#World_Wide_Web. June 01, 2021.
[7] Kate Marek, Web Analytics Overview, Chapter I, 2009.
[8] Istemi Ekin Akkus, Ruichuan Chen, Paul Francis (all from MPI-SWS), Michaela Hardt (Twitter Inc.), and Johannes Gehrke (Cornell University). Non-tracking Web Analytics. In CCS '12 Proceedings of the 2012 ACM conference on Computer and communications security, pages 687-698, 2012.
[9] Peterson, Eric. "What Is Web Analytics." Xml.com. N.p., 2017. Web. 17 Oct. 2017.
[10] Kenny, Rebecca, Pierce, Justin and Pye, Graeme 2012. Ethical considerations and guidelines in web analytics and digital marketing : a retail case study. In AiCE 2012 : Proceedings of the 6th Australian Institute of Computer Ethics conference, Australian Institute of Computer Ethics, Melbourne, Vic., pages 5-12, 2012.