# Machine Learning - A New Trend in Web User Behavior Analysis

Chinedum Eunice Chibudike
Federal Institute of Industrial Research Oshodi (F.I.I.R.O.) Lagos, Nigeria

Haruna Abdu
Federal University Lokoja Kogi State, Nigeria

Henry Okwudili Chibudike
Federal Institute of Industrial Research, Oshodi, Lagos-Nigeria

Ogochukwu Constance Ngige
Federal Institute of Industrial Research Oshodi (F.I.I.R.O.) Lagos, Nigeria

Olubamike Adetutu Adeyoju
Federal Institute of Industrial Research, Oshodi, F.I.I.R.O., Lagos-Nigeria

Nkemdilim Ifeanyi Obi
National Oil Spill Detection and Response Agency (NOSDRA) Abuja-Nigeria

## ABSTRACT

Many disciplines, including political economy, linguistics, psychology, social science, and marketing and engineering science, have studied human behavior. As a result, machine learning could be a broad theoretical framework with a wide range of applications, especially in the analysis of internet user browsing behavior. - Aben E. NLANR PMA knowledge, 2010. For testing and calibrating their theoretical models, the aforementioned disciplines used surveys and experimental sampling. The internet logs dealings that store and register each visitor's action on an internet website are the most important sources of information in terms of internet user browsing behavior. Such files may contain incalculable registers, based on the traffic of a web site, and represent a valuable source of human behavior data. This paper examines the rise of machine learning as a trend in analyzing online user behavior, as well as some novel approaches such as knowledge collection and preparation, understanding human behavior, characterization of information analysis in human behavior, internet application, and global application for defining what internet users are looking for in a computing machine..

## Keywords
Machine Learning, Human Behavior, Web User, Browsing Behavior, Web Site

## 1. INTRODUCTION
Ecommerce is now a major tool that encourages businesses to spend substantial resources in online strategies to improve their business processes on the Internet. Local methods of measuring Web usage do not provide the volume of data needed for a comprehensive evaluation of such strategies. - Ackley DH,et. al., 1985. Web analytics is a method that can help organizations evaluate their online strategies more effectively. "The better content we provide for our target audience and adjust our strategy and posting schedules accordingly, the more knowledge we have of the visitors frequenting the content, and the more accurate the evaluation knowledge is. A website user is someone who visits, browses, or communicates with a website, and user behavior describes how people use the internet. The path they take through the web, as well as interactions such as clicks, are all examples of behaviors. When it comes to website optimization, it simply means keeping track of how visitors interact with the site. - Adda M et.al 2017. The actual value or result comes from analyzing users' actions to figure out what motivates them to act in certain ways. Website behavior is complex and varies depending on the target audience. This implies that you must research your users in particular. Who are they? What exactly do they require? Which browsers and devices do they prefer? How often do they go shopping? If you want to have a competitive advantage in meeting the needs of your customers and keeping them, the answers to these questions are critical. You will learn what users are trying to achieve, what motivates them to behave in certain ways, where they have friction, and how to improve the user experience by studying customer behavior online - Ismaeel S et.al. 2018. Certainly, understanding how visitors behave on your website allows you to provide a better experience that is in line with user needs, allowing your company to continue to expand. Google Analytics is one of the most widely used research tools. This is free and equally. Ecommerce is currently one of the most important tools that encourages businesses to invest significant resources in on-line for enhancing business processes on the internet. The native strategies of mensuration net usage fall short of the expertise required for the majority of analyses of such techniques - Ackley DH, et al., 1985. After a thorough examination of online methods, web associate analytics is a technique that can help structure your search. "The more data we have about how often visitors visit the content, and therefore the more accurate the analysis data is, the more content we can provide to our target market and adjust our strategy and posting schedules accordingly. A person who accesses, browses, or interacts with an internet site is referred to as a site user, and user behavior refers to how people use the site.

## 2. BACKGROUND
Online user behavior analysis is a vital field of study that allows researchers to examine a variety of user characteristics. Business houses and industries can identify target areas to focus on by studying user behavior and predicting their intentions toward specific products. The journey they take from the location to the relationships, such as clicks, is embodied by their behaviors. When it comes to website optimization, it simply means observing the behavior of visitors to the site - Adda M et.al 2017. The $62,000 value or outcome comes from analyzing users' actions to get to the root of what motivates them to behave in the way they do. Website behavior is sophisticated and varies depending on the target market. This indicates that you should examine your users in particular. They're from the World Health Organization, right? What exactly do they require? Which browsers and devices are the most popular among them? Do they, on the other hand,

generally go for his or her purchases? If you want to gain a competitive advantage and meet and retain your customers, you'll need to know the answers to those questions. You can learn what users are trying to achieve, what motivates them to behave in certain ways, where they encounter friction, and how to improve user. experience by conducting online study of customer behavior- Ismaeel S et.al. 2018. Certainly, understanding how visitors behave on your website gives you the opportunity to provide associate enhanced experience that is in line with user wishes, allowing your company to continue to grow. Google Analytics is one of the most widely used tools for conducting research. This is often free and reasonably priced, as well as simple to arrange. Once installed, you'll have access to useful data on how people interact with your website, including where they land and go next, where they branch off, and what they do next. You'll also use it to predict overall trends and patterns, as well as provide growth opportunities, and it's pretty simple to set up. You'll have access to useful information about how users behave on your site once it's installed, such as where they land and go next, where they branch off, and what they communicate with. You'll also use it to predict overall trends and patterns, as well as identify growth opportunities

# 3. RELATED WORKS

## 3.1 A Machine Learning Application for Sales Forecasting (Danilo Sato 2019)

Since 2016, we've been considering how to apply Continuous Delivery to Machine Learning systems, and we've written and presented a case study from a client project we worked on with AutoScout to predict prices for vehicles that were published on their platform. However, since we are not allowed to use examples from actual client code, we have chosen to create a sample ML application based on a public issue and dataset to demonstrate a CD4ML implementation - Deshpande D et al 2017. This application addresses a common forecasting issue that many retailers face: Using historical data, try to predict how much of a given product will sell in the future. We've combined and simplified their data sets for our purposes, as our objective isn't to find the best predictions — that's a job better left to your Data Scientists — but to show how to use CD4ML.

We train a prediction model using the labeled input data and integrate it into a simple web application, which is then deployed to a cloud production environment using a supervised learning algorithm and the popular scikit-learn Python library. The high-level process is depicted in Figure 2. (https://martinfowler.com/articles/cd4ml.html).

# 4. METHODOLOGY AND ANALYSIS

## 4.1 Data Collection

The data were collected from data logs transaction in the data warehouse, or if you need to populate it with more data sources, you then need to have a way to collect and store data consistently into a database through the website back end.

## 4.2 Data Preparation

Raw data is similar to a fact that incubates the truth; it needs to be refined before it can be considered genuinely valuable. In the world of data, refinement refers to the extraction, cleaning, and transformation of raw data into something useful for the analysis you're doing. The information was divided into users in this case. Before moving on to the actual analysis, we arranged each user's events in chronological order. Clickstream data, unlike other data sequences, can vary in length from user to user.

To convert the previously collected event data log into clickstream data, we'll need to do the following:

- Find events/actions taken by a specific user and group them together.

- Divide them into subgroups of events based on which ones were performed during the same session, as defined by the session definition above.

At this point the dataset that will be obtained for the rest of the analysis looks like this:

**Table 1: Dataset of Users**

| User 25 | A1 | A3 | A9 | A9 | A5 | A6 | | | |
| User 26 | A1 | A1 | A3 | A28 | A29 | A29 | | | |
| User 27 | A1 | A1 | A3 | A10 | A10 | A5 | A10 | | |
| User 28 | A1 | A3 | A30 | A30 | A5 | A30 | A30 | A3 | A1 |

## 4.3 Analytics

User Behavior Analytics (UBA) examines what the user does, including the apps that are released, network activity, and the majority of files that are accessed (when the file or email was opened, who opened it, what was done in it and how often). Whether the operations are the result of a hacker, an insider, malware, or other processes, UBA technology looks for patterns of usage that indicate unusual behavior. While UBA may not be able to prevent hackers or insiders from gaining access to the system, it can still detect their work and limit the harm they cause - Martin Aet.al. 2011. My website is now integrated with Google Analytics, a web analytics platform. The tool that enables users to track their activities and the amount of time they spend on the website's various actions. The analytics data was gathered and stored from a variety of users. I gathered data from approximately 113 users, with different data points such as:

- The amount of time users spends on different pages of the website,

- The filters that users use,

- The total amount of time users spend on a particular product page, and so on.

## 4.4 ML Model and Data Analysis

Many platforms, such as Facebook, rely on the generated data from what a user clicks, so the difficult clickstream data can be identified. We must first be able to capture a user's activity across a web page or application in order to analyze. clickstream data. And any Internet marketer will find this extremely useful. Obtaining a 360-degree view of a customer by determining what he clicks and what he does not can have a significant effect on both your products and your customers' experiences.

## 4.5 The big challenge: To understand human behavior by data analysis

With vast quantities of data being gathered at the moment, gaining access to it is a top priority. Information is being generated and stored at a breakneck speed, and much of the big data gathered is based on human behavior. The information we provide from our credit cards, smartphones, televisions, computers, smart speakers, and insurance tracking devices captures human behavior. Organizations can use this data to make better decisions and implement stronger policies by sifting through it and gaining insights into human behavior - Saleh M. et. al. 2018.

When you examine your customers' actions, you may notice mistakes due to things you can't see, but nothing more than an unorganized and seemingly random jumble of events and actions. Machine learning aids in the detection of hidden patterns and the appropriate handling of them. Data from not only search log history, but also customer interactions that may occur on your website. The concept is simple: if your data set is complete, your target and design efforts will be successful.

# 5. RESULT AND DISCUSSION

We train a prediction model using the labeled input data and integrate it into a simple web application, which is then deployed to a cloud production environment using a supervised learning algorithm and the popular scikit-learn Python library. The high-level process is depicted in Figure 1- *Danilo Sato et.al. 2019.*
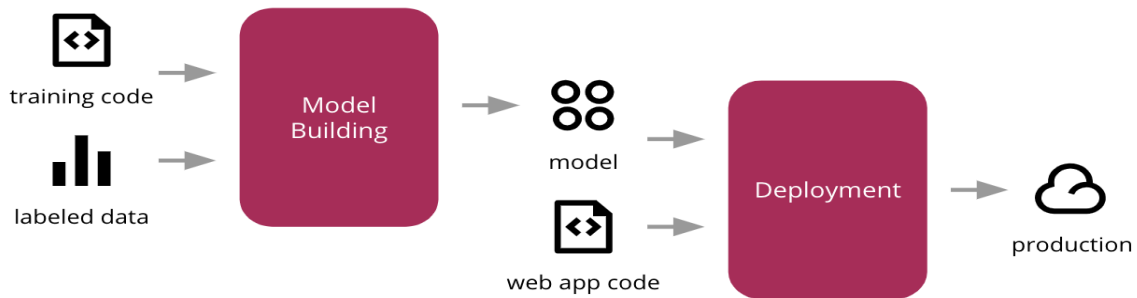


**Fig 1: If necessary, the images can be extended both columns**

Machine Learning applications are becoming increasingly prevalent in our industry, but the process of creating, deploying, and continuously improving them is more difficult than developing, deploying, and continuously improving more conventional software, such as a web service or a mobile application. The coding, the model, and the inputs are all subject to change on three axes. When our web application is launched, users can choose a product and a future date, and the model can predict how many units of that product will be sold on that day - *Danilo Sato et.al. 2019.*

## 5.1 Web Application (RIS)

E-commerce websites are online portals that improve online transactions of goods and services by means of information and funds transfer over the Internet, according to the history of the World Wide Web. In the beginning, e-Commerce was primarily conducted through emails and phone calls. Anything and everything that a transaction requires can now be completed online.

- First and foremost, I purchased a domain name (Kitaab.cloudaccess.host) and hosting from cloudaccess.net, which also provided me with Jumla access.

- After that, I use a pre-existing Jumla theme called Astra and modify it to fit my needs.

- The following pages can be found on my website:

a) Home- This is my website's main page.

b) Store- It includes a variety of goods divided into various categories.

c) Account- It includes the user's account information.

d) Contact Us- This section includes the company's contact information.

e) About- This section provides information about the team.

- The website contains various features:

a) View a list of products

b) View product details

c) Search products

d) Use filters to change the product list (eg. Category, price range, etc.)

e) Add a product to the cart.

f) Payment Gateway.

- User Guiding- I have applied the user guiding tool named WalkMe that will guide a new user through orderly steps on my website to use it in the simplest way to buy a product.
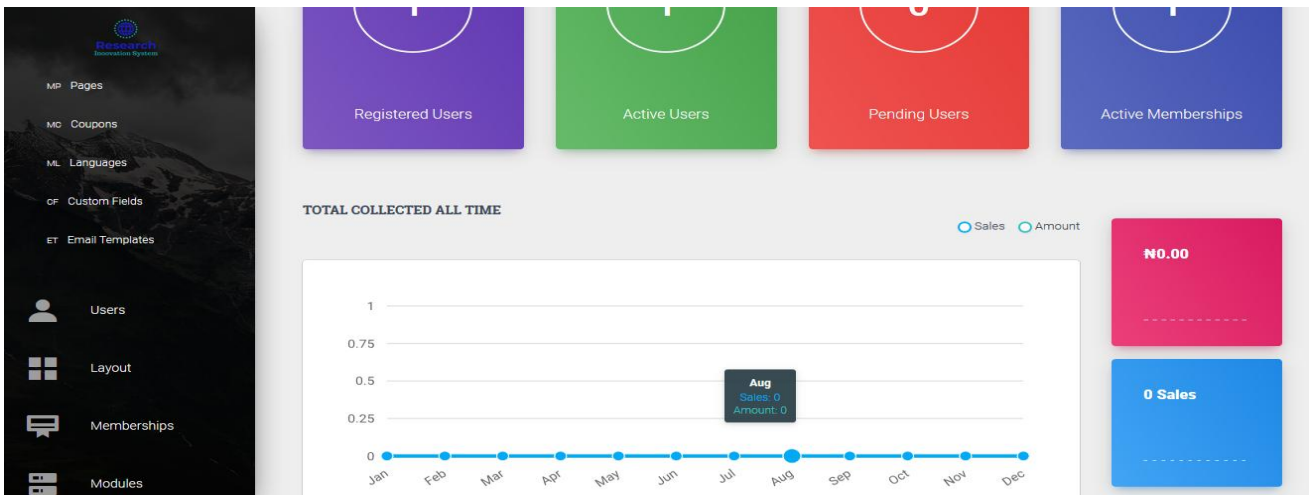
**Figure 2: Administrative page**

## 5.2 Characterizing the Web User Browsing Behavior

In this study, I use search and toolbar logs to conduct a large-scale investigation of online user behavior. Content (news, portals, games, verticals, multimedia), Communication (email, social networking, forums, blogs, chat), and Search (email, social networking, forums, blogs, chat) make up a new CCS taxonomy of pageviews (Web search, item search, multimedia search) - Shirazi F et.al. 2017. We discovered that content accounts for approximately half of all online pageviews, communications for one-third, and search for the remaining one-sixth. We then break down the pageviews further to characterize them within each high-level category. The extent to which the same user revisits certain types of pages over time, as well as the mechanisms by which users move from page to page, within and across hosts, and within and across page types, are investigated. We look at reliable methods for assigning responsibility for a pageview to ancestors in the referral chain. We demonstrate that mail, news, and social networking pageviews are insular in nature, appearing mainly in homogeneous sessions of a single type - Adams R. et al. 2013. Search pageviews, on the other hand, appear to be a significant contributor to a large number of pageviews, but they are not the primary mechanism by which those pageviews were obtained. Finally, we investigate the burstiness of pageviews associated with a URL, demonstrating that "breaking" material with non-uniform visit frequency has little effect on online browsing behavior.
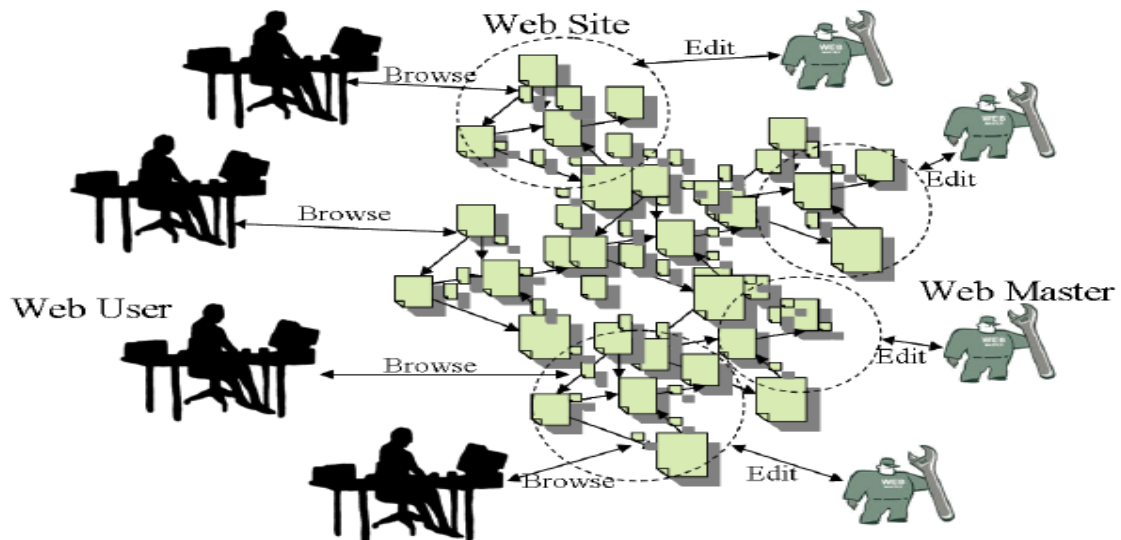


**Figure 3: New Trends in Web User Behaviour Analysis – (Juan Domingo Velasquez)**

Understanding oneself has been a challenge for mankind since the dawn of civilization. Traders anticipate people's needs, politics calculates the best political move, and generals determine the army's position. Human beings coexist in cultures that form complex interdependent systems - *Deshpande D. et al. 2017*. Google is a firm whose principal business is targeted advertising on a search result page. Google's marketing strategy is based on the preferences of web users for web page rankings. For calculating this ranking, a simple stochastic model for web user browsing produces stationary probabilities. As a result, the Internet has emerged as a new frontier in the marketing field, promising commercial success at the expense of the need for precise information about web users. Some successful e-commerce examples could be cited, such as Amazon.com.

**Table 2: Browser statistics in between year 2019/2020**

| 2020 | Chrome | Edge/IE | Firefox | Safari | Opera |
|---|---|---|---|---|---|
| August | 81.2 % | 4.6 % | 7.3 % | 3.4 % | 2.0 % |
| July | 81.3 % | 4.3 % | 7.6 % | 3.4 % | 2.0 % |
| June | 80.7 % | 3.9 % | 8.1 % | 3.7 % | 2.1 % |
| May | 80.7 % | 3.5 % | 8.5 % | 4.1 % | 1.6 % |
| April | 80.7 % | 3.4 % | 8.6 % | 4.2 % | 1.5 % |
| March | 81.4 % | 3.5 % | 8.7 % | 3.7 % | 1.3 % |
| February | 82.0 % | 3.4 % | 8.7 % | 3.4 % | 1.2 % |
| January | 81.9 % | 3.0 % | 9.1 % | 3.3 % | 1.3 % |
| 2019 | Chrome | Edge/IE | Firefox | Safari | Opera |
| November | 81.3 % | 3.2 % | 9.2 % | 3.5 % | 1.4 % |
| September | 81.4 % | 3.3 % | 9.1 % | 3.1 % | 1.6 % |
| July | 80.9 % | 3.3 % | 9.3 % | 2.7 % | 1.6 % |
| May | 80.4 % | 3.6 % | 9.5 % | 3.3 % | 1.7 % |
| March | 80.0 % | 3.8 % | 9.6 % | 3.3 % | 1.7 % |
| January | 79.5 % | 4.0 % | 10.2 % | 3.3 % | 1.6 % |
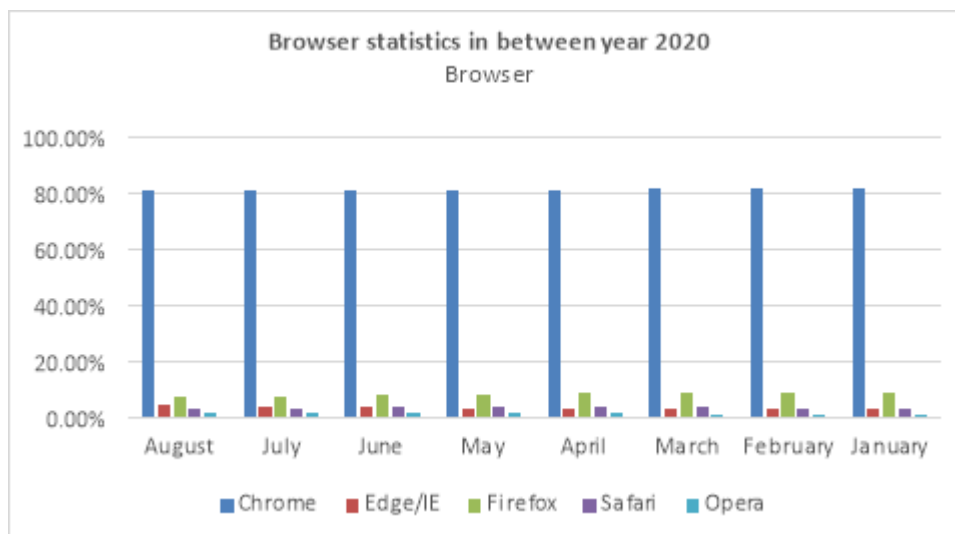


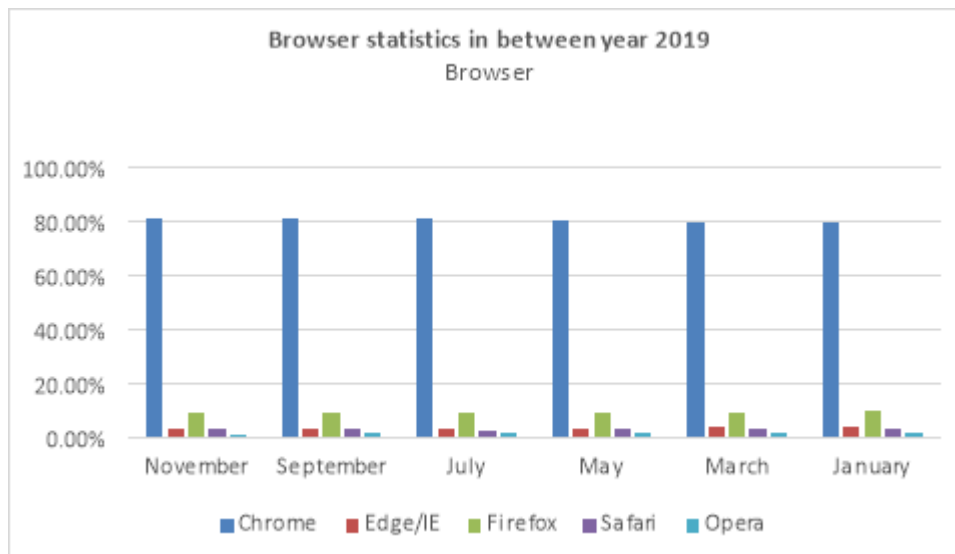**Figure 4: Browser statistics in between year 2020 Browser**



**Figure 5: Browser statistics in between year 2019 Browser**

## 5.3 Real world application

### 5.3.1 Inform Content Creation

Perhaps no digital initiative has been as popular in recent years as content marketing. Brands are churning out content at a breakneck pace in the hopes of increasing brand loyalty, awareness, and, ultimately, conversion. According to a recent study by the Content Marketing Institute, only 35% of business-to-business advertisers have a recorded content marketing strategy - *ACM Special Interest Group on KDD, 2017*. Furthermore, less than half of respondents (48%) claim to have a content marketing plan, but it isn't recorded. The tide is turning, which is good news for data-driven businesses. Tableau, a business intelligence and analytics software provider, has named social intelligence as the second most important trend in business intelligence, written, "Tracking conversations at scale through social will make companies find out when there is an introduction of topic as a new trend and what their customers are talking about." Brands monitor social conversations (via Hootsuite, Sprout Social, Brandwatch, and others) in order to make content creation choices based on what has happened (what consumers are talking about) and what could happen (if they could create content owing to the existing social conversations). Uprising of it is a new trend in business intelligence that allows content marketers to find and analyze high-performing content within their vertical, see which social networks react best to content, and see who influences the trend and who shares it across verticals, publishers, and particular articles. (see sample of image

https://www.shutterstock.com/search/web+content+creation).



**Figure 6: Content Image**

## 6. RECOMMENDATION

Machine learning is a new trending tool that has proven to be effective. It enables us to revolutionize work in a variety of seemingly unrelated fields. Machines now assist humans in everything from market forecasting and algorithmic trading to predicting when a bridge is likely to collapse, so it must be adopted for improved industry performance. Machine learning has yielded a slew of unexpected benefits for web developers, and it promises to have even more impact as the associated technologies improve. While machine learning-powered Artificial Intelligence (AI) is unlikely to replace human programmers and web developers anytime soon, the fact that machines have demonstrated the ability to sift through large amounts of data and discover critical patterns already shows their importance. It should be urged to improve prediction accuracy. If given the chance, machine learning has a good chance of fundamentally altering the process of website and web app development.

## 7. CONCLUSION

Recent events have demonstrated that Machine Learning can be used to automate a variety of tasks previously thought to be only human-capable, such as image recognition, text generation, game play, and web browsing behavior analysis. We used toolbar logs from a commercial search engine in this study to characterize the browsing behavior of young users, particularly to understand the activities on the Internet that trigger searches - Aben E. et.al 2017. We determined the probability of a user conducting a search on the Web and multimedia verticals (i.e. videos and images) owing to the previous event being another search or browsing event by measuring the percentage of browsing and search activity in toolbar sessions. These metrics clearly show an increase in the level of uncertainty and unsuccessful search sessions among children, according to our findings. We also discovered a strong link between the reading level of the pages that were clicked and user traits such as educational attainment and age. In terms of browsing behavior, children are more likely than adults to start their Internet activities with a search engine (rather than directly browsing content) - Ackley DH et al 1985. We also found that teenagers engage in far more browsing activity than adult users. Furthermore, we discovered that when children visit knowledge-related websites (i.e., pages with a lot of information, like Wikipedia articles), they appear to conduct more web searches than adults. Furthermore, children, especially adolescents, were found to have a higher proclivity for multimedia search, implying that the aggregation of multimedia results into the current search result should be improved. Machine Learning is going to have a big impact on the economy and everyday life in general. Entire sectors and work assignments will be programmed, and the job market will be forever altered - Adda M et.al 2017. Machine Learning Engineers are desperately needed everywhere, and many businesses want to get their foot in the door of Machine Learning and Artificial Intelligence, so now is the best time to start learning about it. Now that Machine Learning is the norm, it's up to us, the engineers and researchers, to push for even more significant advances in the field of Machine Learning.

## 8. ACKNOWLEDGMENTS

## 9. REFERENCES

[1] Aben E. NLANR PMA data. 2010. https://labs.ripe.net/datarepository/data-sets/nlanr-pma-data. Accessed 27 Dec 2017.

[2] Ackley DH, Hinton GE, Sejnowski TJ. A learning algorithm for boltzmann machines. Cogn Sci. 1985; 9(1):147–69.

[3] ACM Special Interest Group on KDD. KDD cup archives. 2016. http://www.kdd.org/kdd-cup. Accessed 22 Nov 2017.

[4] Adams R. Active queue management: A survey. IEEE Commun Surv Tutor. 2013; 15(3):1425–76.

[5] Adda M, Qader K, Al-Kasassbeh M. Comparative analysis of clustering techniques in network traffic faults classification. Int J Innov Res Comput Commun Eng. 2017; 5(4):6551–63.

[6] Deshpande D, Deshpande S (2017) Analysis of various characteristics of online user behavior models. Int J Comput Appl 161(11):5–10.

[7] Ismaeel S, Karim R, Miri A (2018) Proactive dynamic virtual-machine consolidation for energy conservation in cloud data centers. J Cloud Comput 7(10):2018.

[8] Martin A, Anutthamaa NB, Sathyavathy M, Manjari M, Francois S, Venkatesan P (2011) A framework for predicting phishing websites using neural networks. Int J Comput Sci Issues 8(2):330–336.

[9] Saleh M, SHETTY P, Nisha (2018) Analysis of Web Server Logs to Understand Internet User Behavior and Develop Digital Marketing Strategies. Int J Eng Technol 7(4.41):15–21.

[10] Shirazi F, Iqbal A (2017) Community clouds within M-commerce: a privacy by design perspective, Community clouds within M-commerce: a privacy by design perspective. J Cloud Comput 6:22