Quality Enhancement of Web APP using Big Data analytics

Kavita Tandon[1], D. P. Mishra[2]

[1]M.Tech Scholar, Department of Computer Network

[2]Associate Professor, Department of Computer Science and Engineering

[1,2]Bhilai Institute of Technology, Durg, India

(Email :[1]tinkytandon@gmail.com)

Abstract— Big data includes those data sets which are so complex that traditional data mining and data processing application software is not enough to deal with them. Studying of server log data can provide remarkable and useful information and facts about logs. Stored information and facts can help to find out user perception and behavior towards the web application. This can improve the effectiveness of the websites by adapting the information structure to the user's behaviour. In this paper, we propose a novel technique for select big data analytics approach since it can enhance the worth of web contents. Several web usage mining methods for extracting useful features are discussed employing all these techniques to cluster users of the domain to study their behaviours comprehensively. Our main aim is to develop a web application that gives web log and access information which will be used further to analyze user behaviour by using BIG DATA HADOOP technology for mining enriched web access log data. The contributions of this research are a data enrichment that is content and source based, and a tree-like visualization of frequent navigational sequences.

Keywords— Big data, Hadoop, Web Mining, Web App, etc.

I.   INTRODUCTION

Nowadays, Internet has become a household entity without which our daily life cannot be imagined. Web information has turned out to be more prominent and because of that web mining has pulled in part of consideration [1] for web content mining and web utilization mining. The 'Web Utilization Mining' is defined as the programmed disclosure of client's interest usage patterns from web servers. Web utilization mining is an imperative innovation for understanding client's practices on the web. This can be utilized as a part of assortment of uses, for instance, one can monitor beforehand visit to web pages of a client. These pages can be utilized to distinguish the ordinary conduct of the client and to make forecast about sought pages [4], webpage by making groups of clients with comparable designs and by including navigational connections.

Visit conduct for the clients can be utilized to recognize required connects to enhance the general execution of future accesses. Perfecting and reserving approaches can be made on the premise of much of the time in using pages to enhance inactivity time. In addition, utilization examples can be utilized for business knowledge in request to enhance deals and notice.

Web structure mining is the way toward finding the association between site pages. Web content mining incorporates mining, extraction and combination of valuable information and learning of Web page content. Web Usage Mining is a strategy of extricating valuable data from the Web Log, e.g. the example in which a client experiences distinctive Web pages [2, 3]. Web applications are expanding at a colossal speed and its clients, are expanding at exponential speed [4].

A.   Web Usage Mining

Web mining is the reconciliation of data accumulated by conventional information mining strategies and systems with data assembled, as shown in Fig. 1. Web Usage Mining is the utilization of information mining procedures to find intriguing use designs from Web information keeping in mind the end goal to comprehend and better serve the requirements of Web-based applications. Utilization information catches the character or root of Web clients alongside their perusing conduct at a Web website [5, 6].

B.   Big Data Approach

There is a lot of growth in web data. The data has been so large it wouldn't be handled by traditional mining methods. Big data is a terminology that express the large volume of data – both structured and unstructured. The 3Vs have been expanded to the characteristics of Big Data [7].

Volume - size of data is now larger than terabytes and petabytes. This large scale makes it difficult to analyze using conventional methods.

Velocity - Big Data should be used to mine large amount of data within a pre-defined period of time. The traditional methods of mining may take huge time to mine such a volume of data.

Variety - Big Data comes from various sources. It is designed to handle structured, semi-structured as well as unstructured data.

Hadoop: Apache Hadoop is open source software which processes on large scale storage on commodity hardware. Hadoop has been designed such that its software framework can automatically manage and deal with hardware failure.

C.   Big Data Research History

Though Big data is being seen in almost all organization, it is still in developing phase with its uncertain origin. According to researchers, it got introduced in mid-nineties only, but as shown in Fig. 2. It became widely popular since current decade. IBM and organizations are doing lots of promotional events and activities for Big Data to build the niche analytics market. Enormous information definitions have advanced quickly, which has raised some perplexity.

Obviously, estimate is the main trademark that rings a bell considering the inquiry "what is huge information?" However, different qualities of enormous information have risen as of late. For example, its recommended in [10] that Volume, Variety, and Velocity (or the Three V's) are the three
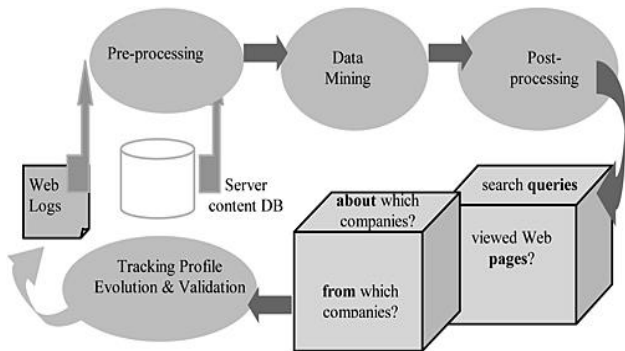

Figure 1 Web Usage Mining Approach

measurements of difficulties in information administration. The Three V's have risen as a typical structure to depict enormous data. For instance, Gartner, Inc. characterizes huge information in comparable terms: "Huge information is high-volume, high-speed and high-assortment data resources that request practical, imaginative types of data preparing for upgraded understanding and basic leadership."

II.   RELATED WORK

A. *Understanding Requirements for Online Services based on Users Behavioral Data Analysis*

Client behavioural information gives critical prompt to shortcomings in the present programming outline and focuses for change in future [8]. The present online programming administrations suppliers could gather several types of administration use information, which is a decent wellspring of data which can help engineers better comprehend client's practices. This paper takes the case of the development of online lexicon administrations, investigates answers for inquiries, for example, what are the key pointers as to the nature of programming item outline and administration substance, and how to improve novel item accomplishment by understandings got from client behavioural information examination.

In this paper, they initially outline the conceivable sorts of web log information that can be gathered up until this point. Then a client conduct produced show word reference space in light of client behavioural information examination. From this we talk about how information gathering capacity should co-develop with the online administrations, and how particular information examination methodologies such as mining calculations and factual capacities can help create fascinating examination after-effects of significant worth to plan choice producers.

Because of the fast improvement of Web applications, increasingly and more programming administrations are conveyed over the Internet [9]. Clients of the present web applications are profoundly differentiated, whose needs are advancing continually. Keeping in mind the end goal to cover the assortment of clients and stay aware of the dynamic changes of their needs, specialist organizations need to suit clients with distinctive aptitudes, information, age, sexual

orientation, inabilities, education, culture, pay, and foresee what they will require in light of the current accessible procedures. For instance, site interface fashioner requirements to think about the clients' processing abilities and learning.

Fashioner of web indexes should incorporate essential and progressed scans for various clients. However, most existing investigations depend on strategies or exercises led physically, other than robotized information investigation. There are numerous run of the mill methodologies of client examine. Ease of use testing is a broadly embraced approach for assessing whether a product item is anything but difficult to utilize, which can be led through the life cycle of programming to get data from clients. During the time spent ease of use testing, just a few clients are insufficient to discover all ease of use issues [10].

A few others think expanding the quantity of clients included will increment of the shot of recognizing issues with low event rate. It can track the mouse development of clients on the site to perceive how the clients collaborate with the framework when displaying. Eye-following framework, which can yield a flood of focuses at which the eye was looking, has likewise been utilized to see the clients' concentration when utilizing the programming. Customary client techniques, for example, centre gather, poll inquire about and inside and out meeting with clients are additionally broadly utilized. These techniques have been utilized for numerous years and have been demonstrated helpful. In this paper, we initially condense the conceivable kinds of web log information that can be gathered up until this point. Next it discusses about how information gathering capacities should co-evolve with the related online administrations, and how particular information examination methodologies, for example, information mining calculations and measurable capacities can help create intriguing investigation consequences of significant worth to plan chiefs [11].

This paper is a preparatory advance taken to get a few bits of knowledge into prerequisites elicitation in light of client behavioural information investigation. It set out from condensing the distinctive sorts of utilization information that is at present available to general on the web specialist co-ops, this incorporate solid client activities arrangement, eye development following time spent on various capacities, content related follows, and connections made, and clients profiles, and furthermore measurable information, for example, data about the sellers, the distinctive adaptations of items, the IP addresses mirroring the geological area of the clients, the characters of the clients, the day by day benefits solicitations of the clients, the time and frequencies of administration asked for, and client's criticisms and appraisals on the item or administrations, and so forth. In light of the information open to an online word reference venture, they have recognized a couple of situations and made some instinctive investigation to general understandings can be acquired by information investigation in area.

Their investigation demonstrates that, because of the reality that both the data on utilizations of new advances and data on dormant buyer needs are required. In the disclosure stage, while strategies that rely upon the "visionary aptitudes" of customers are most suitable to certain degree, the most proper strategies rely upon the innovative knowledge get from a dynamic information investigation to customers or try different things with a few variants of item ideas. By investigating client's behavioural information, we are including clients in the item considering naturally. The discoveries of our contextual

Figure 2 Big data Analysis
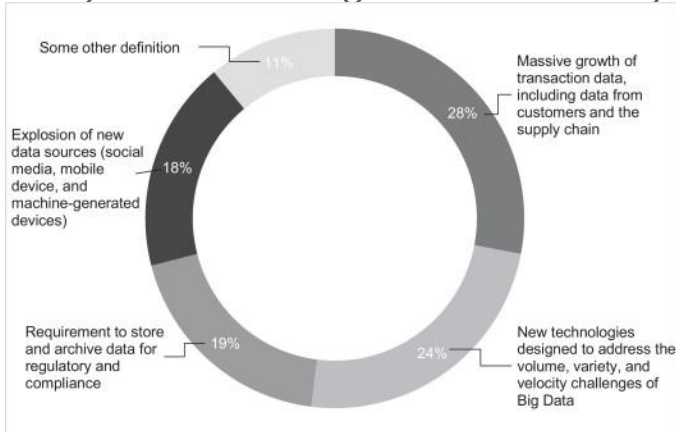
analyses in the online word reference give prove that organizations proactively utilizing information examination to help the advancement of new item highlight are useful. In addition, the outcomes demonstrate that data that is created by the inclusion of buyers is a standout amongst the most essential factors providing guidance to the improvement of item developments. Constraints of the present examination are both routes: To begin with, we will work even more intimately with item planners to utilize information examination results to coordinate their choices on new discharges and highlights. At

that point we investigate client information investigation for results of various natures

B. Customer behavior analytics and data mining

Client conduct examination depends on purchaser purchasing conduct, with the client assuming the parts of client, payer and purchaser. The worry of numerous associations is no longer on the individual purchaser yet rather on aggregate or authoritative purchasing conduct which helps in figuring out which client's merit creating and overseeing by setting up exceptional techniques keeping in mind the end goal to pull in particular clients. Through examination of clients' conduct, exact profiles are being created by indicating needs and intrigue and enabling business to give clients what they need it, when they need, prompting a superior consumer loyalty accordingly holding them to return for additional. While expansive scale data innovation has been developing separate exchange and logical frameworks, information mining gives the connection between the two. Information mining programming investigates connections and examples in put away exchange information in view of open-finished client inquiries. Considering past examinations creators discovers the extension to go for explore in showcase container examination utilizing three unique calculations in particular Association Rule Mining, Manage Induction Technique and Apriori Algorithm. Creators will make a relative report of three methods and receive the best conclusion [12].

In well-run independent venture associations, learning associations with clients is framed normally; the associations take in their client's conduct through individual associations with every one of them. They take in more and more about their clients over time and after that utilization the information to serve them better, therefore, clients are faithful to the associations and the business benefit increments. Bigger organizations with thousands or a large number of clients despise this extravagance of having individual associations with every client, these vast associations must depend on other

methods for learning client conduct that will enable them to anticipate effectively what clients like, for example, their necessities. Assessing the execution of any association is a fundamental part to overcome these shortcomings [13].

Clients who visit destinations abandon significant data about their conduct, client conduct examination means to enhance business execution through an understanding for at various times clients in order to decide and recognize future clients and their conduct. Cooperating with your clients is additionally not as straightforward as it has been before [14]. Clients and forthcoming clients need to cooperate on their terms, implying that you have to look at numerous criteria while assessing how to continue. Such as:

- The Right Offer
- To the Right Person
- At the Right Time
- Through the Right Channel

This has given us the chance to build up an application that investigates the database and concentrate profitable data which will help administration with basic leadership as respects client conduct, deals design and potentially foresee future deals precisely. Comparison of online investigation preparing (OLAP) and information mining, we depicted OLAP as a PC procedure that empowers clients to effectively concentrate and view information from various perspective and furthermore OLAP server is required to compose furthermore, look at data. OLAP gives synopsis yet information mining gives understanding and insights about the conduct of singular client. However, Data mining and client conduct, both require a comprehension of past client's utilization conduct and information extraction made conceivable by mechanical headway [15].

Customer relationship administration is an innovation that deals with the connection amongst associations and their clients, the key individuals being the database advertisers. Additionally, specified in this work are the phases of information mining starting with the beginning investigation, trailed by show building or example distinguishing proof and ultimately organization. Comprehending the above points, distinctive strategies for information mining procedures has also been discussed like neural systems; choice tree and affiliation lead referring to a few cases of their application.

C. Advertisement Posting based on Consumer Behavior

Long range informal communication and other online internet business destinations enable buyers and dealers to convey specifically with respect to their things to be acquired/sold, brands and items. The promotions posted in these sites regularly not intriguing to the customers as these locales pour in pointless promotions. Its proposed a shopper conduct demonstrate utilizing which pertinent promotions can be presented on the shopper at whatever point they visit the site. This model breaks down the procedure and exercises individuals take part in while hunting down, choosing, obtaining, utilizing, assessing furthermore, discarding items and administrations to fulfill their necessities and wants. We have developed a versatile model system in light of this customer direct show using Hadoop Framework which picks and shows critical notification in the site according to the general slant of the customers. Long range casual correspondence and other online web business goals empower customers and merchants to grant particularly concerning

things to be purchased/sold, brands and things. The ads posted in these locales every now and again not captivating to the clients as these goals pour in silly ads. In case the plugs are posted in light of the interests of the customers by then there is a likelihood that these advancements can accomplish customers and in this way the business would increment be able to for the ones who posts these commercials [16].

In light of the past joint efforts of the customers with the web-based systems administration site page the promoters can look at the direct of the customers and interpret territory, interests and economics (e.g. 'sports enthusiasts') of the customers and after that mean to introduce material notification on give a better trial than the customers. The interests of every individual are extraordinary when contrasted with others. For instance, a portion of the clients (In this paper, we have utilized clients and purchasers conversely) might be occupied with watching satire motion pictures, blood and gore flicks, watching parody appears, watching features of cricket matches, data in regards to birthday events of VIPs, getting data in regards to new mobiles in the market, new advancements and their applications, propelling of rockets and satellites, web based looking for dress and adornments, and so forth.

In this paper, client activities and interests are recorded in a log. In this log design, user ID and its details, day, login time, logout time, page subtle elements (pages went by amid a session by the client), time spent on page and different points of interest are saved which is put away in the Hadoop Distributed File System (HDFS). These log points of interest are gathered by the accompanying following projects which are executing in the online networking site [17] at the foundation. These projects will catch the conduct of the clients, point of interest and afterward record the same in the log. One of those systems which records the above is Clicking Advertisements Program. All the client taps on commercials in the site are followed and put away in HDFS by this module. Different modules like following preferences, following aversions, following shared subtle elements, and so forth can track the client activities and record the same in the log. At that point we bunch the clients in light of their basic advantages, obtaining or offering examples, conduct and as needs be conveying significant ads to the objective clients. We trust this kind of commercial will upgrade the matter of the association which posts ads in these sites.

In this paper, we have proposed a shopper conduct demonstrate utilizing which important promotions can be presented on the shoppers at whatever point they visit the site. This model breaks down the procedure and exercises individuals take part in while scanning for, choosing, acquiring, utilizing, assessing and discarding items and benefits to fulfill their requirements and wants. We have built up a versatile model framework in view of this buyer conduct demonstrate utilizing Hadoop Framework. We have likewise demonstrated that this replica framework chooses and shows significant promotions in the site as per the general inclination of the purchasers.

D.   A Data Mining Based Approach to a Firm's Marketing Channel

Firms need to gather and dissect promoting information with a specific end goal to have an upper hand in the division. The point of this exploration is to separate information from a universal company's promoting channel to enhance the effectiveness of the advertising framework. The Cross Industry Standard Process for Data Mining (CRISP-DM) is utilized to examine the review information. Information are bunched by

applying a Kohonen Self Organizing Map (SOM) to lessen the traits. Oddity discovery investigation is connected. We create a C5.0 Choice Tree (DT) display utilized for anticipating the showcasing channel firms' grievances with high exactness. Choice standards are additionally extricated [18].

An advertising channel, which conveys an association's items and administrations to shoppers, is one of the basic achievement factors in a showcasing framework to accomplish advertising destinations. On the off chance that a firm doesn't know its advertising channel well, it can't oversee it. Right off the bat, the firm should gather information from its showcasing channel. The firm needs to break down information genuinely to have an upper hand in the area. Information Mining (DM), which is the procedure of naturally looking huge volumes of information to separate learning from them an out of a human-justifiable structure, causes examiners to perceive connections inside information. Applying DM strategies to promoting information is amazingly valuable to discover fascinating, already obscure, covered up designs, which would then be able to be better characterized, in monstrous datasets. In this way, DM finds vital learning from the showcasing channel [19]. The accomplished learning has a vital significance as far as rivalry and change of promoting and creation for the firm. This is on the grounds that learning accomplished can enhance the correspondence between the promoting channel and the firm by better controlling the procedures, and by knowing the insights about them.

Today, advertising channel choices are as essential as the choices organizations make about the highlights and costs of items [20]. In this examination, we connected a DM structure and we introduced a choice tree acceptance from showcasing channel information to enhance the effectiveness of the advertising framework. DM methods were actualized to showcasing overview information. We investigated the utilization of various pre-handling and DM methods including inconsistency discovery examination, Kohonen SOM, and C5.0 DTs. This examination included quality diminishment utilizing KNs. A C5.0 DT which was utilized for the grouping of the informational index with 10 tree profundities was produced. The exactness rate of the model was 92.67%. The DT display lays out the information grid obviously with the goal that all choices can be investigated. This obtained information might be utilized to foresee the future practices of the promoting channel firms. Information are prepared into a reasonable configuration, and choice tenets are likewise produced. The DT demonstration causes administrators to comprehend the advertising channel firms. This examination is likewise imperative to evaluate the future objections of the organizations and to design future advertising advancements. On the off chance that we assess the present promoting channel and plan for limit needs we will accomplish a superior promoting framework execution. Therefore, the information acquired will enhance the execution of the advertising framework. Elective DM systems utilizing computerized reasoning strategies can be examined in future research to analyze different methodologies and to actualize this system.

E.   Market Basket Analysis: Identify the changing trends of market data using association rule mining

Market Basket Analysis(MBA) otherwise called affiliation manage learning or partiality examination, is an information mining system that can be utilized as a part of different fields, for example, advertising, bioinformatics, training field, atomic science and so on[21].

The fundamental point of MBA in promoting is to give the data to the retailer to comprehend the buy conduct of the purchaser, which can help the retailer in redress basic leadership. There are different calculations are accessible for performing MBA. The current calculations take a shot at static information and they don't catch changes in information with time. Be that as it may, proposed calculation mine static information as well as gives another approach to consider changes occurring in information. This paper examines the information mining method i.e. affiliation manage mining and give another calculation which may accommodating to look at the client conduct and aids expanding the deals.

Today, the extensive measure of information is being kept up in the databases in different fields like retail showcases, managing an account division, medicinal field and so on. Be that as it may, it isn't essential that the entire data is valuable for the client. That is the reason, it is important to separate the valuable data from substantial measure of information. This procedure of separating valuable information is known as information mining or A Knowledge Discovery and Data (KDD) process [22].

The general procedure of finding and translating designs from information includes numerous means, for example, determination, preprocessing, change, information mining and interpretation. Information mining helps in the business for advertising. Market crate examination is otherwise called affiliation run mining. It encourages the advertising investigator to comprehend the conduct of clients e.g. which items are being purchased together. There are different systems and calculations that are accessible to perform information mining. There are many different techniques of data mining are as follow:

Classification Technique: In classification, first look at the highlights of recently exhibited protest and relegate it to a predefined class for instance group the acknowledge candidates as low, medium or high hazard.

Association technique: The fundamental objective is to build up the connection between things which exist in the market. The normal cases of affiliation demonstrating are Market bin Analysis and strategically pitching programs. The devices utilized for affiliation run mining are apriori calculation and weak toolbox.

Prediction technique: Expectation of some obscure or missing characteristics esteems in light of other Information. For instance: Forecast the deal an incentive for one week from now in light of accessible information.

Clustering Technique: In this, Data Mining composes information into important sub-gatherings (bunches) to such an extent that focuses inside the gathering are like each other, and as various as conceivable from the focuses in alternate gatherings. It is an unsupervised.

Outlier Analysis: In this, Data Mining is done to recognize and clarify special cases. For instance, if there should arise an occurrence of Market Basket Data Analysis, anomaly can be some exchange which happens curiously.

At introduce numerous information mining calculations have been created and connected on assortment of functional issues. However occasional mining is another approach in information mining which has picked up its importance nowadays. This field is developing because of necessities in various applications and constraints of information mining. This would improve the energy of existing information mining strategies.

Discovering the examples because of changes in information is in itself an intriguing region to be investigated. Author of this paper proposed that, a few territories are still there which should be centered around. Right off the bat, comes about have impacted extraordinarily by the manual edge esteems for score, so it is expected to computerize the limit esteems for better acknowledgment of anomalies. Furthermore, this approach is particularly focused at Market Basket Data, it might maybe be reached out to different zones.

F. Deep-level Quality Management Based on Big Data Analytics with Case Study

The Big information investigation gives new opportunities to the endeavors to upgrade their administration and assembling levels. An answer with contextual analysis is proposed to achieve deep level quality administration in view of enormous information examination. Initially, the execution of enormous information examination in view of mechanical process information is delineated with contextual investigation outline [23].

Through the examination and highlight extraction of disconnected information, the comparing reference demonstrate library is developed, which can be utilized for ongoing preparing of unlabeled information in the modern field. The group, anomaly and other information pointers we can get from the computation and investigation would have a specific directing essentialness for the undertaking's profound level quality administration.

At that point, the disseminated memory registering motor in light of Spark, furthermore, the execution of web application stage in light of Spring MVC system is additionally portrayed. This can get higher execution and unhindered capacities for information examination. In the meantime, human-information interface can be accomplished.

With the fast advancement of innovation and economy today, producing industry has turned into a critical connection amongst innovation and utilization, which is the essential assurance of changing over logical and mechanical development into beneficial powers and social esteems as quickly as time permits. In addition, it turns into the foundation of social economy and the principle body with the improvement of the circumstances. In the interim, with the rise of Industry 4.0 and Made in China 2025 idea, the fourth mechanical insurgency is on the rise.

Assembling is additionally confronting the difficulties of change what's more, change. Smart assembling has turned into a new pattern in the worldwide assembling industry. Particularly in China, it is earnest to accomplish the profound level reconciliation of data and savvy for some endeavors which confront the chance to grow generation and update hardware [24]. Besides, the utilization of enormous information innovation can upgrade the quality administration, process redesigning and other angles' capacity of administration and basic leadership.

In this paper, the attributes of mechanical information are broken down, and the general examination procedure of modern information is given right off the bat. At that point the execution technique for profound level quality administration is proposed with contextual investigations. An arrangement of answers for profound level quality administration and the stage developments are recommended in light of the enormous information investigation.

The reference demonstrates library is built by displaying the disconnected information and can be utilized for ongoing handling of online information streams. Meanwhile, to finish the investigation of a lot of information, elite registering bunch comprising of numerous PCs is built in light of the Start distributed computing structure, which gives speed ensure for the continuous huge measure of information [25]. With a specific end goal to influence the framework to have simpler human-PC interface, we assemble Web applications in light of the Spring MVC system for the cooperation amongst clients and the stage. The applications pass on the client's task directions to the figuring stage and demonstrate the outcomes friendlily on the page or in programmed created reports. The framework is completed and test running with Weichai Power and can be the case for customary assembling endeavors to fabricate profound level quality administration stage in view of enormous information examination.

III. CONCLUSION AND FUTURE WORK

There is a need for managers and top-level management employees to explore the hidden data and those relevant facts and data that will help to grow the business like online news, advertisement and marketing agency. The following objectives should be considered for working with Web Usage Mining:

1. Avoid sampling / aggregation;

2. Decrease movement, duplicity and replication of data;

3. Data must be closest possible to the analytical study.

4. Optimize computation speed using Big Data program.

5. Study of User Behaviour

6. Frequent item set

7. Location centric analysis

As mentioned above, this paper gives design and development of front end for data collection and preparation with collecting, sorting, storing, conversing the obtained data. The analytics work using big data approach is the future work.

REFERENCES

[1] Raorane, Abhijit, and R. V. Kulkarni. "Data mining techniques: A source for consumer behavior analysis." *arXiv preprint arXiv:1109.1202*, 2011.

[2] Padmanabhan, Balaji, and Alexander Tuzhilin. "On the use of optimization for data mining: Theoretical interactions and eCRM opportunities." *Management Science* 49.10 , 2003. Pp. 1327-1343.

[3] Victor, Haastrup Adeleye, et al. "Customer behaviour analytics and data mining." *American Journal of Computation, Communication and Control* 1.4 , 2014, pp. 66-74.

[4] Yeo, A. C., et al. "A mathematical programming approach to optimise insurance premium pricing within a data mining framework." *Journal of the Operational research Society* 53.11, 2002, pp. 1197-1203.

[5] Al-Radaideh, Qasem A., Emad M. Al-Shawakfa, and Mustafa I. Al-Najjar. "Mining student data using decision trees." *International Arab Conference on Information Technology (ACIT'2006), Yarmouk University, Jordan.* 2006.

[6] Geeta, R. B., Shashikumar G. Totad, and Prasad Reddy PVGD. "Amalgamation of web usage mining and web structure mining." *International Journal of Recent Trends in Engineering* 1.2, 2009, pp. 279.

[7] Shirgave, Suresh, and Prakash Kulkarni. "Semantically enriched web usage mining for predicting user future movements." *International Journal of Web & Semantic Technology* 4.4, 2013, pp. 59.

[8] Qi, Xiaoguang, and Brian D. Davison. "Web page classification: Features and algorithms." *ACM computing surveys (CSUR)* 41.2, 2009, pp. 12.

[9] Berson, Alex, Stephen Smith, and Kurt Thearling. *Building data mining applications for CRM*. New York: McGraw-Hill, 2000.

[10] Baglin, Carol Ann. *Linking eligibility for early intervention services to school age educational disability*. Diss. Johns Hopkins University, 2001.

[11] Zhou, Qing, and Lin Liu. "Understanding Requirements for Online Services Based on Users' Behavioural Data Analysis." *Computer Software and Applications Conference Workshops (COMPSACW), 2013 IEEE 37th Annual*. IEEE, 2013.

[12] Victor, Haastrup Adeleye, et al. "Customer behaviour analytics and data mining." *American Journal of Computation, Communication and Control* 1.4, 2014, pp. 66-74.

[13] Thearling, Kurt. "Data mining and customer relationships." *Building data mining applications for CRM. New York, NY: McGraw Hill. Available: http://www3. shore. net/~ kht/text/whexcerpt. htm (April 15, 2002)* 2000.

[14] "Developing a cost-benefit analysis of mining sites in Mongolia" , annual report -2011, Ulaanbaatar, 2012. .

[15] Chen, Yen-Liang, et al. "Market basket analysis in a multiple store environment." *Decision support systems* 40.2, 2005, pp. 339-354.

[16] Ragunathan, Thirumalaisamy, et al. "Advertisement posting based on consumer behaviour." *Procedia Computer Science*50, 2015, pp. 329-334.

[17] Vinerean, Simona, et al. "The effects of social media marketing on online consumer behavior." *International Journal of Business and Management* 8.14, 2013, pp. 66.

[18] Ozyirmidokuz, Esra Kahya, Kumru Uyar, and Mustafa Hakan Ozyirmidokuz. "A Data Mining Based Approach to a Firm's Marketing Channel." *Procedia Economics and Finance* 27, 2015, pp. 77-84.

[19] Dragomir, Otilia Elena, Florin Dragomir, and Marian Radulescu. "Matlab application of Kohonen self-organizing map to classify consumers' load profiles." *Procedia Computer Science* 31, 2014, pp. 474-479.

[20] Gómez-Pérez, Gabriel, et al. "Assigning discounts in a marketing campaign by using reinforcement learning and neural networks." *Expert Systems with Applications* 36.4, 2009, pp. 8022-8031.

[21] Kaur, Manpreet, and Shivani Kang. "Market Basket Analysis: Identify the changing trends of market data using association rule mining." *Procedia computer science* 85, 2016, pp. 78-85.

[22] He, Zengyou, et al. "FP-outlier: Frequent pattern based outlier detection." *Computer Science and Information Systems* 2.1, 2005, pp. 103-118.

[23] Khan, Mohiuddin Ali, Sateesh Kumar Pradhan, and M. A. Khaleel. "Outlier Detection for Business Intelligence using data mining techniques." *International Journal of Computer Applications* 106.2 , 2014.

[24] Zhao Quiankun, Bhowmick Sourav. Association Rule Mining: A Survey. Technical Report CAIS Nanyan Technological University, Singapore, 2003, pp. 1-20.

[25] Kaur Paramjit, Attwal Kanwalpreet S. Data Mining: Review.International Journal of Computer Science and Information Technologies, 2014, pp. 6225-6228.

Kavita Tandon has done her graduation from Institute of Technology, Guru Ghasidas Vishwavidyalaya (ITGGV), Bilaspur which is affiliated to Guru Ghasidas Vishwavidyalaya (GGV), Bilaspur, Chhattisgarh, India. She is currently pursuing her post-graduation from Bhilai Institute of Technology, Durg, in the Department of Computer Networks. Her current area of research is Data mining, Web mining, Hadoop, Sensor Networks, Network security and applications.

D. P. Mishra completed his graduation in Computer Science and Engineering from Amravati University, and M. tech in Computer Technology & Application from Chhattisgarh Swami Vivekanand Technical University. He is currently pursuing his Ph.D from Chhattisgarh Swami Vivekanand Technical University, Bhilai. He is experienced in designing, developing, implementing and testing computer-based hardware & software solutions. He is an expertise in dealing with Networking and Security solutions (Hardware & Software base) used in wired and wireless networking. He is proficient in design & deployment of wireless network. Currently, he is an associate professor at Bhilai Institute of Technology, Durg, in the Department of Computer Science & Engineering.