iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** VIII **Month of publication:** August 2019

DOI: http://doi.org/10.22214/ijraset.2019.8036

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VIII, Aug 2019- Available at www.ijraset.com

Analysis of Users Behaviour in Structured E-Commerce Websites using Web Usage Mining Techniques

Nikhat Fatima[1], V Mallesi[2]
[1]*M.Tech Scholar,* [2]*Associate Professor & Head, Department of Computer Science and Engineering, Bheema Institute of Technology & Science, Adoni. Kurnool Dist A.P, India*

Abstract: ***With the fast improvement of Internet in India, the industry's/company business model or plan of action has changed. At present, extraordinary procedure has been made in web based business stage for its benefit and exchange quick. Rivalry for clients is the key factor for online business in the inexorably furious challenge. If we can grasp the requirements of customer and targeted business activities is developed then we can provide trading mode in convenient way and a wide selection for customers on website and provides better retain of customers for e-commerce business. One of the key approaches to this issue is web data mining techniques. We can obtain the behaviour of user by viewing the browsing behaviour of customers on websites, analysis it and then finding out solution to it. With this sellers will identifies the needs of customers and gives personalized in accordance with preferences of customers and then competitive advantage can be taken. In this paper we propose to use methodology and tool to improve accuracy of interpretation of user's behaviours. Also we propose the usage of Temporal Logic and model checking techniques for structured e-commerce websites. We perform the analysis of the usage of e-commerce websites and to find out customers' complex behavioural patterns with the help of checking temporal logic formulas describing such behaviours against the log model.***

Keywords: E-commerce, behaviour, website, pattern, model checking techniques, user, linear temporal model, data mining technique

I. INTRODUCTION

The rapid growth of internet and smart phone has changed the life style of people and making purchasing and selling online more flexible and comfort. The online shopping is made with the use of e-commerce websites. The customer uses e-commerce website to brows endless product catalogues, to compare prices, to create wish list and to have continuous information of different products and offers. This growing electronic market (e-commerce website) is very competitive, making the option for a customer to easily move from one e-commerce to another when their requirements are not fulfilled [1], [2]. As a result, business analysts of e-commerce need to analyse and understand behaviour of customers when navigated through the e-commerce website and identification of reasons for product purchase or not. This behavioural information is useful to permit e-commerce websites to send a more adapted service to customers, allow ecommerce website to retain customers and increasing benefits. But determining behaviour of customer and the reasons for guiding buying process of customer is a extremely complex task [3][4][5]. With the help of e-commerce websites user can have verity of options for navigation and actions: movement of user through variety of category of product is easy and follows many paths for navigation for specific product visited or utilize different methods for buying products [7][8]. These activities of users will be recorded into the logs of web server [3], [8]. The sequences of web events (click streams) generated by every user are stored in a order way in web server logs. The most valuable behaviour of users is hided in web logs and it must be determined and analysed [9]. The accurate analysis gives the scope to improve the contents of website and website structure [10], to adapt and personalize the website contents [11], [12], [13], to products recommendation [14], [15], or to know the user interest on specific products [16], for example.

The web data mining techniques are very much useful for find outing patterns in log files. The main goal of these techniques is to find out user usage patterns that give the detail explanation of user's interests. The various data mining techniques example: classification techniques, clustering techniques association rules chronological patterns [18], [19] have been utilized in e-commerce field. In several application areas such techniques (data mining) are employed along with process mining techniques. Such approaches (techniques) are component of the domain called business intelligence and uses algorithms that are very specific to find out the patterns which are hidden and its relation in sets with large data [20]. The behaviour of any customer is possible in e-commerce website because e-commerce website is open system. Such flexibility has led to find out a model called process oriented

©IJRASET: All Rights are Reserved

indicating behaviour of customers a very complex task [21]. This is due to availability of various likely interactions to final process model can be over fitting spaghetti model or an under fitting flower model [20], from which no constructive analysis can be carried out. As a result, the data mining techniques have been selected for e-commerce websites analysis. However, at present there are certain constrains with the view point of analysis for data mining techniques. On the other side this techniques will not operate directly with sequences of events (click stream) that are generated while performing navigation of user's through the e-commerce website and with this an abstraction of this sequence. Such abstraction along with use of summarized data set describes the details of session of customer on websites. The summery of this data called number of web pages visited, frequent visit of customer for product category, and the time spent by customers on web page, for example. The many of mining techniques are capable to classify the abstractions and find out simple relation between certain interests of high level events.
Our contributions to this work are:

A. The analysis of methodology and tool to improve accuracy of interpretation of user's behaviours.
B. The use of Temporal Logic and model checking techniques in structured e-commerce websites.
C. Analysis of the usage of e-commerce websites and identified the complex behavioral pattern of customers with the help of checking temporal logic formulas illustrating this kind of behaviors against, the log model.

The rest of the paper is organized as follows. In Section 2, we describe related work of techniques to analyse e-commerce web-logs, in section 3 we present the proposed system, models and system architecture. In section 4 results and discussion and we conclude the paper in Section 5.

II. RELATED WORK

The related work to techniques employed to analyse e-commerce web logs is reviewed.
In paper [10] have done work on improvement of E-commerce website by employing web usage mining. They have analysed the customer usage of web browser, WebPages visited by customer and the time spent by customer on each webpage or usage of keywords in search engine. In paper [14, 16] the authors done carried out research on use of product taxonomy and web usage mining to e-commerce for collaborative recommendations. They studied the user's interest in various categories of product, visited categories and its frequency of visits. In [12] the authors have proposed the user modelling system based on web usage and content mining. The most frequent words present in web pages due to customer visits is computed by employing techniques called text mining to generates session characterization. They have identified user's interest by checking the contents of the web pages that has been visited by customers. In paper [26] have proposed soft clustering technique for customer segmentation of various category of data in e-commerce. This is carried out to construct the categorization of data by using questionnaires satisfied by customers and in [15] categorization is done with the help of purchasing made by customer's, customer demographic and customer personal data. In both the cases after computing the characterizations of customers a clustering algorithms are employed to determine sessions sets indicating similar behaviour and few common interests and this sessions sets is used for improving contents and structure of website[10]. Also sessions sets are used for content adaption and content personalize [12][27] and also for products recommendation. This sets also gives the understanding of buying process behaviour and interest in specific products by users/customers [16] [26], [28], [29]. In paper[30][31] authors has presented the concept of click stream prediction using click stream data by employing a mining technique called sequential stream with markov chains. They have computed the patterns called mining significant. With this techniques prediction of click stream is carried out. This sequential stream mining technique is employed for prediction of user's behaviour and to extract user's navigation sequences. This user's navigation sequences are used to create models called Markov chains (statistical and probabilistic) models. These models are utilized for prediction of user's next click. Creating of Markov chain models is very expensive and this kind of models responds to very short term reasoning. The prediction of statistical models can be improved by combine use of clustering algorithms and Markov chains [31]. The concept is to apply clustering algorithms for grouping user sessions first and clusters are obtained. After this a specific markov chain is generated for each cluster. In [21] the authors presents about process mining techniques for obtaining causal relations among the users sessions events. The proposed technique can also utilizes events of abstraction of very high level and for identification difficulties of infrequent behaviours pattern in the events of buying. In paper [22, 23, 24] presents about a declarative method can be adopted and a set of constraints is considered for anything does not violates and set can be acceptable. The constraints can be represented by employing temporal logic. The language called declare property description[32] is used to define a patterns set and this patterns sets which is related to usual workflow structures are employed to easily analyse the task of specifying properties. The causality relations between the events in session of user can be simply described by declares. In paper [33] it has been proposed MP-declare

©IJRASET: All Rights are Reserved

which defines the data and time constraints in patterns of declare with set of patterns which are predefined and checking of general formulas not made possible. In this case the addition of novel pattern may require the implementation and desired behaviour. In this case certain tool called model checker is used to check the properties such as described temporal logic against the web log. Presently, the available and widely used powerful tools for logs analysis of e-commerce websites and one of the main tools is Google analytics [34]. The Google analytics carries controlling the network traffic, information collection of user sessions such as visiting fist and last pages of websites and each pages time spent by users and displays the synthesis report of users behaviours. The data based on the traffic may be combined with information of users personal and geographic. Google analytics works on analysis of information collected with the use of page tagging technique. The analysis can be used to improve the design of website, advertising and marketing. The other tools are Clicky, Piwik, Adobe Analytics o W3Counter Web Analytical tool [35, 36, 37, 38]. Form the above discussed previous work we have analysed and identified the issue related to accuracy of interpretation of user's behaviours and so there is scope for improvement in user's behaviours interpretation. In our work we carry the improvement of accuracy in interpretation in behaviour of users by proposing novel methodology and tool.

III. PROPOSED SYSTEM

In this section we discuss about the details of proposed methodology and tool to have very accurate user's behaviour interpretation. We propose a mining technique to obtain causal relations between user trace events without using web pages tagging. This proposed mining technique provides easy global analysis of behaviour of user in e-commerce website. In this work we apply LTL (linear-temporal logic)-based model checking techniques to analyse e-commerce web logs and analysis can be enabled with the use of representation of event types commonly and attributes of web structure of e-commerce, categorization of product and possibilities of user navigation. After this the pre-process of web server logs is carried out and finally different queries are performed for logs analysis and for interpretation of logs. We propose a methodology for usage of temporal logic and model checking techniques in a structured e-commerce, web-sites. In this work we carry out the analysis of utilization of e-commerce websites and we find out complex behavioural patterns of customers with the use of formulas called checking temporal logic. This will describe the behaviours against the log-model. Initially the pre-processing of logs of web server is carried out for extracting the traces details. The trace in the sense user session events sequences. The events can be user actions or system actions done when a product is visited by client and also when product category page is visited when product is added to wish-list, when the usage of search engine is done etc. The business analyst will use a pre-defined set of temporal logic patterns for queries formulation and it will help to find out and understand the way of usage of website by client. The existence of relation of complex causality between events present in the sessions of client. The control over the way the applications of checking algorithms, performance results disappointment, we can have when few model checking tools at our disposal when used beside big models, this leads to development of specific model checking tool. We carried out this with the usage of the SPOT libraries for L.T.L model checking [25]. Modules of proposed system are Behavioural Module and Linear-Temporal Logic (LTL) model. The behavioural model describes about the knowledge of user behaviour permits e-commerce websites to deliver (convey) more personalized services to customers. Also it provides increase in benefits and customers retaining. The e-commerce websites usage is analysed in order find out complex behavioural patterns of customers with the help of formula called checking temporal logic describing complex behaviour against log mode. Linear temporal logic (LTL) it is also abbreviated as linear-time temporal logic is a kind of modal temporal logic (TL) with time as modalities. In this model the future of paths formulae is encoded. The future paths present in model are true conditions (eventually) and true condition based on until another becomes true. LTL model is a fragment of the more complex behaviour which permits branching time and quantifiers. Also LTL is a fragment of SIS monadic second order logic of one successor and with these characteristics it is also called as propositional temporal logic.

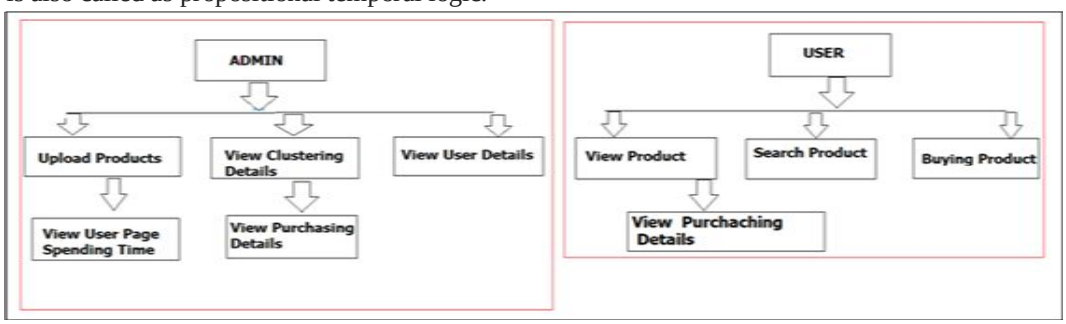


Figure 1 shows the architecture of proposed system

©IJRASET: All Rights are Reserved

The figure 1 illustrates the architectural diagram of proposed work. It consists of admin section and user section. The admin can perform different operation such as upload products, view clustering details and view user details. After upload products admin can further view the spending time of user on page and view clustering details further step is followed to view purchasing details of user on website. In user section the user will perform three different operations such as view product, product searching and product buying. After completion of viewing product by user further he/she can view purchasing details.

A. Advantages of Proposed System

1) The proposed method is applicable for much bigger logs directly.
2) It provides improvement in product categorization.
3) Method and Tool scale very well

IV. RESULTS & DISCUSSIONS

The results are shown in terms of snapshots and snapshots are self-explanatory. The snapshot makes the user understand easily the working operations in the proposed system. Below are snapshots of execution of designed proposed system. First open NETBEANS application and in projects tool right click on analysis of users and run the project. The following screen will display as shown in figure 2. This is admin section for performing different operations. First admin has to login using name and password which is generated after admin registration.

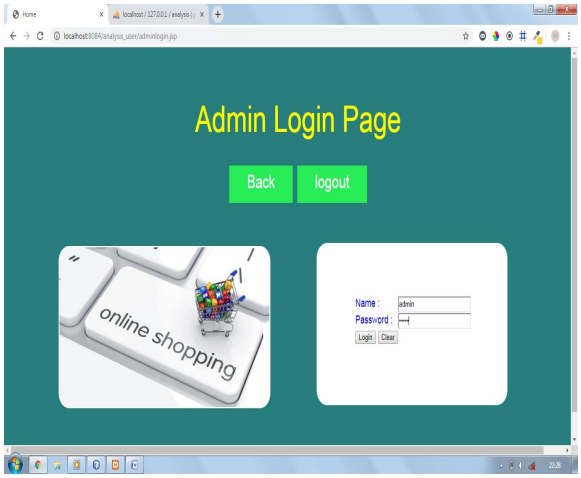


Figure 2: Admin Login with name and password

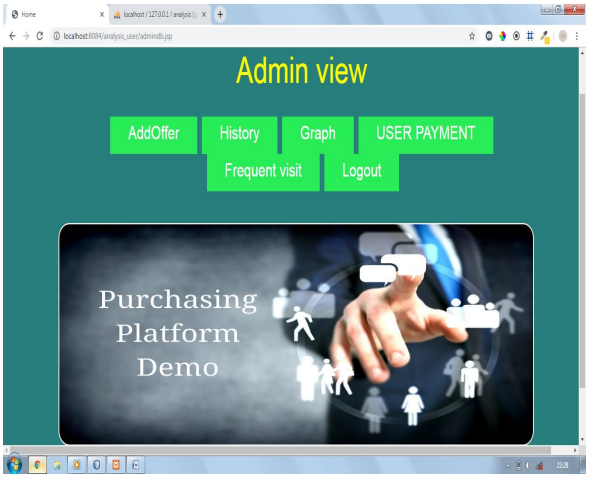


Figure 3: shows the Admin view details

Figure 3 illustrates the admin view section and in this admin can do various operations such as add offer, view history, graph, user payment made details, visitor frequency. In adder offer menu admin can add various offers for customers and in menu of use payment the admin can verify payment details made by customers. Admin can view the visitor frequency and analysis the user behaviors.

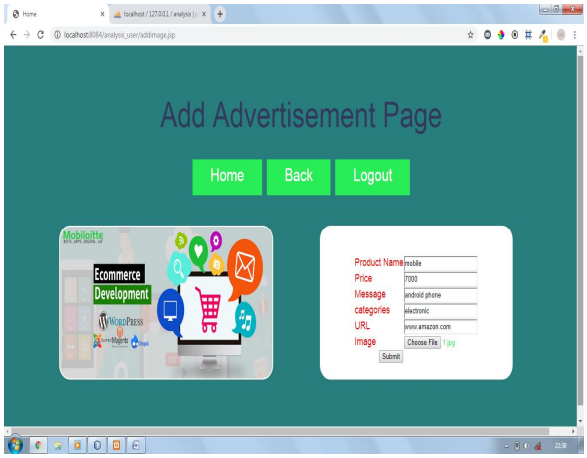


Figure 4: shows advertisement pages for adding advertisement

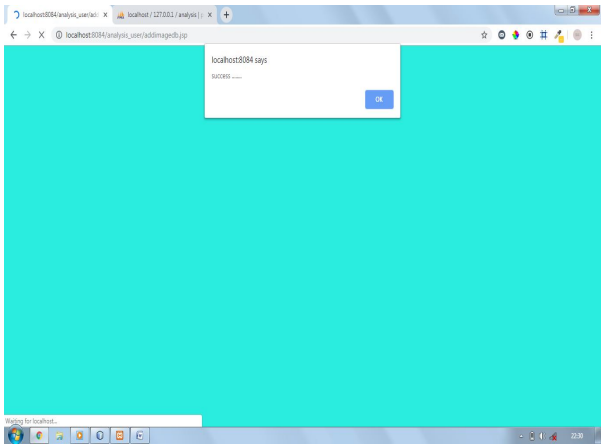
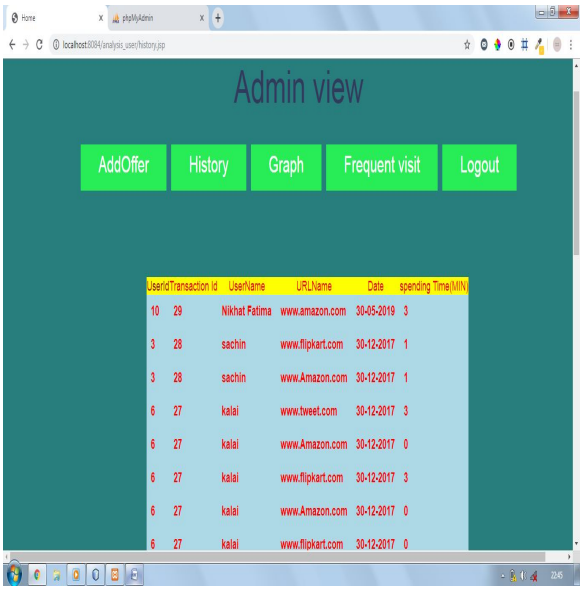


Figure 5: shows the status of addition of product

©IJRASET: All Rights are Reserved



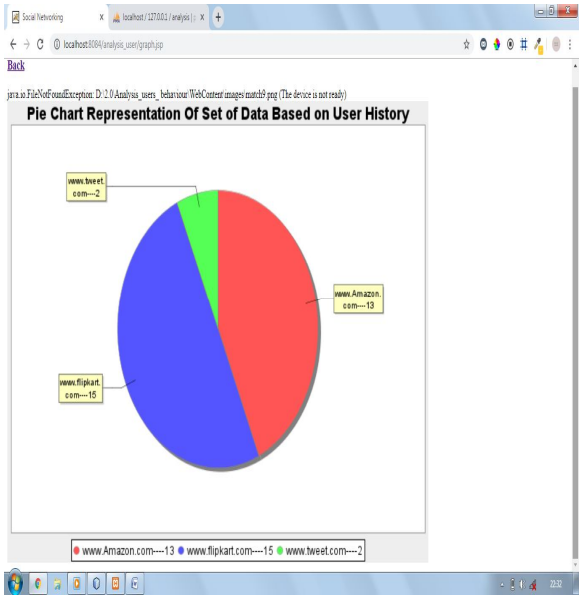


Figure 6: shows the details of user visited on different websites Figure 7: shows the pie chart illustrates set of data based on user history

Figure 4 illustrates the e-commerce webpage and in the admin can add advertisement. The details required for posting the advertisement are name of product, price, messages and product categories and advertisement images. Figure 5 illustrates the webpage showing the status of product successfully added to the website and a local host display the message of success. Figure 6 shows the web page containing the history of users purchasing product from different sites. The history contains user ID, transaction ID, username, URL name, date and time spent. Figure 7 indicates the set of data based on user history in terms of pie chart. This web page displays the user history graph for different websites such as Amazon, flipchart and tweet etc.

Figure 8: shows the users payment details such as amount, payment status and address for different website.

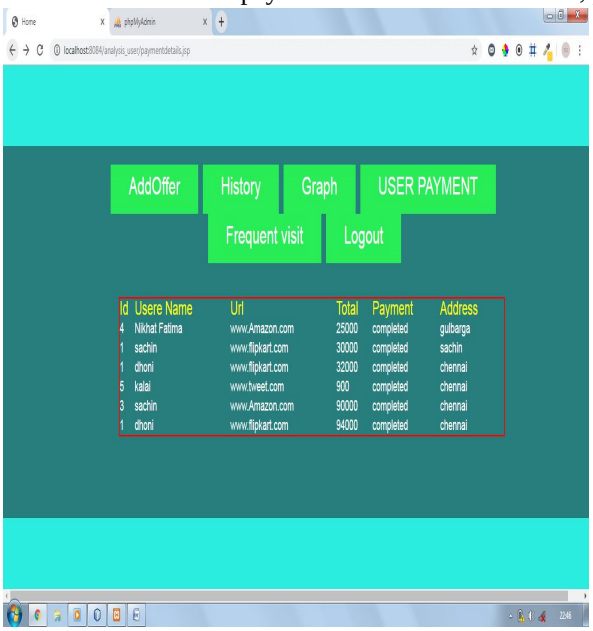


Figure 8: shows the users payment details for different website

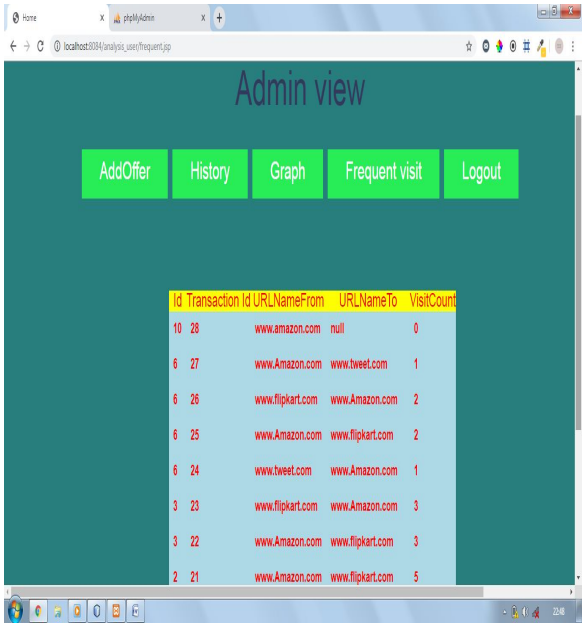


Figure 9: shows the frequency of user visited to websites

Figure 9: shows the frequent visit of user on e-commerce websites with details such as ID, transaction ID, URL name from, URL name to and visitor count. After viewing all the websites from figure 3 to 9 the admin can able to perform analysis of behaviors of user on e-commerce website (web logs).

©IJRASET: All Rights are Reserved

V. CONCLUSION AND FUTURE WORK

In this paper, we have successfully carried out the study and analysis of methodology and tools for improvement of interpretation accuracy of user behaviors. Also the usage of temporal logic and model checking techniques is analyzed for structured e-commerce web sites. Finally we analyzed the usage of e-commerce websites by customers and with the use of checking temporal logic formulas the complex behavioral pattern of customers is identified with the help of behaviors against, the log model. The proposed system is successfully implemented using Netbeans and Mysql and tomcat tools.

In future, we will propose an analysis tool with graphical interface for the input of properties to be analyzed and the output of results. Also we want to extend this work for further analysis of more behavioral patterns and to have automatic discovery.

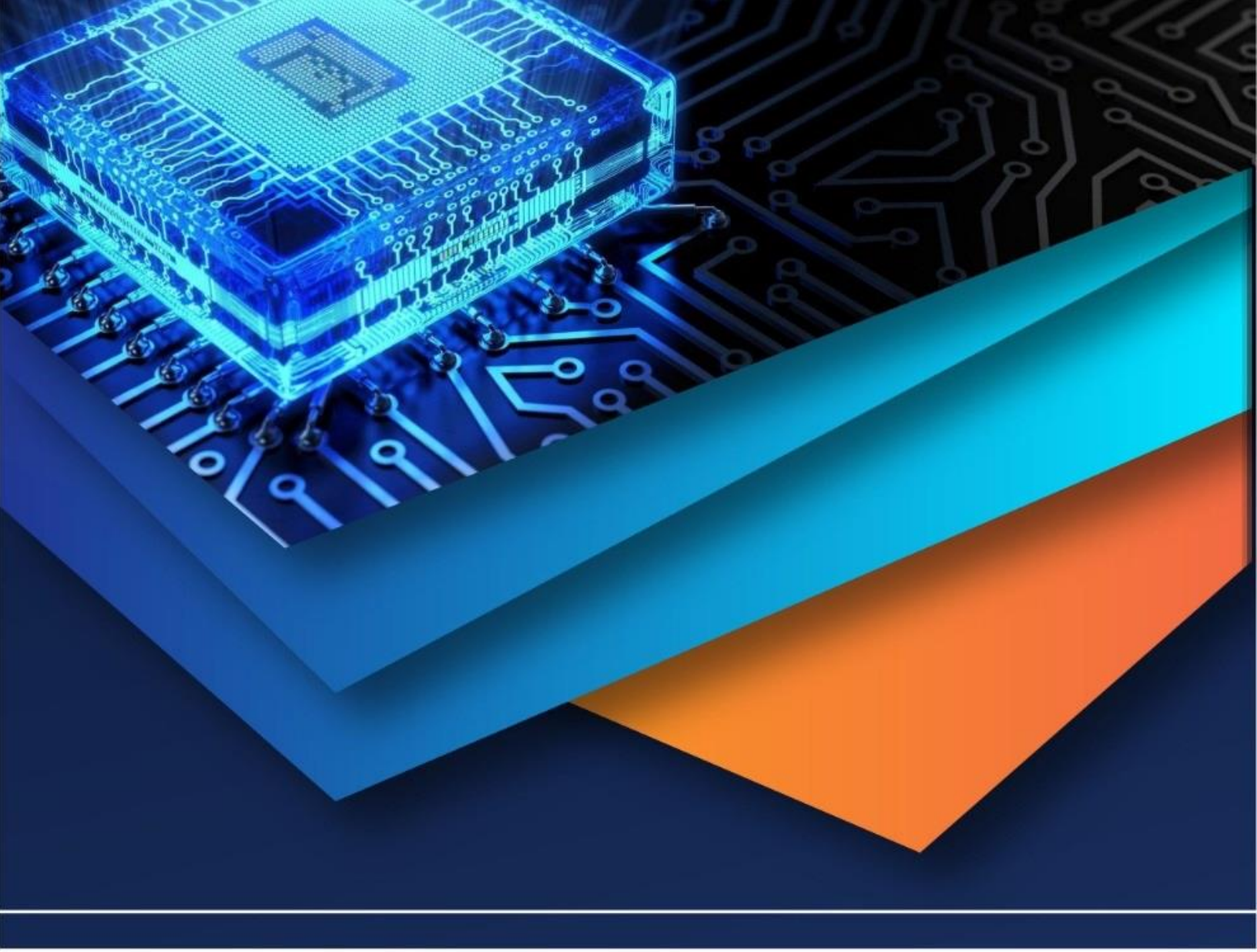
REFERENCES

[1] J. B. Schafer, J. A. Konstan, and J. Riedl, “E-commerce recommendation applications.” Hingham, MA, USA: Kluwer Academic Publishers, Jan. 2001, vol. 5, no. 1-2, pp. 115–153
[2] N. Poggi, D. Carrera, R. Gavalda, J. Torres, E. Ayguad´e, “Characterization of workload and resource consumption for an online travel and booking site,” in Workload Characterization (IISWC), IEEE International Symposium on. 2010, pp. 1–10.
[3] R. Kohavi, “Mining e-commerce data: the good, the bad, and the ugly,” in Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2001, pp. 8–13.
[4] W. W. Moe and P. S. Fader, “Dynamic conversion behavior at ecommerce sites,” Management Science, vol. 50, no. 3, pp. 326–335, 2004.
[5] G. Liu, T. T. Nguyen, G. Zhao, W. Zha, J. Yang, J. Cao, M. Wu, P. Zhao, and W. Chen, “Repeat buyer prediction for e-commerce,” in Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD ’16. New York, NY, USA: ACM, 2016, pp. 155–164.
[6] J. D. Xu, “Retaining customers by utilizing technology-facilitated chat: Mitigating website anxiety and task complexity,” Information & Management, vol. 53, no. 5, pp. 554 – 569, 2016.
[7] Y. S. Kim and B.-J. Yum, “Recommender system based on click stream data using association rule mining,” Expert Systems with Applications, vol. 38, no. 10, pp. 13 320–13 327, 2011.
[8] R. Kosala and H. Blockeel, “Web mining research: A survey,” SIGKDD Explor. Newsl., vol. 2, no. 1, pp. 1–15, Jun. 2000.
[9] F. M. Facca and P. L. Lanzi, “Mining interesting knowledge from weblogs: a survey,” Data & Knowledge Engineering, vol. 53, no. 3, pp. 225–241, 2005.
[10] C. J. Carmona, S. Ram´ırez-Gallego, F. Torres, E. Bernal, M. J. del Jes´us, S. Garc´ıa, “Web usage mining to improve the design of an ecommerce website: Orolivesur. com,” Expert Systems with Applications, vol. 39, no.12, pp. 11 243–11 249, 2012.
[11] Q Song M. Shepperd,Mining web browsing patterns for ecommerce, Computers Industry, vol.57, no.7, pp.622–630, 2006.
[12] O. Arbelaitz, I. Gurrutxaga, A. Lojo, J. Muguerza, J. M. Prez I. Perona, “Web usage content mining to extract knowledge for modelling the users of the bidasoa turismo website and to adapt it.” Expert Syst. Appl., vol. 40, no. 18, pp. 7478–7491, 2013.
[13] J. K. Gerrikagoitia, I. Castander, F. Reb´on, and A. Alzua-Sorzabal, “New trends of intelligent e-marketing based on web mining for e-shops,” Procedia-Social and Behavioral Sciences, vol. 175, pp. 75–83, 2015.
[14] Y. H. Cho and J. K. Kim, “Application of web usage mining and product taxonomy to collaborative recommendations in e-commerce,” Expert Systems with Applications, vol. 26, no. 2, pp. 233 – 246, 2004.
[15] K.-J. Kim and H. Ahn, “A recommender system using fGAg kmeans clustering in an online shopping market,” Expert Systems with Applications, vol. 34, no. 2, pp. 1200 – 1209, 2008.
[16] Q. Su and L. Chen, “A method for discovering clusters of e-commerce interest patterns using click-stream data,” Electronic Commerce Research and Applications, vol. 14, no. 1, pp. 1 – 13, 2015.
[17] J. Srivastava, R. Cooley, M. Deshpande, and P.-N. Tan, “Web usagemining: Discovery and applications of usage patterns from web data,” SIGKDD Explor. Newsl., vol. 1, no. 2, pp. 12–23, Jan. 2000.
[18] Q. Zhang and R. S. Segall, “Web mining: a survey of current research, techniques, and software,” International Journal of Information Technology & Decision Making, vol. 7, no. 04, pp. 683–720, 2008.
[19] B. Singh and H. K. Singh, “Web data mining research: a survey,” in Computational Intelligence and Computing Research (ICCIC), 2010 IEEE International Conference on. IEEE, 2010, pp. 1–10.
[20] W. M. P. van der Aalst, Process Mining: Discovery, Conformance and Enhancement of Business Processes, 1st ed. Springer Publishing Company, Incorporated, 2011.
[21] N. Poggi, V. Muthusamy, D. Carrera, and R. Khalaf, “Business process mining from e-commerce web logs,” 11th International Conference on Business Process Management, ser. BPM’13. Berlin, Heidelberg: Springer-Verlag, 2013, pp. 65–80.
[22] F. M. Maggi, R. P. J. C. Bose, and W. M. P. van der Aalst, Efficient Discovery of Understandable Declarative Process Models from Event Logs. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 270– 285.
[23] M. R¨aim, C. Di Ciccio, F. M. Maggi, M. Mecella, and J. Mendling, “Log-based understanding of business processes through temporal logic query checking,” in On the Move to Meaningful Internet Systems: OTM 2014 Conferences: Confederated International Conferences: CoopIS, and ODBASE 2014, Amantea, Italy, October 27-31, 2014, Proceedings, R. Meersman, H. Panetto, T. Dillon, M. Missikoff, L. Liu, O. Pastor, A. Cuzzocrea, and T. Sellis, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 75–92.
[24] A. Burattin, M. Cimitile, F. M. Maggi, and A. Sperduti, “Online discovery of declarative process models from event streams,” IEEE Transactions on Services Computing, vol. 8, no. 6, pp. 833–846, 2015.
[25] A. Duret-Lutz, A. Lewkowicz, A. Fauchille, T. Michaud, E. Renault, and L. Xu, “Spot 2.0—a framework for LTL and !-automata manipulation,” in Proceedings of the 14th International Symposium on Automated Technology for Verification and Analysis (ATVA’16), ser. Lecture Notes in Computer Science, vol. 9938. Springer, Oct. 2016, pp. 122–129.

©IJRASET: All Rights are Reserved

[26] R.-S. Wu and P.-H. Chou, "Customer segmentation of multiple category data in e-commerce using a soft-clustering approach," Electronic Commerce Research and Applications, vol. 10, no. 3, pp. 331–341, May 2011.
[27] L. G. Vasconcelos, R. D. C. Santos, and L. A. Baldochi, "Exploiting client logs to support the construction of adaptive e-commerce applications," in 2016 IEEE 13th International Conference on e-Business Engineering (ICEBE), 2016, pp. 164–169.
[28] Y.-L. Chen, M.-H. Kuo, S.-Y. Wu, and K. Tang, "Discovering recency, frequency, and monetary (rfm) sequential patterns from customers' purchasing data," Electronic Commerce Research and Applications, vol. 8, no. 5, pp. 241–251, Oct. 2009.
[29] S. Kim, J. Yeo, E. Koh, and N. Lipka, "Purchase influence mining: Identifying top-k items attracting purchase of target item," in Proceedings of the 25th International Conference Companion on World Wide Web, ser. WWW '16 Companion. International World Wide Web Conferences Steering Committee, 2016, pp. 57–58.
[30] S. D. Bernhard, C. K. Leung, V. J. Reimer, and J. Westlake, "Clickstream prediction using sequential stream mining techniques with markov chains," in Proceedings of the 20th International Database Engineering & Applications Symposium, ser. IDEAS '16. New York, NY, USA: ACM, 2016, pp. 24–33.
[31] L. Lu, M. Dunham, and Y. Meng, "Mining significant usage patterns from clickstream data," in Proceedings of the 7th International Conference on Knowledge Discovery on the Web: Advances in Web Mining and Web Usage Analysis, ser. WebKDD'05. Berlin, Heidelberg: Springer- Verlag, 2006, pp. 1–17.
[32] W. M. van Der Aalst, M. Pesic, and H. Schonenberg, "Declarative workflows: Balancing between flexibility and support," Computer Science- Research and Development, vol. 23, no. 2, pp. 99–113, 2009.
[33] A. Burattin, F. M. Maggi, and A. Sperduti, "Conformance checking based on multi-perspective declarative process models," Expert Systems with Applications, vol. 65, pp. 194 – 211, 2016.
[34] (2017) Google analytics. Accessed 22nd May 2017. [Online]. Available: https://analytics.google.com/analytics/web/
[35] (2017) Clicky. Accessed 22nd May 2017. [Online]. Available: https://clicky.com
[36] (2017) Piwik open-source analytics platform. Accessed 22nd May 2017. [Online]. Available: https://piwik.org
[37] (2017) Adobe analytics. Accessed 22nd May 2017. [Online]. Available: https://analytics.google.com/analytics/web
[38] (2017) W3counter. Accessed 22nd May 2017. [Online]. Available: https://www.w3counter.com

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)