

Identification of User Behaviour by Web Usage Mining

Bommi Harika¹, Dr. T Sudha^{2*}

¹Research Scholar, Department of Computer Science, Sri Padmavati Mahila Viswavidyalam, Tirupati *E-mail: harika.bommy@gmail.com

^{2*}Professor, Research supervisor, Department Of computer science, Sri Padmavati Mahila Viswavidyalam, Tirupati

*E-mail: thatimakula_sudha@yahoo.com

Article Info

Page Number: 678-692

Publication Issue:

Vol. 71 No. 4 (2022)

Article History

Article Received: 25 March 2022

Revised: 30 April 2022

Accepted: 15 June 2022

Publication: 19 August 2022

Abstract

In the current scenario, information is multiplied everyday due to the technological development in World Wide Web and Internet. Millions of customers are daily accessing the Internet to search the information and fulfill their requirements. Information was overload in the web servers and receiving the right idea was a difficult task of the customers. So, the research in web mining is focused and key research area to facilitate the information searching process. Web mining has special place as it helps the organizations and website analyst to make clever decisions about their customers. Web usage mining collects the website visitor's browsing details and recorded in the kind of plain text files in web servers. This information helps to identify the website visitor's navigation behaviors. Web recommendation, one such methods of web personalization that endorses the web pages to its website visitors depends on its earlier browse history. Web recommendation schemes aid the website companies for better navigation of web pages, rapidly attain their terminus and to attain pertinent idea. In this research work an efficient web page recommendation technique is developed by using web usage mining and pattern mining algorithms to improve the accuracy of the existing web page recommendation techniques. The aim is to categorize user behavior in recognizing the designs of the browser and navigation data of web users as well as to evaluate the performance of the Frequent Pattern (FP) Growth algorithm, Apriori algorithm and modified web log searching algorithm. The association rule mining method was executed to its clustered web logs to identify the regularly visiting web pages through the users. Realtime datasets are evaluated. The proposed recommendation technique is compared against the existing recommendation techniques to prove that that the proposed recommendation technique improves the accuracy of the web page recommendation.

Keywords: Algorithm, Apriori, Mining, Sparsity, Webpage

INTRODUCTION

Web Usage Process used to extract usable as the use of data mining methods to huge Web data sources in order to obtain user behavior, such as the visitor behavior described in [1.] Pattern detection and analysis are used to profile consumers and individual preferences as a next step. A crucial role is played by statistical methods in this regard. It is certain that acceptable traits and primary features describing a typology of users can be identified,

resulting in Web personalization [2, 3]. Classifier, association rules, and grouping studies are examples of statistical procedures.

Mining of Web is used to determine interesting use designs from Web logs which help in understanding and serving the different needs of the Web users. A web log captures/gathers the information of web users by their browsing behavior at a Website [4, 22]. The rapid advancement of web technology has resulted in an overabundance of information. Online users face a significant barrier in locating relevant and refined information in this overburdened environment. Recommendation systems are critical in providing such information to users [23-28]. Recommendation systems are used in a variety of web services, including online news, e-commerce, online music, Netflix, YouTube, Facebook, and scientific research, to help users make better decisions. The difficulties that arise in recommendation systems are addressed by the content-based filtering approach as well as the collaborative filtering technique [5, 6]. To get better advice, some researchers have employed these methods to improve a hybrid filtering system. The content-based strategy is based solely on the resemblance of the objects' features, and thus necessitates thorough information about just the things. It scores the items based on the similarities and recommends the top N to the users. Just the star rating, which would be the user preference value, is used in the information retrieval approach. An automatic, potential analyses offered by data mining tools and technologies [29-35]. Data mining is collection of demonstration tools representative for decision support system. Data mining tools and techniques are able to answer any complex, ad-hoc business queries that usually time consuming to resolve [7]. To know the previously unknown facts and hidden patterns used for market forecasting analysis, trend analysis, medical trial, science exploration, detection of fraud, etc. Data mining methods and tools can be analyzed as a result of the accepted progress of information technology [21, 40].

I. LITERATURE SURVEY

RapidMiner was created and deployed in 2001 with Ralf Klinkenberg, Ingo Mierswa, and Simon Fischer at the Dortmund University of Technology's Artificial Intelligence Unit [8]. RapidMiner is a software that is very well for machine learning, data mining, prescriptive modelling, and business analytics, and is also suggested by plenty of researchers. It is utilized for future studies, analytic education and training, web development, and industrial uses [9]. KDnuggets, a well-known forum on Data Mining and Artificial Intelligence, conducted a poll on latest events in data mining, and RapidMiner, a data mining newspaper, rated second in data analytic tools utilized for existing projects in 2009 and the first in 2010. It is published and made accessible underneath the AGPL open source license, and has been managed by Source Forge since 2004 for the past ten years. RapidMiner may be used to construct analytical processes and analyze data for web-, text-, multimedia-, feature engineering, data stream-, ensemble methodology, and networked data mining. It could be used to analyze data produced by implanted devices with higher throughput. Other plugins can be used to enhance the capability of RapidMiner [10, 11]. RapidMiner is open-source and fully available as freeware underneath the GNU AGPL as a Community Edition. RapidMiner has a user interface that is well adapted for designing an investigative-pipelines. The engine can also be

employed as an API or accessed from other programmers. The command prompt can be used to call certain functions. The GUI generates an XML file that specifies the data analysis procedures that the investigator or user wishes to utilize. There is also a commercial Enterprise Edition available for incorporation with closed codebases. RapidMiner 5beta is the version we use. For forecasting the user navigation behavior, this study effort provides two methodologies: web log search engine and improved web log searching method.

[12] shown transactions based on web as innovations and also examine its factors. Accuracy of Second order Markov model is high in comparison with First Order Markov study. In the same way as higher order model is built the accuracy level will be increased. [13] describe about the comparison analysis of different pattern recognition techniques, filtering, IP address to domain name and by recognizing the bandwidth/Hit comparison for image files to improve the website performance by structure content delivery and presentation. There are various Tools like OutWit Hub which can be used to search, fetch and organize data gathered from the Web. [14] describes about the hierarchical clustering for large databases which is an efficient approach for generating access patterns from the given Web logs. A first order Markov Model is used when a user request for a web page and in it Clustering techniques are also applied. For reducing the user perceived latency Dependency Graph (DG) is used, where graph consist of most access patterns is implemented. It also provides some drawbacks when pre-fetching policy is not designed cautiously then excessive network traffic is occurred. The long-term pre-fetching describes about the pattern statistics to identify valuable objects by using global access pattern statistics and it has application in mobile computing environments, Content Distribution Network (CDN) etc. Various advantages of web pre-fetching are described in [7, 15], whereas research in web caching grow immensely. [16] proposed two mining algorithms first clustering it through k-means and prefetching the patterns by Apriori algorithms of web proxy server. [17] have described about web usage mining where preprocessing of log data is done by data cleaner, user and session documentation. Data clean is a process where unwanted data is removed like .gif, .jpeg, .java script, .css, .txt file extensions. In session identification process the combination of number of pages accessed by the user is defined and in user identification process IP address, agent and OS act as single user. [18] describes about the preprocessing, data cleaning, session and user identification process. In data cleaning process irrelevant data is removed like .jpeg, txt, .css file extension and also with failed HTTP status. They also present the rules for identification of user and session, with path completion approach for missing page references.[19] explain two approaches for web log preprocessing through XML & text file. In XML file preprocessing is done which gives better structures to understand the web log file. In text file they describe about the attribute, record which are separated by a delimiter space for better understanding of web log files. [20] describes about extraction of data field and data storage where data field extraction is done through extraction algorithm implemented in java platform, after that data is stored through log table created by SQL query.

The suggested technique uses data sources like 1. User profile, 2. Item profile, 3. Tags, and 4. Rating, to identify user preferences. It investigates the user and item matrix's underlying pattern, which combines user behavior of desired characteristics. The data sparsity method is

limited by using the miner program [21]. The proposed approach consists of three steps. The missing data is deleted in the first stage of pre-processing, and the items are aggregated to lower the item space's dimensionality. In this phase, order to determine the number of cluster sizes is a difficult process. The second phase is uncovering underlying patterns in user preference behavior or a tangible-things. The apriori and FP Growth methods identify the frequent item set from the transactional database, which is necessary for generating frequent items which are closely associated to one another. The associative classifier is created with the help of the training data in the final phase, which is the suggestion. The proposed method is capable of analyzing the most often recurring things, revealing underlying link, association, and pattern of behind selected items. Rather than proposing popular goods, the proposed system analyses the user's hidden interests and recommends items based on past interest patterns. Because users' interests change over time, the suggested method analyses and ranks users' interests based on recent preferences and does not promote things that are no longer in use.

II. METHODOLOGY

In online usage mining, connection criteria are utilized to discover the connection among user sessions that occur together regularly [12]. Each association rule discussed here is based on the relationship between the Support and Confidence counts. The amount of transactions that also include that ordered pair equals the Support of that item set. The web usage mining begins when mining variables like support and confidence are configured. The number of databases scans necessary for the development of item sets, the data preprocessing technique employed, and how sensitive the method would be to changes in user settings, i.e. support, have all been examined.

Although the Apriori approach uses an effective process for trimming candidate item sets, it takes a long time to compute and requires numerous dataset scans to produce contender item sets. In terms of developing common patterns, the FP-growth method only requires two database scans. To describe the whole database, this technology utilizes a tight tree-structure. It doesn't necessitate candidate creation, which cuts down on calculation time. Web Log searching method mines log data obtained from user engagement over an amount of time. Figure 1 depicts the recommended methodology. To determine user behavior, real-time datasets from Sci-Hub, PUBG, Amazon, and Bitcoin are analyzed. Nonetheless, the suggested web log algorithms have a higher overall performance and memory consumption for various data architectures. When compared with conventional techniques, the proposed Modified Web log search technique requires less effort.

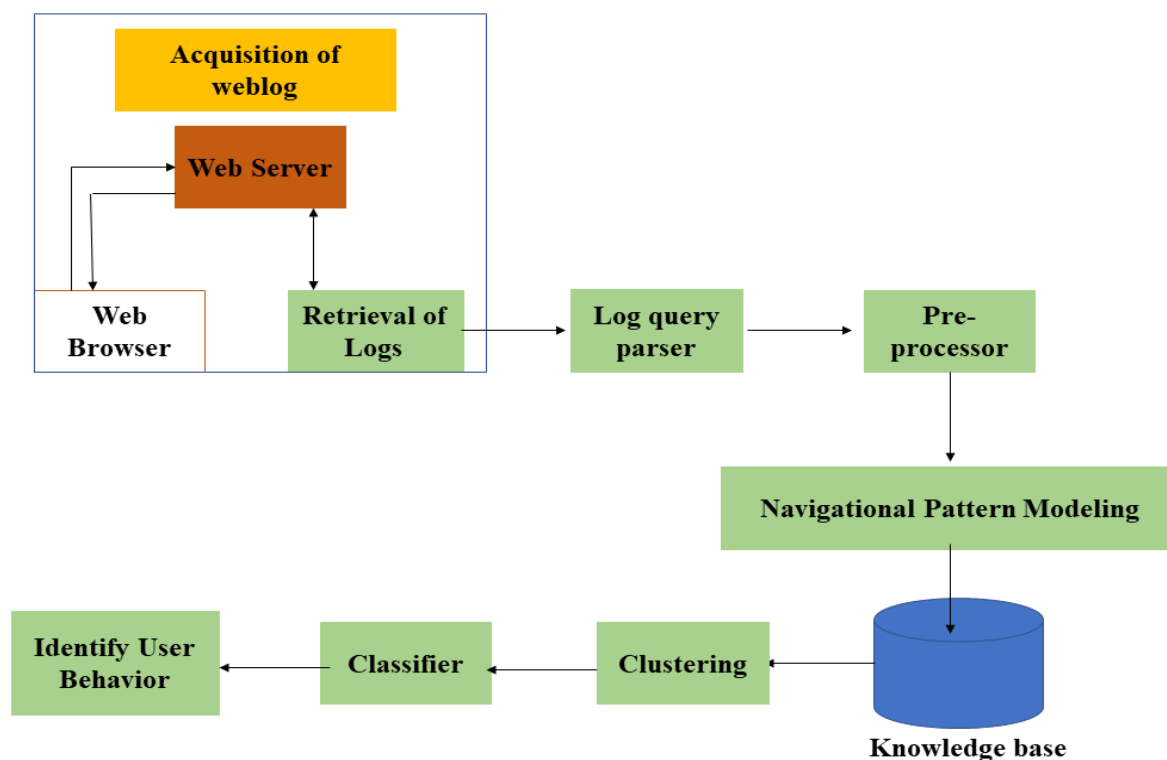


Figure 1. Proposed methodology

III. RESULTS AND DISCUSSION

Following were real life datasets that selected were shown in Table 1. The real time datasets of Sci-Hub, PUBG, amazon and bitcoin are evaluated to find the user behavior.

Table 1. Datasets

Dataset	Website	Number of logins	Number of Items	Average number of transactions per item
Sci-Hub	search.dataone.org	49208	96	17
PUBG game logs	figshare.com	126234	109	12
Amazon e-commerce	data.world	112208	105	9
Bitcoin Blockchain Historical Data	www.kaggle.com	251159	110	13

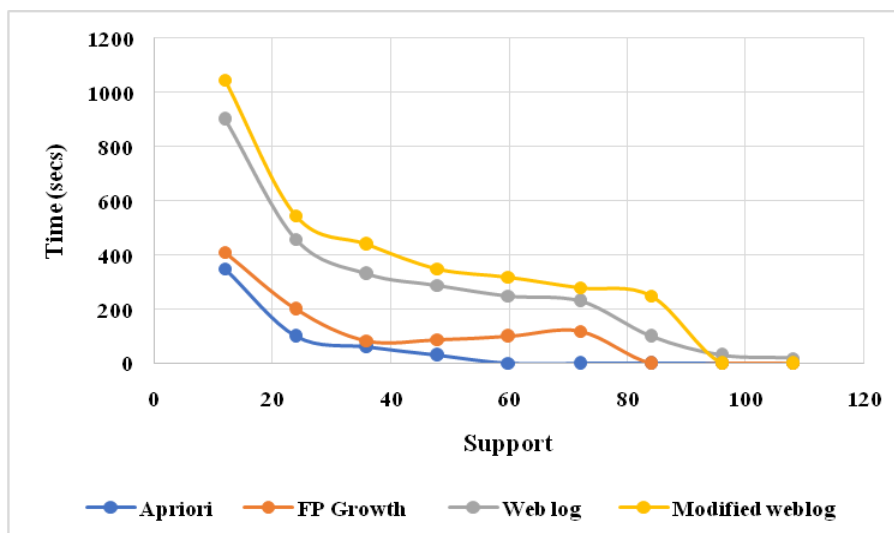


Figure 2: Analysis of algorithms for Sci-Hub dataset

Table 2: Comparing Sci-Hub datasets

Support in per cent	Apriori Algorithm	FP-Growth Algorithm	Web Searching Algorithm	Log	Modified Web Log Searching Algorithm
10	562.352	13.66	1012.53		1045.32
20	32.141	8.12	538.99		544.87
30	6.510	4.91	239.22		243.11
40	2.491	3.42	66.47		68.54
50	1.80	2.60	34.365		38.633
60	1.210	1.5	19.701		19.78
70	0.77	0.89	8.477		8.961
80	0.599	0.53	4.125		4.99
90	0.451	0.4	1.911		1.96
100	0.15	0.31	1.157		1.727

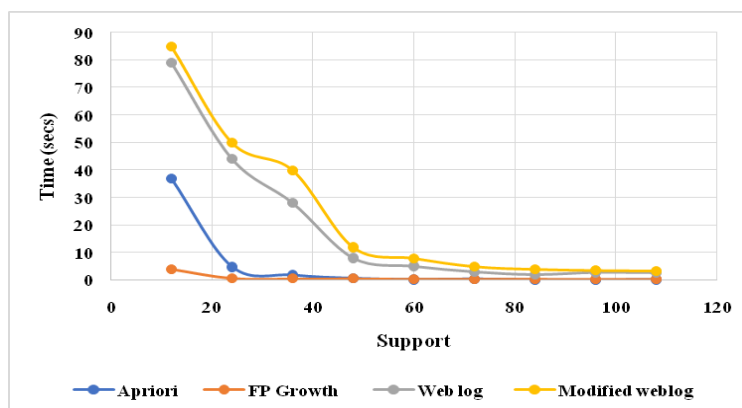


Figure 3: Analysis of algorithms for PUBG game logs dataset

Table 3: Comparing PUBG game logs datasets

Support in per cent	Apriori Algorithm	FP-Growth Algorithm	Web Log Searching Algorithm	Modified Web Log Searching Algorithm
10	37.621	3.9	79.717	82.476
20	4.512	0.83	44.83	48.53
30	1.933	0.60	28.083	37.26
40	0.837	0.64	8.65	9.86
50	0.485	0.51	5.342	6.675
60	0.6	0.50	3.987	4.564
70	0.342	0.49	2.965	3.94
80	0.341	0.48	2.866	3.68
90	0.341	0.46	2.749	3.43
100	0.338	0.44	2.744	3.42

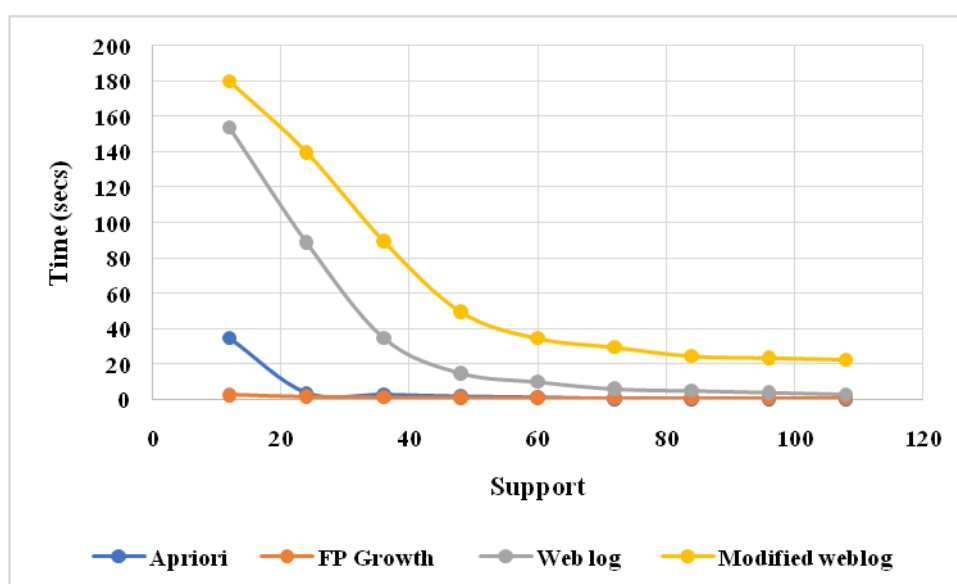


Figure 4: Analysis of algorithms for Amazon e-commerce dataset

Table 4: Comparing Amazon e-commerce datasets

Support in per cent	Apriori Algorithm	FP-Growth Algorithm	Web Log Searching Algorithm	Modified Web Log Searching Algorithm
10	35.43	2.88	154.57	180.6
20	4.33	1.76	89.54	137.74
30	2.56	1.68	35.65	88.98
40	1.48	1.57	15.66	47.51
50	0.75	1.42	9.521	30.63
60	0.606	1.38	5.322	28.76

70	0.599	1.37	3.587	25.36
80	0.462	1.29	3.476	24.43
90	0.459	1.27	3.451	24.30
100	0.458	1.25	3.398	23.95

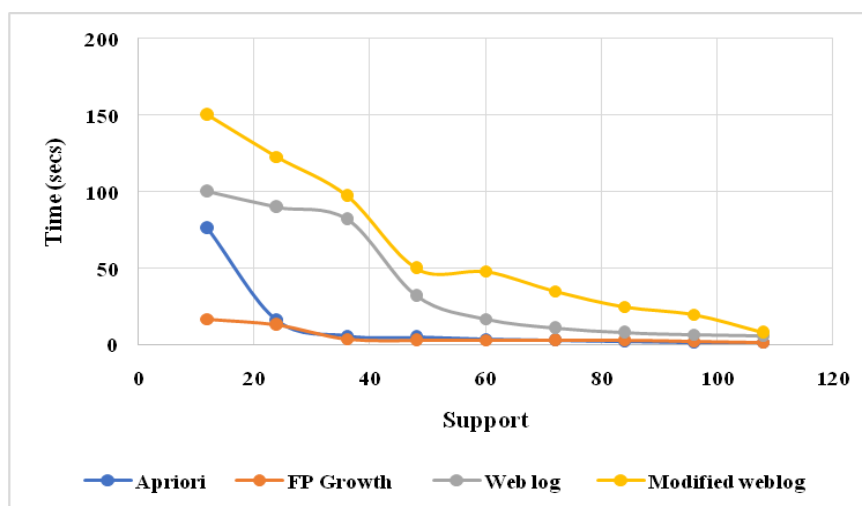


Figure 5: Analysis of algorithms for Bitcoin log dataset

Table 5: Comparing Bitcoin log datasets

Support in per cent	Apriori Algorithm	FP-Growth Algorithm	Web Log Searching Algorithm	Modified Web Log Searching Algorithm
10	76.521	17.71	190.20	199.31
20	15.919	13.65	182.76	192.00
30	5.754	3.32	82.73	97.10
40	4.87	3.29	32.16	45.84
50	3.666	3.11	17.21	18.35
60	2.98	2.97	11.74	17.60
70	2.57	2.94	8.62	12.75
80	1.95	2.73	6.98	11.10
90	1.89	2.42	6.03	8.06
100	1.84	1.98	5.99	7.95

The Analysis of figure 2, 3, 4 and figure 5 shows that modified weblog searching algorithm was better in all the four methods for research datasets [8, 36, 37]. Table 2, 3, 4 and 5 depicts the values at various support. When the and throughout is raised, the execution efficiency increased. For better amounts of support, the efficiency of the FP Growth and Apriori algorithms is nearly identical. For same datasets, FP-growth and Apriori take longer than other techniques. Since frequent pattern sets (patterns) are valuable for applying numerous data mining methods including such classification, clustering, or association rule, the project

plan is broad. Pattern mining methods can be used to create frequent itemsets in a variety of ways [38,39]. We have the opportunities to learn the full breadth of this topic in today's reality when researching on this investigation. We implemented algorithms on the Java language, and the results achieved may differ depending on the programming languages, technique, and machine architecture employed.

The accuracy of association rule mining has been evaluated using a variety of data structures, including the Frequent Pattern (FP) Growth, Apriori, weblog searching and modified web log searching algorithm. Partition-based techniques are also used for analysis in this study. Parallel processing is possible with the partition-based technique. It lowers the cost of calculation and improves the accuracy of both the association rule mining method. One of the simple examples of a transaction-based method is the apriori algorithm. The assessments were carried out using a variety of support count metrics, with the execution time of each method being recorded.

Comparative analysis with the existing method

Figure 6 (a), (b), and (c) show the results of the experiment, which indicate that the proposed strategy improves suggestion preciseness of accuracy, recollection, and F-score, respectively. Various CF-based approaches are evaluated to the suggested technique. A previous study used techniques and K Means cluster analysis approaches to solving the information complexity [17,21]. When dealing with huge amounts of data, this approach suffers from certain data leakage and demands a high computing cost. However, the proposed solution uses two-level clustering algorithms to tackle the data scalability issue. It analyses the hidden information using association rules to determine the user's unknown choice. The latent characteristics emphasis link between the user, rating, and tag is discovered using the Probabilistic matrix factorization approach. Both overlapping conditions and multiple tags associated with objects are effectively handled by the suggested method. As compared to conventional techniques, it increased the precision value by 5%. The number of relevant recommended items is represented by recall values. It demonstrates that it improves suggestion accuracy. The experiment shows, performance is efficient during the levels of data sparsity, based on the results. The findings reveal that user choice is determined not just by the rating score, but also by a thorough examination of the concealed pattern information. As a result, analyzing user preference behavior improves suggestion accuracy.

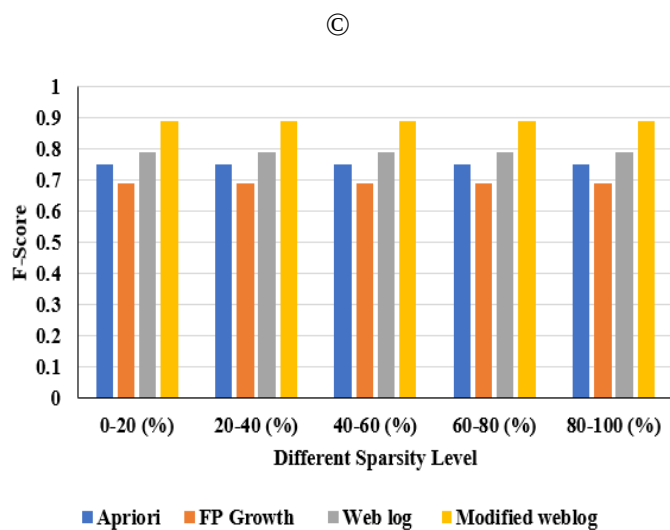
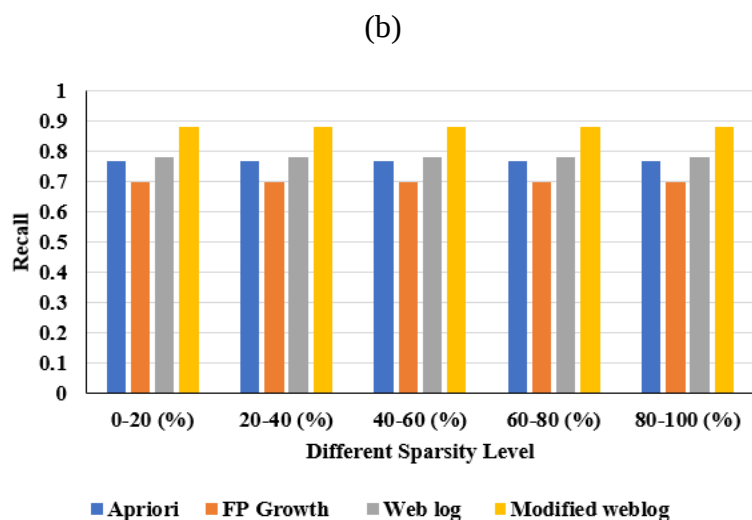
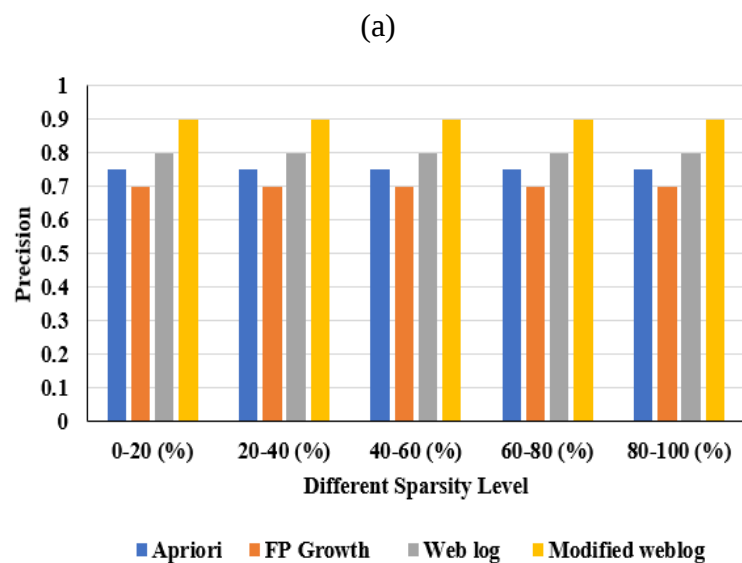


Figure 6: (a) Precision (b) Recall (c) F-Score

The obtained results show that the technique attains an average value of 5 per cent increased precision rate, 3% greater recall value or 4% greater F-value after compared to conventional CF technique.

Analysis of Algorithms

The proposed algorithm is analyzed to detect network anomalies and to identify four basic assaults like Denial of Service (DoS), Probe, Remote to Local (R2L) as well as User to Root (U2R). However, the prevailing models are efficient in detecting just a few of the aforementioned attacks while having inadequate detection rates for the rest. This deficiency makes it difficult to choose an appropriate IDS model when a user does not know what attacks to expect. Thus, there is a need for an intrusion detection systems (IDS) model that can detect, with uniform efficiency, all the four main classes of network intrusions. The proposed model can detect DoS, Probe, R2L, and U2R attack classes with uniform and high efficiency as shown in Figure 7. The resultant model achieved over 85% detection rates for each of DoS, probe, R2L, and U2R attacks.

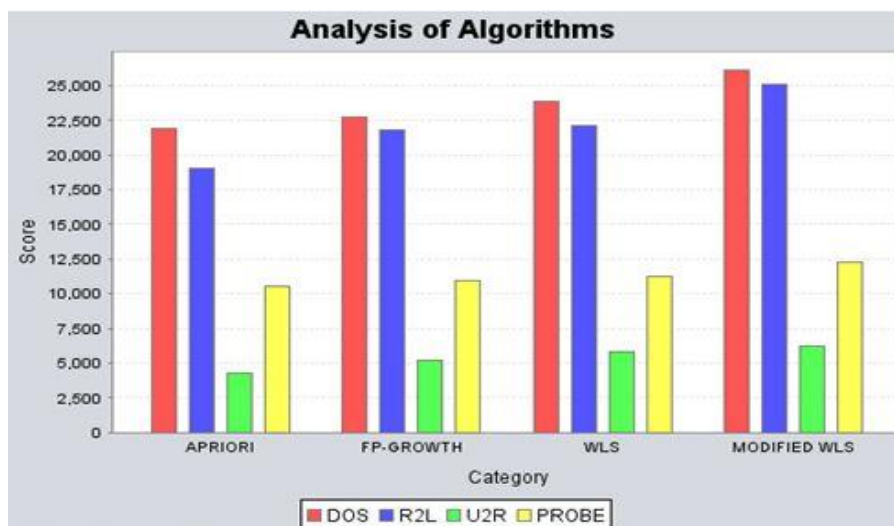


Figure 7: Analysis of intrusion detection systems

The suggested technique has the important benefit of enriching the user account by grouping item attributes like item tagging idea, item category, and object novelty depending on time. In addition, the technique extracts hidden insights from item preferences to forecast unknown preferences. The predicted technique has the advantage of assessing the most often recurring things, revealing underlying links, association, as well as the structure behind the selected items. Rather than proposing popular goods, the proposed system analyses the user's hidden interests and recommends items derived from past interest patterns. Because users' interests change over time, the suggested method analyses and ranks users' interests based on recent desires and do not recommend things that are no longer in use.

IV. CONCLUSION

This study classifies user behavior by recognizing patterns in web users' browsing and movement data, as well as comparing the performance of the Apriori algorithm, FP Growing

algorithm, web log search algorithm, and modified web log searching algorithm. Web Log searching algorithm mines log data (Web server access logs, browser logs, profiles, registration data, user sessions or transactions, cookies, user searches, mouse clicks, or any data resulting from Web interaction) obtained from user engagement over an amount of time. In addition, the suggested Web Log searching technique improves data set trimming. In reality, it has been clearly tested that it mines valid data set results, when earlier algorithms failed to do so. The algorithm was created without the use of any earlier algorithms. Repetitive log database searching is no longer necessary. The proposed Web Log searching technique has been shown to be extremely time efficient. The suggested approach will help with the efficiency, authorization, and correctness of log information sets retrieved from a log repository. Modified Web log searching algorithm removes the loopholes in the Apriori, FP Algorithm and web log algorithms. The total performance of predicted web log techniques was better, and memory utilization possible for a variety of data types. In comparison with existing strategies, the proposed Modified Web log searching algorithm takes less time. The proposed technique has the enormous value of enriching the user account by combining item attributes like item and exclusive, item classification, and object uniqueness depending on time. In addition, the technique extracts secret information from item preferences in ability to forecast new desires.

REFERENCES

1. Al, A. N. J. (2021, December). Efficient Visualization Framework for Real-Time Monitoring Network Traffic of High-Speed Networks. In 2021 IEEE International Conference on Big Data (Big Data) (pp. 5839-5842). IEEE.
2. Ali, S., Hafeez, Y., Humayun, M., Jamail, N. S. M., Aqib, M., & Nawaz, A. (2022). Enabling recommendation system architecture in virtualized environment for e-learning. *Egyptian Informatics Journal*, 23(1), 33-45.
3. Al-Shargabi, A. A., & Siewe, F. (2020, November). A Lightweight Association Rules Based Prediction Algorithm (LWRCCAR) for Context-Aware Systems in IoT Ubiquitous, Fog, and Edge Computing Environment. In Proceedings of the Future Technologies Conference (pp. 22-30). Springer, Cham.
4. Bellini, P., Palesi, L. A. I., Nesi, P., & Pantaleo, G. (2022). Multi Clustering Recommendation System for Fashion Retail. *Multimedia Tools and Applications*, 1-28.
5. Cahyati, M., & Mandiri, R. H. S. N. (2018). Implementasi Data Mining Penjualan Tas Pada Toko Fabella Shop Menggunakan Algoritma Apriori. *Speed-Sentra Penelitian Engineering dan Edukasi*, 10(4).
6. Celdrán A.H., Pérez M.G., Clemente F.J.G., and Perez G.M., "Design of a recommender system based on users' behavior and collaborative location and tracking", *Journal of Computational Science*, vol. 12, pp.83-94, Jan. 2016.

7. Chailes, A., & Hermawan, A. (2020). Application of Data Mining Method to Determine Purchasing Patterns Using Apriori Algorithms and Fp-Growth in Mukara Stores. *ALGOR*, 1(2), 1-8.
8. Chandrashekar, D. K., Srikantaiah, K. C., & Venugopal, K. R. (2021). Spark-Based FP-Growth Algorithm for Generating Association Rules from Big Data. In *Communication and Intelligent Systems* (pp. 659-668). Springer, Singapore.
9. Jalali, M., Mustapha, N., Sulaiman, M. N., & Mamat, A. (2010). WebPUM: A Web-based recommendation system to predict user future movements. *Expert Systems with Applications*, 37(9), 6201-6212.
10. Kanna, R. R. (2018). An Efficient Web Log File Classification Techniques to Identify the Fault Data Identification Using Multi-Class Support Vector Machine Algorithm. *Journal of Computational and Theoretical Nanoscience*, 15(9-10), 2831-2835.
11. Karthik, E., & Sethukarasi, T. (2022). Sarcastic user behavior classification and prediction from social media data using firebug swarm optimization-based long short-term memory. *The Journal of Supercomputing*, 78(4), 5333-5357.
12. Költringer, C., & Dickinger, A. (2015). Analyzing destination branding and image from online sources: A web content mining approach. *Journal of Business Research*, 68(9), 1836-1843.
13. Kumar, A., Bhushan, B., Pokhriya, N., Chaganti, R., & Nand, P. (2022). Web Mining and Web Usage Mining for Various Human-Driven Applications. In *Advanced Practical Approaches to Web Mining Techniques and Application* (pp. 138-162). IGI Global.
14. Mariani, M. M., Perez-Vega, R., & Wirtz, J. (2022). AI in marketing, consumer research and psychology: A systematic literature review and research agenda. *Psychology & Marketing*, 39(4), 755-776.
15. Masih, F. J., & Punhani, R. (2017). Analyzing E-Commerce website using Google Analytics Tool to enhance Customer Interaction. *JIMS8I-International Journal of Information Communication and Computing Technology*, 5(1), 280-284.
16. Nakayama, T., Kato, H., & Yamane, Y. (2000). Discovering the gap between web site designers' expectations and users' behavior. *Computer Networks*, 33(1-6), 811-822.
17. Rani, L. N., Defit, S., & Muhammad, L. J. (2021). Determination of Student Subjects in Higher Education Using Hybrid Data Mining Method with the K-Means Algorithm and FP Growth. *International Journal of Artificial Intelligence Research*, 5(1), 91-101.
18. Ranjan, R., & Sharma, A. (2019). Evaluation of frequent itemset mining platforms using apriori and fp-growth algorithm. *International Journal of Information Systems & Management Science*, 2(2).

19. Rao, K. S., Krishnamurthy, M., & Kannan, A. (2015). Extracting the user's interests by using Web log data based on Web usage mining. *Journal of Computational and Theoretical Nanoscience*, 12(12), 5031-5040
20. Sowmya, H. K., & Anandhi, R. J. (2022). Semantic Based Weighted Web Session Clustering Using Adapted K-Means and Hierarchical Agglomerative Algorithms. *Journal of Web Engineering*, 239-264.
21. Tommasi, F., Catalano, C., Corvaglia, U., & Taurino, I. (2022). MinerAlert: an hybrid approach for web mining detection. *Journal of Computer Virology and Hacking Techniques*, 1-14.
22. Zeniarja, J., Mangkawa, D. L. A., & Salam, A. (2021). Data Mining Applications for Violence Pattern Analysis with FP-Growth Algorithm. *Journal of Applied Intelligent System*, 6(1), 32-39.
23. Velásquez J.D, "Web mining and privacy concerns: Some important legal issues to be consider before applying any data and information extraction technique in web-based environments," *Expert Systems with Applications*, vol. 40, no. 13, pp. 5228-5239, 2013.
24. Wandri, R., & Hanafiah, A. (2022). Analysis of Information Technology (IT) Goods Sales Patterns Using the FP-Growth Algorithm. *IT Journal Research and Development*, 130-141.
25. Wisely Joe, J., Ghinaiya, M., & Syed Ibrahim, S. P. (2020). Effective Mining of High Utility Itemsets with Automated Minimum Utility Thresholds. In *Intelligent Computing and Innovation on Data Science* (pp. 277-285). Springer, Singapore.
26. Wu, T. Y., Lin, J. C. W., Yun, U., Chen, C. H., Srivastava, G., & Lv, X. (2020). An efficient algorithm for fuzzy frequent itemset mining. *Journal of Intelligent & Fuzzy Systems*, 38(5), 5787-5797.
27. Sidhu, S., Meena, U. K., Nawani, A., Gupta, H., & Thakur, N. (2014). FP Growth algorithm implementation. *International Journal of Computer Applications*, 93(8).
28. Shaikh, A. (2015). Web Usage Mining Using Apriori and FP Growth Alogrithm. *IJCSIT International Journal of Computer Science and Information Technologies*, 6(1), 354-357.
29. Santra A.K., and Jayasudha S., "Classification of web log data to identify interested users using Naïve Bayesian classification", *International Journal of Computer Science Issues (IJCSI)*, vol. 9, no. 1, p.381, 2012.
30. Shabtay, L., Fournier-Viger, P., Yaari, R., & Dattner, I. (2021). A guided FP-Growth algorithm for mining multitude-targeted item-sets and class association rules in imbalanced data. *Information Sciences*, 553, 353-375.

31. Rachman, R. (2021). Penentuan Pola Penjualan Media Edukasi dengan Menggunakan Metode Algoritme Apriori dan FP-Growth. *Paradigma-Jurnal Komputer dan Informatika*, 23(1). 2070.
32. Parvatikar, S., & Joshi, B. (2014). Analysis of user behavior through web usage mining. *Int. J. Comput. Appl.*, 09750-8887.
33. Nafie Ali, F. M., & Mohamed Hamed, A. A. (2018). Usage Apriori and clustering algorithms in WEKA tools to mining dataset of traffic accidents. *Journal of Information and Telecommunication*, 2(3), 231-245.
34. Mishra R., and Choubey A., “Discovery of frequent patterns from weblog data by using FP-growth algorithm for web usage mining”, *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 2, no. 9, Sep. 2012.
35. Mishra R., Kumar P., and Bhasker B., “A web recommendation system considering sequential information”, *Decision Support Systems*, vol.75, pp.1-10, July 2015.
36. Mishra, S., Patnaik, S., & Mishra, B. B. (2022). Use of Big Data Analysis in Data Management Aspects. In *Advances in Data Science and Management* (pp. 19-28). Springer, Singapore.
37. Kumar, S. (2020). Relevance of Buddhist Philosophy in Modern Management Theory. *Psychology and Education*, Vol. 58, no.2, pp. 2104–2111.
38. Allugunti V.R (2022). A machine learning model for skin disease classification using
39. convolution neural network. *International Journal of Computing, Programming and Database Management* 3(1), 141-147
40. Carmona, C. J., Ramírez-Gallego, S., Torres, F., Bernal, E., del Jesus, M. J., & García, S. (2012). Web usage mining to improve the design of an e-commerce website: OrOliveSur.com. *Expert Systems with Applications*, 39(12), 11243-11249.
41. Tanasa, D., & Trousse, B. (2004). Advanced data preprocessing for intersites web usage mining. *IEEE Intelligent Systems*, 19(2), 59-65.
42. Mobasher, B., Cooley, R., & Srivastava, J. (2000). Automatic personalization based on web usage mining. *Communications of the ACM*, 43(8), 142-151.
43. Nasraoui, O., Soliman, M., Saka, E., Badia, A., & Germain, R. (2007). A web usage mining framework for mining evolving user profiles in dynamic web sites. *IEEE transactions on knowledge and data engineering*, 20(2), 202-215.