A Proposed Paradigm for Enhancing Customer Retention using Web Usage Mining

Merna Ashraf
Business Information System Dept
Faculty of Commerce and Business Administration

Shimaa Ouf
Business Information System Dept
Faculty of Commerce and Business Administration

Yehia Helmy
Business Information system Dept
Faculty of Commerce and Business Administration

ABSTRACT
Due to the rapid growth of the internet and the emergence of the World Wide Web, there has been a huge amount of data stored in databases which increase the opportunities and relationships between companies and their customers. Although Companies find difficulty in satisfying customers with different backgrounds, so they find that it is important to adopt a strategy that helps them to understand and manage the needs of the organization's current and potential customers which are called customer relationship management(CRM). The purpose of CRM is to create value for customers and help organizations to gain "competitive advantage" over competitors. Because many businesses face problems in how to benefit from this huge amount of data that come from the internet, this leads to the emergence of web mining, web mining uses data mining techniques to improve business. Because customer retention is the core of CRM. The purpose of this research is to improve customer retention by merging five data mining techniques in the pattern discovery phase of web usage mining in order to enhance the recommender system that will increase customer retention.

Keywords
Web usage mining, data pre-processing, web server log, Pattern discovery.

1. INTRODUCTION
Because of rapid growth of technology, most people replace their routine transactions to online transactions, which lead to an enormous amount of data everywhere, businesses have difficulty in how to acquire and use these data to retain their customers and increase business's revenues(Gupta and Kohli 2019).

Web usage mining is the process of finding out and extracting what users are looking for on the internet, by using data mining techniques to discover usage patterns from web data and to better serve the needs of web-based applications(Chavda, Jain et al. 2017).

Data mining means extracting knowledge from a large set of data and transform it into an easily interpretable structure that helps organizations to analyze large and complex situations that result into customer satisfaction and turn over to firms(Ngai, Xiu et al. 2009) .

Retaining customers and increasing their satisfaction is the core of CRM, applying data mining techniques in CRM help organizations to become customer-focused, since that not all customers are equally important, building relationship with valuable customers such as high lifetime customer is a critical situation (Vannieuwenborgh 2018).

Retention of customers gives greater benefits over the acquisition of new customers. Researchers have found a 5% increase in customer retention boosts life time customer profits by 50% on average across multiple industries such as insurance(Soltani and Navimipour 2016).

This research is organized as follows: In Section2, the research methodology used in this study is reported. Section3 Research papers on web usage mining are analyzed. In Section4, a proposed framework is introduced and analyzed which helps in enhancing quality of recommendations offered to uses and finally the Conclusion.

2. RESEARCH METHODOLOGY
There is a lot of academic researches focus on web usage mining in predicting customer behaviors. Many scientific databases are searched to achieve a full bibliography of research papers on this field especially the most popular scientific databases: **Science Direct, Springer Link, and IEEE Library.**

The purpose of the search is to find research papers that relate to web usage mining and its role in increasing customer retention through recommender systems. The search was performed based on four keywords "web usage mining and customer relationship management", "web usage mining and customer retention", "web usage mining and data mining", "web usage mining and recommender system".

The results of the search process on the 3 most famous scientific databases were 111 articles. These papers published between 2015 and 2019.

2.1 Classification of research papers
The papers are categorized based on publication year and scientific databases.

2.1.1 Classification by publication's year
Classification of research papers by publication's year between 2015 and 2019 has shown in **Figure1**.

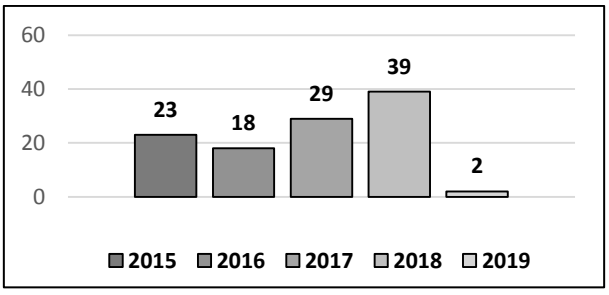

Fig1.classification of research paper by publication's year.

2.1.2 Classification by scientific databases

The Research papers are chosen from many scientific databases but most of them are chosen from the 3 most popular databases.

- Science Direct introduced (46 research papers).
- Springer Link introduced (54 research papers)
- IEEE Library introduced (11 research papers).

Classification of research papers for the most 3 scientific databases shown in **Figure2**.

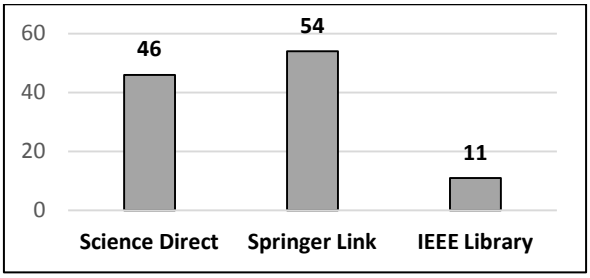

Fig2.classification of research paper by scientific databases.

3. LITERATURE REVIEW

Understanding customers' preferences and behavior is important for the success of any organization because by understanding their wants, hobbies, tastes, and interests. Organizations can provide them with products and services that make them feel satisfied.

Using new technologies to understand customers and predict their needs are in increasing to retain customers by recommending them with what they want and that leads to higher revenues to organizations.

Applying web usage mining to improve the accuracy of recommender systems is very important because most of the websites such as Amazon and Netflix rely on how to recommend the right product and service to retain their customers.

In this research web, usage mining will be used by merging the most known data mining algorithms to increase customer retention by providing them with a more accurate recommender system.

There are several studies in the area of applying web usage mining to understand customers. These studies focus on web usage mining and its role in customer relationship management or recommender system.

(Ismail, Ibrahim et al. 2015)Review the application of data mining in e-commerce by focusing on structured and unstructured data, collected through various resources and cloud computing services to justify the importance of data mining. They present a clear guide to e-commerce companies sitting on a huge volume of data to easily manipulate the data for business improvement.

(Chavda, Jain et al. 2017)indicate future trends of web mining by giving a brief idea regarding web mining concerned with its techniques, tools, and applications.

(Soltani and Navimipour 2016)mention that there is no comprehensive and systematic study about reviewing and analyzing that information system and CRM is the overall process of building and maintaining profitable customers. They present a comprehensive study and survey on the state of the art mechanisms in the scope of Customer relationship management.

(Bahari and Elayidom 2015)introduce the CRM-data mining framework establishes close customer relationships and manages the relationship between organizations and customers in today's advanced world of businesses.

(Lopes and Roy 2015)address how the organization uses web log files data to attract new customers and retain the existing ones. Providing real-time dynamic recommendation techniques to all the visitors of websites.

(Bose 2016) mentions that leading e-commerce retailers such as Amazon, E-Bay, and others have been leveraging the power of data to make informed decisions. He provides a useful, in-depth about data mining applications.

(Vaish, Vaish et al. 2016)mention how customer relationship management makes the process of acquiring, developing and maintaining relationships with customers more effective and efficient By describing the most recent tool in mining customers relationship management.

(Vannieuwenborgh 2018)indicate that companies must pay attention to the factors influencing the adoption of data mining in CRM by introducing an overview of data mining techniques, data mining tools, and its role in CRM.

(Kaur and Aggarwal 2015)mention why Web usage mining is crucial for Customer Relationship Management (CRM) by Analyze web log files.

(Isinkaye, Folajimi et al. 2015)address How to overcome the problem of information overload by exploring the different characteristics and potentials of different prediction techniques in recommendation systems.

(Nigam, Tokekar et al. 2015)mention that the Prediction of web user behavior is the demand of today competitive edge of the World Wide Web by evaluated and compared various models for predicting the next web page accessed by the web user.

(Gupta and Kohli 2019)Handling outliers in web data by introducing a framework to find outliers in the output of a regression algorithm is being formulated with the help of Ordered Weighted operators.

(Lu, Wu et al. 2015)Handling information overload problem by reviews up-to-date application developments of recommender systems.

(Suchacka, Chodak et al. 2017)identify e-commerce behavior characterization based on web server log data by Applying association rule mining.

(Agrawal and Jawdekar 2016)address what is the user-based data that can be fetched from the weblog by Explore user data from weblog based on time-dependent and time-independent.

(Sellamy, Fakhri et al. 2018)indicate a Short presentation of web mining its techniques, tools, and applications by Propose an approach for cross-analysis between skills acquired in university and skills sought by employees .

There are different data mining techniques to enhance customer retention. The proposed model will merge them to increase accuracy of recommendations offered to customers.

4. PROPOSED FRAMEWORK

As viewed in the previous section, the research papers focus on the web usage mining concept, its phases and data mining techniques that are used in the pattern discovery phase. Also, researchers proposed algorithms to clean data, identify user and session. Many researchers focus on using two or more

techniques to solve a particular problem. The researchers do not focus on providing a more personalized recommendation to users by creating an environment that applies three phases of web usage mining and merging data mining techniques to better understand users and find the most patterns that can be used to predict user's future needs.

Figure3 presents a proposed framework for providing more personalized recommendations to the user to increase user satisfaction. Following subsections discuss the work:

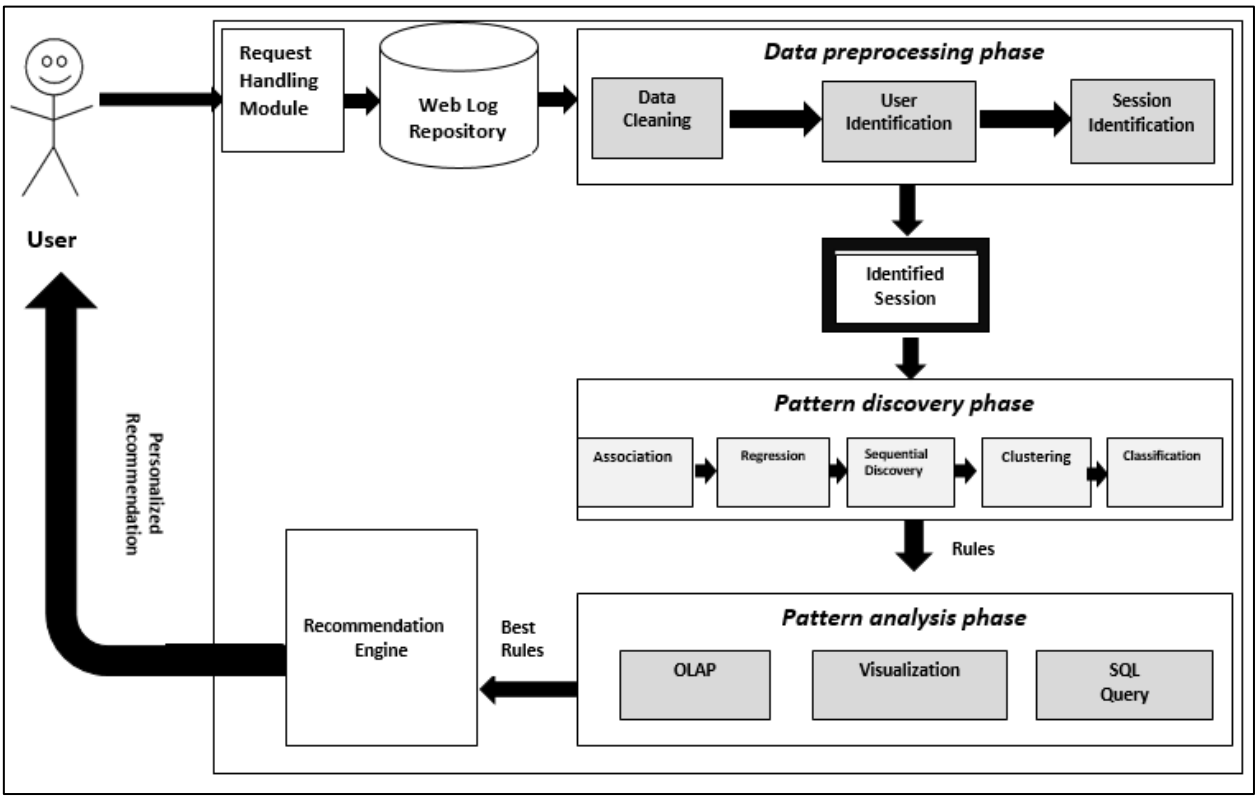

Fig3. Framework for Personalized Recommendations.

A. **Request Handling Module:** handles the requests made by users. Users' requests are logged in the weblog file.

B. **Web log repository:** web log file include access information such as IP address ,Date/ Time ,Method ,URL ,Protocol ,Status code ,Bytes ,Referrer and Agent.

C. **Data pre-processing phase**: this phase retrieves raw data from the log file and processed them. Raw data must be processed to increase effectiveness and decrease the mining time. Data pre-processing phase includes three important processes which are data cleaning, user identification and session identification.

- **Data cleaning:** log file containing a large number of errors, incomplete information, irrelevant data and noise that need to be removed, such as graphics file, unsuccessful status code.
- **User identification:** identifying each distinct user by IP address, Cookies and so on.
- **Session identification:** The session is the sequence of pages viewed by a single user during a specified time. The session can be identified by time heuristics or navigational heuristics.

D. **Pattern discovery phase:** in this stage pre-processed information is analyzed to extract valuable patterns, these patterns are used to recognize paths that users follow on the website. This phase depends on data mining techniques such as Association rule, Sequential discovery, Clustering, Regression, and Classification.

5. CONCLUSION

Organizations are using technology to improve their business and retain their customers. The web is the most important medium to conduct business and commerce. Web usage mining allows organizations to understand their customers by tracking their behavior in order to know the customer 'needs and predict their future needs and wants. This paper proposed a framework for providing a personalized recommendation to customers by merging data mining techniques which result in enhancing customer retention which is an important part of customer relationship management (CRM).

6. REFERENCES

[1] Agrawal, N. and A. Jawdekar (2016). User-based approach for finding various results in web usage mining. 2016 Symposium on Colossal Data Analysis and Networking (CDAN), IEEE.

[2] Bahari, T. F. and M. S. J. P. c. s. Elayidom (2015). "An efficient CRM-data mining framework for the prediction of customer behaviour." 46: 725-731.

[3] Bose "A Practical Guide to Data Mining for E-Commerce Business", 2016.

[4] Chavda, S., et al. (2017). "Recent Trends and Novel Approaches in Web Usage Mining."

[5] Gupta, A. and S. J. I. F. Kohli (2019). "FORA: An OWO based framework for finding outliers in web usage mining." 48: 27-38.

[6] Isinkaye, F., et al. (2015). "Recommendation systems: Principles, methods and evaluation." 16(3): 261-273.

[7] Ismail, M., et al. (2015). "Data Mining in electronic commerce: benefits and challenges." 8(12): 501.

[8] Kaur, N. and H. J. I. J. o. C. A. Aggarwal (2015). "Web log analysis for identifying the number of visitors and their behavior to enhance the accessibility and usability of website." 110(4): 25-30.

[9] Lopes, P. and B. J. P. C. S. Roy (2015). "Dynamic recommendation system using web usage mining for e-commerce users." 45: 60-69.

[10] Lu, J., et al. (2015). "Recommender system application developments: a survey." 74: 12-32.

[11] Ngai, E. W., et al. (2009). "Application of data mining techniques in customer relationship management: A literature review and classification." 36(2): 2592-2602.

[12] Nigam, B., et al. (2015). "Evaluation of models for predicting user's next request in web usage mining." 4: 1-13.

[13] Sellamy, K., et al. (2018). Web mining techniques and applications: Literature review and a proposal approach to improve performance of employment for young graduate in Morocco. 2018 International Conference on Intelligent Systems and Computer Vision (ISCV), IEEE.

[14] Soltani, Z. and N. J. J. C. i. H. B. Navimipour (2016). "Customer relationship management mechanisms: A systematic review of the state of the art literature and recommendations for future research." 61: 667-688.

[15] Suchacka, G., et al. (2017). "Using association rules to assess purchase probability in online stores." 15(3): 751-780.

[16] Vaish, A., et al. (2016). "Customer relationship management (CRM) towards service orientation in hospitals: A review." 13(4): 224-228.

[17] Vannieuwenborgh, S. (2018). Data Mining in Customer Relationship Management, Open Universiteit Nederland.