CHRIST
(DEEMED TO BE UNIVERSITY)
BANGALORE · INDIA

Notice for the PhD Viva Voce Examination

Ms Gokulapriya R (Registration Number: 1560078), PhD scholar at the School of Engineering and Technology, CHRIST (Deemed to be University), Bangalore will defend her PhD thesis at the public viva-voce examination on Saturday, 7 January 2023 at 10.00 am in the CDI Conference Room, Block V, Bangalore Kengeri Campus, Bengaluru 560074.

Title of the Thesis	**:**	**Web User Access Log Analytics Using Neural Learning, Regression, and Logit Boost Clustering Techniques for Accurate User Behavioural Pattern Identification**
Discipline	**:**	**Computer Science and Engineering**
External Examiner (Outside Karnataka)	**:**	**Dr R Suguna** Professor Department of Computer Science and Engineering Vel Tech Rangarajan Dr. Sagunthala R & D Institute of Science and Technology No. 42, Avadi-Vel Tech Road, Vel Nagar, Avadi Chennai - 600 062 Tamil Nadu
External Examiner (Within Karnataka)	**:**	**Dr H S Guruprasad** Professor Department of Information Science and Engineering BMS College of Engineering P.O. Box No.: 1908, Bull Temple Road Bengaluru – 560019, Karnataka
Supervisor	**:**	**Dr Ganesh Kumar R** Associate Professor Department of Computer Science and Engineering School of Engineering and Technology CHRIST (Deemed to be University) Bengaluru – 560074 Karnataka

The members of the Research Advisory Committee of the Scholar, the faculty members of the Department and the School, interested experts and research scholars of all the branches of research are cordially invited to attend this open viva.

Registrar

Place: Bengaluru
Date: 02 January 2023

ABSTRACT

Web Usage Mining (WUM), is the process of mining user behaviour patterns from huge log files. Weblogs provide substantial input to learning the identity of an online user. Analysis of these patterns extracted from the weblog datasets is currently being explored by various researchers. Due to the recent advent of automation, mining patterns from weblogs are automated. These automated mining processes focus on browsing habits and usage patterns. To make this process of gathering better, there are many ways to look at howusers act and put them into relevant groups.Identifying, detecting, and classifying features that demarcate specific traits that are related is an important task. Conventional research is designed to discover web usage mining strategies through clustering and classification methods. However, there is a need to focus on and improve the accuracy of the prediction systems that classify acquired features to figure out the patterns of web users. Deep learning methods are used to mine weblog data to improve accuracy and precision. To improve user behaviour pattern mining, a two-level clustering process is introduced as Ensemble Fuzzy K-Meanswith Logit Boost Clustering (EFK-LBC) technique to extract the weblog. In this technique, a preprocessing step is included to remove redundant data and choose reliable log files. The Fuzzy-K means clustering technique is used to identify behavioural patterns exhibited by recurrent users. Finally, the Logit Boost Clustering method is introduced to the data,that help in generating a strong cluster. Clustering of web users' frequent behavioural patterns using the Logit Boost ensemble technique helps the proposed EFK-LBC method to improve the accuracy up to 88% and reduce the clustering time by 20% compared with existing approaches. Though the proposed EFK-LBC technique performs better for user identification, the different initialization of clusters provides various final clustering results. This increases the error rate in the EFK-LBC technique. To solve this issue, the similaritybased grouping process was introduced. Based on the assimilation of patterns collected, identification of users based on web access behavioural patterns is essential. A Normal Discriminant Tanimoto Similarity Based Convolution Feedforward Neural Learning Classification (NDTS-CFNLC) method with different layers was proposed.

In the input layer, normal discriminant preprocessing is used to remove the redundant patterns from the access weblog files. Tanimoto Similarity at layer two supports fetching the relevantweb pattern. This process detects redundant patterns and eliminates them at the initial stages. A sigmoid function is appended to the output layer to classify frequently accessed patterns. It has helped in reducing the error rate by 12% and also improved the accuracy of 93% of web user identifications. NDTS-CFNLC method performs better, still, the web user behaviour pattern mining needs to be enhanced. To facilitate recognizing the frequently accessed web pattern a Mutual Information Pre-processing based Broken-Stick Linear Regression (MIP-BSLR) is used. The dependence is measured by MI-P in the weblog dataset. Based on the dependency measures results, the redundant patterns are isolated and the Broken-Stick Linear Regression analysis (BSLR) is implemented to recognize frequently accessed web patterns. This method predicts the web users' behaviour accurately.

TheMetrics used to measure the accuracy suggest that the model is able to get 95% accuracy compared with existingmethods. The proposed EFK-LBC technique, NDTS-CFNLC technique, and MIP-BSLR techniques are simulated on the NASA and ApacheWeblog Datasets for frequent user access pattern detection, analysis, and prediction. Based on the results obtained, it is evident that the proposed methodology proves to be more accurate than the existing models. The proposedMIP-BSLR method achieves an increased patternmining performance by 15% on the Apache Weblog dataset and a 16% significant increase over the NASA data set compared with other existing models. The time requirements of web pattern mining were reduced on both datasets to 33% and 63%, respectively. The proposed work has reduced the false positive rate by 66% and 52% on both datasets, respectively. A detailed comparative analysis is carried out to compare the accuracy and precision to establish the significance of the contribution made in this research work.

Keywords*: preprocessing, mutual information, linear regression, neural learning, fuzzy k- means, classification, clustering, user behaiour patterns*

PUBLICATIONS:

1. Gokulapriya R and Ganesh Kumar R, "Behavioral Pattern Mining for User Identity and Access Control A Cluster based Ensemble Model", International Journal of Engineering and Technology, Volume 7, Issue 3, 2018, Pages 438-444.
2. Gokulapriya R, and Ganesh Kumar R, "Research Aligned Analysis on Web Access Behavioral Pattern Mining for User Identification", in International Journal of Engineering and Advanced Technology (IJEAT) Vol. 8, Issue. 6, 2019 International Journal of Engineering and Advanced Technology, 8(6), 2249-8958.
3. Gokulapriya R, and Ganesh Kumar R, "Normal DiscriminantDeep convolution neural classification based web behavioral pattern mining for user identification", in Journal of Ambient Intelligence and Humanized Computing, Vol.12, Issue.6 , 2021.
4. Gokulapriya R, and Ganesh Kumar R, "Mutual Information based preprocessing with Broken stick linear regression Technique for web user behaviour pattern mining", in International Journal of Intelligent Engineering and Systems (IJIES) Vol. 14, Issue 1, 2021.