International Journal of Engineering & Technology, 7 (2.33) (2018) 518-521

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Discovering the behavior of the students using data mining techniques

S. Kamalakkannan [1] *, S. Prasanna [2]

[1] *Research Scholar, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, India*
[2] *Associate Professor, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, India*
**Corresponding author E-mail: kamalindia81@yahoo.com*

Abstract

The real issue of numerous online sites is the introduction of numerous decisions for the different users at once. This normally comes about into tedious undertaking in discovering the correct item or data on the site. The user present intrigue relies on the navigational conduct which causes the associations to control users in their perusing exercises and acquire some applicable data in a limited ability to focus time. Since, the subsequent examples, which are acquired through data mining systems, did not perform well in the forecast of future temples designs due to the low coordinating rate of coming about tenets and of user's perusing conduct. This paper centers around the investigation of the pro-grammed web use data mining and proposal framework, which depends on current user conduct through his/her, snap stream information. In this paper, we attempt to show signs of improvement understanding on how Internet utilization of understudy's conduct in Engineering College can influence on their everyday scholarly exercises additionally it thinks about the use examples of various department' understudies. What's more, we endeavor to discover similitudes and dissimilarities of use examples of understudies on different branches and discovering connections between Internet utilization examples of understudies and their student performance CPI (Cumulative Performance Index). This paper displays the consequences of an investigation for a time of three months, in regards to the behavior mining of understudies identified with their Internet use designs with examining access log documents.

Keywords: *Behavior Mining; Internet Usage Behavior; Web Log File; Web Usage Mining.*

1. Introduction

Education is a tremendous and critical concern, in current years the amount of data stored in academic database is developing swiftly. The saved database consists of hidden data approximately the development of the scholar's performance and behavior. The capability to are expecting the student overall performance in education could be very critical in academic environments. From this net mining the scholars' behavior is founded way by using the internet log files the scholars searched web pages and their behavior are noted.

Data mining is the system which comes underneath the class of pc innovative know-how so as to research expansive data collections which has a place with the example. Here colossal information set stands for Big Data. Information mining is a programmed way which is utilized to separate noteworthy data from the data carport and further utilize this insights for numerous purposes .The extraction of critical data might be performed through coordinating styles and it's far executed through group examination, peculiarity investigation, and reliance assessment. Spatial records are utilized to complete all above functions or strategies. The coordinated example is a state of brief abstract of the information put away inside the insights distribution center and these styles is utilized for predetermination expectation and different choice making frameworks to take legitimate choice.

Social Network sites permit people to:

1) Develop an open or semi-open profile interior of a partial framework,
2) express a list of various users and to whom the users share a courting,
3) Vision and traverse their listing of institutions and those complete by others within the structure.

Data mining is the method of studying big data sets to discover unpredicted interactions and to précis the facts in unique approaches that are each comprehensive and beneficial to the facts holders. These approaches are useful for examining Internet usage behavior and patterns. Methodologies for discovering and exhibiting relationships in

- Huge amount of data
- Discovering hidden information in a database
- Acceptable data to a model

To observe the usage of internet access and their behavior of students related to their Internet usage patterns with examining access log files. We use student's data to examine their learning behavior to predict the outcomes and to advise students at risk before their finishing exams and also it can help educational scholars and progress planners for improvement of coaching techniques with respect to Internet usage behavior connected to branch and gender variation.

2. Literature review

1) Esa Heikkinen, Timo D. Hämäläinen proposed that a new log file analyzing framework, (i.e) LOGDIG, for checking expected system behavior from log files. LOGDIG be capable of also be configured to examine other systems log files

by its flexible metadata formats and a new behavior mining language In the year 2015 [1].

2) Jincy B. Chrystal and Stephy Joseph said that the feature extraction and class of such textual content documents require an efficient device getting to know algorithm which plays automated textual content class. Text mining and classification of product reviews using structured support vector machine. This research defines the type of product estimate files as a multi-label ordering situation and reports the problem the use of Organized Support Vector Machine. The final results of this work is the categorized phrases inside the assessment text into more than one magnificence labels in step with the extracted structures. The accurateness and performance of the machine is restrained and located to be an enhanced method within the case of a Multi-label textual content class situation 2015 [2].
3) Er. Jyoti1, Er.Amandeep Singh Walia said that these paper emphases on the study of the automatic web usage data mining and commendation system which is based on present user behavior through his/her click stream information. The K-Nearest-Neighbor (KNN) classification Technique has been skilled for use in actual-time and on line to find users and visitors click on circulation data, equivalent it to a selected consumer institution and commends a tailored surfing option that happen the wishes of the unique user at specific time. In this paper, the hassle and numerous methods are explained for commendation fashions. In this paintings, knn algorithm is used. Their overall performance will be compared in phrases of error charge, memory required 2017 [3].
4) P. Tamiljeselvy, Sangavi.S, Suvetha. T, Umashankari. T, Web usage mining using Improved KNN Algorithm they said The Classification is done using improved K-NN classification algorithm and it has been skilled to be used online and in real time to identify peoples click stream data, matching it to a specific user cluster and suggest a tailored browsing possibility that meet the requirement of the specific user at a specific time. K-Nearest Neighbors Algorithm, Decision Tree Classification automatic Real-Time recommendation system. The system performs classification of users on the simulated active sessions extracted from testing sessions by collecting active users' click stream and matches this with similar class in the data mart, so as to generate a set of recommendations to the client in a Real-Time basis using improved k- NN classification 2017 [4].
5) Manisha Kumari, Sarita Soni, A Review of classification in Web Usage Mining using K- Nearest Neighbour., classification algorithms in pattern discoverty phase This paper has provided a survey about WUM processes and one of its classification technique KNN. KNN is very simple and if compared with other algorithms KNN again maintains its efficiency. KNN is a satisfactory classification technique used in WUM, but is a lazy learner and the accuracy depends on the value of k. There are bundles of paper available about KNN. After examining these papers we can be concludes that there are two approaches to improve the performance of KNN 2016 [5].

3. Proposed system

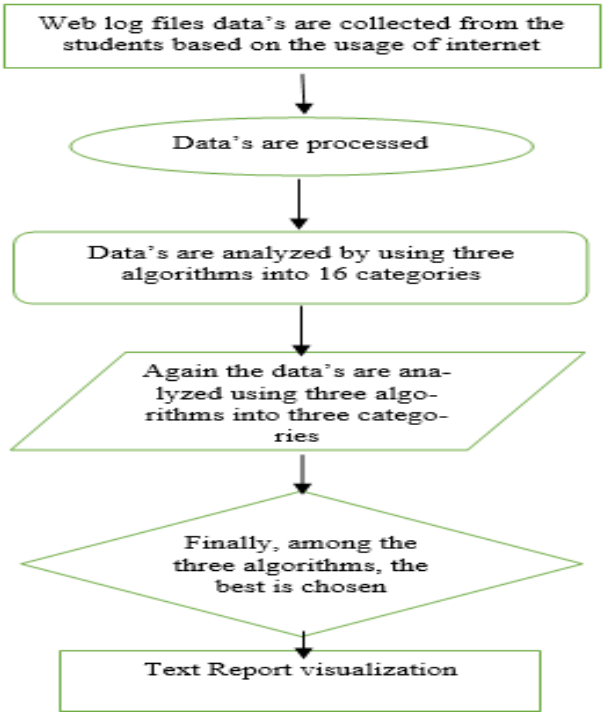

Fig. 1: Proposed Algorithm Flowchart.

3.1. Web log file

Web log document is a log record automatically made and kept up with the guide of a web server. Each "hit" to the Web page, alongside each perspective of a HTML archive, image or other object is logged. The raw net log record organize is generally one line of content for each hit to the web website. This includes information about who transformed into visiting the site, where they arrived from, and precisely what they had been doing at the web page.

3.2. Web mining

Web mining is the one of the data mining techniques to automatically find out and extract information from Web documents and services. There are totally three methods are relevant for web mining

- Web Content Mining
- Web Structure Mining
- Web Usage Mining

4. Data preprocessing

In data mining and statistical data evaluation, before models may be built or algorithms may be used information needs to be prepared. In this context, getting ready the statistical approach reworking them previous to the evaluation as a way to ease the algorithm's task. Often, the purpose might be to modify the statistics in order that the hypotheses, on which the algorithms are primarily based, are confirmed, while at the identical time retaining their information content intact. One of the maximum fundamental transformation is normalization [7].

In this paper, we use Third Normal Form and Boyce Codd Normal Form for finding the dependency among the dataset.

During the implementation process, Amassed raw net log files by the net mining strategies are preprocessed by way of normalization method to extract useful statistics from the raw web log documents where the unwanted and duplicate uncooked documents are eliminated. To discover the sample of those statistics, we apply a predictive version that is a data mining technique.

This predictive model is taken in which three algorithms are implemented

- K-Nearest Neighbor
- Naïve Bayes
- Support Vector Machine.

5. K-nearest neighbor algorithm

K nearest neighbor is classification algorithm that stores all to be had cases and categorizes new cases created on a similarity measure. The nearest neighbor algorithm (KNN) belongs to the class of pattern respect geometric strategies. The technique does now not execute a priori any rules approximately the delivery from which the modeling sample is drained. It includes a preparation set with each high-quality and poor values. A new sample is classed by means of manipulative the space to the closest neighboring education case [8].

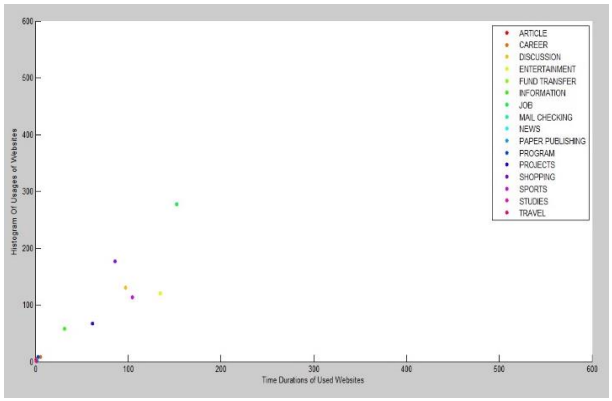

Fig. 2: K Nearest Neighbor Algorithm Output Screenshot.

From these 16 categories, the majority of the students frequently used internet access for the study purposes. Finally, a text report is generated from that we can analyze the behavior of students.

6. Naïve bayes algorithm

Naïve Bayes multi label classifier is a type classification algorithm it is based on Bayes theorem with an predict of independence amongst predictors. A Naive Bayes multi- label classifier predicts that the life of a specific function in a category is numerous to the presence of another function. This version is easy to build and in particular, this version may be used for terribly massive data units. Besides simplicity, Naive Bayes multi-label classifier is likewise may be used to carry out particularly state-of-the-art class strategies [9].

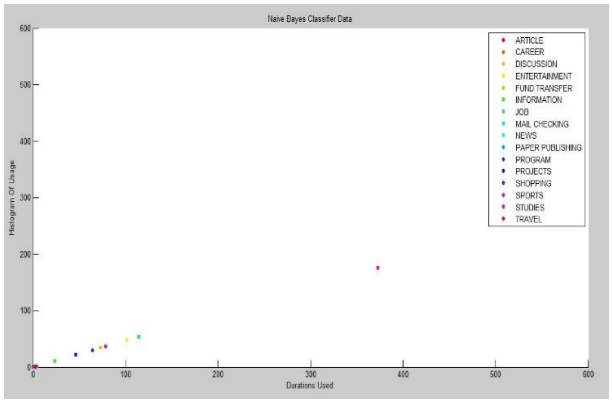

Fig. 3: Naïve Bayes Algorithm Output Screenshot.

By this algorithm, Students spent their time, largely on an article and career purposes is identified.

7. Support vector machine support

Vector Machine" (SVM) is a directed gadget studying set of rules which may be used for both type and regression demanding situations. It is primarily used in type troubles. In this algorithm, we plot each statistics item as a point in the n-dimensional area (in which n is the range of features you have got) with the value of each feature being the price of a particular coordinate. SVM is specially used to detect and make the most complicated styles in data via clustering, classifying and rating the facts. Learning machines which are used to perform binary classifications and regression estimations. They usually use kernel based techniques to use linear category techniques to nonlinear class troubles. There are a number of forms of SVM including linear, polynomial, sigmoid and so forth [10]. By using this algorithm, entire students' performance can be found.

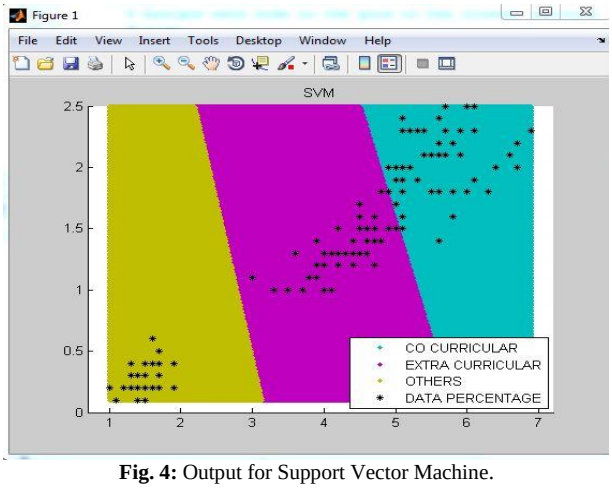

Fig. 4: Output for Support Vector Machine.

Table 1: SVM Percentage Category

Category	Percentage
Co- curricular	43%
Extra- curricular	36%
Others	21%

The above figure 4 explains the results of Students mining behavior. Here, 16 website categories is grouped into 3 categories, i.e. Co- curricular, Extra Curricular and others. 43% of students are using the internet for co-curricular purpose, 36% of students are using the internet for Extra curricular purpose and remaining 21% of students are using the internet for other purposes.

8. Conclusion

In this paper, we have discovered the use examples of understudies on different branches and discovering connections between Internet utilization examples of understudies. Alternate factors, for example, understudies' use practices of Internet as far as normal time spent every day on the Internet, the quantity of hits (in parts of opened or went by pages) and number of one of a kind Websites went to every day by understudies of various department or branches are also identified with the help of three algorithms such as Naïve Bayes classification, Support Vector Machine, K Nearest Neighbor Algorithm. From these text report is generated and also learning behavior of students are predicted to warn students before their final exams in terms of results. With the help of Receiver Operating Characteristics Curve, It is proven that the Naïve Bayes algorithm shows better results when compared to other two classifiers.

References

[1] Esa Heikkinen, Timo D. Hämäläinen LOGDIG, (2015). Log File Analyzer for Mining Expected Behavior from Log File.

[2] Jincy B. Chrystal and Stephy Joseph (2015) said that the feature extraction and classification of such text documents require an efficient machine learning algorithm
[3] Jyoti1, Er.Amandeep Singh Walia said that this paper focuses on the study of the automatic web usage data mining and recommendation system Theint Theint Aye, (2017). Web Log Cleaning For Mining of Web Usage Patterns.
[4] P. Tamiljeselvy, Sangavi.S, Suvetha. T, Umashankari. T, Web usage mining using Improved KNN Algorithm (2017) they said The Classification is done using improved K-NN classification algorithm
[5] Manisha Kumari, Sarita Soni (2016), A Review of classification in Web Usage Mining using K- Nearest Neighbour..
[6] Rozita Jamili Oskouei, Chaudhary, B.D... (2010). Internet Usage Pattern by Female Students: A Case Study, ITNG, Seventh International Conference on Information Technology, pp.1247-1250.
[7] Santhisree,K Dr Damodaram, A Appaji,S Nagerjunadevi,D (2010). Web usage data clustering using Dbscan algorithm and set Similarities.
[8] Tanasa, D et.al, Advanced data preprocessing for inter sites Web Usage mining, IEEE computer society.
[9] Wang, X Ouyang, Y Hu,X Zhang,Y Discovery of User Frequent Access Patterns on Web Usage Mining, IEEE 8th International Conference on Computer Supported Cooperative.
[10] Bhaskar, S., Singh, V. B., &Nayak, A. K. (2014, March). Managing data in SVM supervised algorithm for data mining technology. In IT in Business, Industry and Government (CSIBIG), 2014 Conference on (pp. 1-4). IEEE.