

# A Web Navigation Frame Work to Identify the Influence of Faculty on Students using Data mining Techniques

Sundari R\*

Department of Computer Science Engineering, GITAM University, India

## Abstract

Analyzing student web browsing behavior is a challenging task. This paper mainly focuses on a methodology to identify the influencing factor that has driven a student towards navigating a particular web site. Most of the research in this direction is untouched to estimate the influence of faculty on the student's behavioral patterns. In this work we focus on a novel statistical approach based on adaptive Gaussian mixture model, where the data clustered is given as input to the model to classify the student navigating pattern. The concept of regression analysis is used to find the relationship between student's navigational behavior and faculty's experience and rating. This article considers a real time dataset of GITAM University for experimentation.

**Keywords:** Gaussian mixture model; Regression; Clustering; Classification

## Introduction

Present day students flock to the internet as the primary tool for researching any topic. In most of the cases, the students get influenced by different factors and these influences make them drive towards goal setting. This paper examines the student behavior on web and estimates the influence of a faculty teaching on their behavior [1]. Perfect procedures are needed to find out this inclination, thus, faculty rating given by the students and experience of the faculty in a particular field of specialization were taken into consideration. In this work in-depth analysis of different kinds of students specifically related to the engineering group is concentrated [2,3].

Despite the fact that students join in an engineering college with the goal to receive degree, their future dreams are different [4,5]. This being the backdrop, we at GITAM University, have conducted a Brain storming session of about one week to motivate the students regarding various opportunities that are ahead before them. To understand the impact of this session on the students, we have provided specific IP address and monitored their web navigation pattern thereof.

A statistical framework is developed for clustering the students into different groups basing on their navigation pattern. The objective of this prediction is to characterize the student behavior relating to a particular cluster. After classifying the students to one of the predefined groups, regression analysis is measured to find out the relationship between the students browsing pattern and the influence of an experienced teacher or teacher who has been rated good.

## Related work

The experience of the faculty plays a vital role in impressing, motivating and educating the students in different ways. There is a lot of difference between experienced and inexperienced faculty in the way they elucidate the topic to the students. In the brain storming session conducted to the students to guide them to the future endeavors we tested practically how the experience of the faculty helps in motivating the students towards their journey of success. Kunyanuth et al. [6] proposed a model in order to guide the students in choosing their track in the field of computer science. To select appropriate fields student registration data, course data and class learning were analyzed using data mining techniques. This research aims at developing a decision support system for guiding the students in choosing the correct

field according to their abilities and interests. The data used in the experiments was collected from computer science program, Suan Sunandha Rajabhat University, during the period of 2006-2012. In the data gathering phase 4 quizzes based on computer science fields were conducted to the students in the subjects of data base, software engineering, multimedia and network and communication fields. The equal width method was used to partition the value of continuous attributes into five nominal values: VERY POOR, POOR, FAIR, GOOD and VERY GOOD. The data was analyzed by using naive Bayesian and decision trees classification techniques and the experimental results shown that naive Bayesian is more efficient than Decision trees.

This research helps in deciding whether a teacher with good experience can have the effect on student's behavior or a teacher rated well impresses the student well towards the lecture. Now a days rating had become a common measure for measuring the quality of everything from person to goods, this research confirms that a qualitative teacher or faculty can't be judged with his rating but with his experience.

## Dataset

This data set contains the sessionised data for the gitam.edu web server (<http://www.gitam.edu>). This data is based on the students navigating pattern for a period of one week during which the motivating session is delivered. The following snap shot presents a view of the dataset before preprocessing (Sample 1).

During preprocessing data is cleaned by removing whitespaces, images, audio and video files. The cleaned data is preprocessed for identifying sessions, different users; unique URL's or page views and time spent by them in each page view using different preprocessing techniques. As a first step in preprocessing all the Unique URL's in the dataset are identified and assigned unique identification numbers. In the second step, the dataset is presented with the student id, together

\*Corresponding author: Sundari R, Department of Computer Science Engineering, GITAM University, India, Tel: 9848375001; E-mail: [rekhasundari@yahoo.com](mailto:rekhasundari@yahoo.com)

Received March 20, 2015; Accepted July 27, 2015; Published July 31, 2015

**Citation:** Sundari R (2015) A Web Navigation Frame Work to Identify the Influence of Faculty on Students using Data mining Techniques. J Comput Sci Syst Biol 8: 239-242. doi: [10.4172/jcsb.1000195](https://doi.org/10.4172/jcsb.1000195)

**Copyright:** © 2015 Sundari R, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

with his access sequence. As sequence is not of priority in the proposed work, the third step is carried out, in which each entry in the dataset is redesigned such that if the student visits the page the entry is represented by 1 else 0. The Tables 1-3 below elucidate the outputs of the three preprocessing steps respectively.

**Clustering**

As the dataset under consideration consists different navigation patterns, to mine the relevant browsing patterns clustering is considered. In our work we use this technique to identify groups in students with browsing patterns [7].

**Latent class analysis**

Latent class analysis is considered for clustering the students into


groups. This is a model based cluster analysis technique that uses mixture of probability distributions to assign a data point to a cluster.

The basic latent class cluster model is given by

$$P(y_n|\theta) = \sum_1^S \pi_j p_j(y_n|\theta_j)$$
 (1)

Where  $y_n$  is the nth observation of the manifest variables, S is the number of clusters and  $\pi_i$  is the prior probability of membership in cluster j.  $p_j$  is the cluster specific probability of  $\theta_j$  given the cluster specific parameters  $\theta_j$ . For each data point LCA calculates the probability to the cluster membership. After the model is built data points are assigned to the clusters that have higher probability.

After performing clustering, we identified six different groups in 400 students and when we analyzed the URL's the groups have browsed



**Sample 1:** Dataset before preprocessing.

| ID | URL                                 |
|----|-------------------------------------|
| 1  | www.goabroad.com                    |
| 2  | www.afsusa.orglstudy-abroadlfaq     |
| 3  | www.uniguru.co.inl                  |
| 4  | www.usnews.com                      |
| 5  | www.i20fever.com                    |
| 6  | www.ustraveldocs.com                |
| 7  | www.topuniversities.comluniversity  |
| 8  | rankingslworld...rankingsl2013      |
| 9  | www.4icu.org > Asia                 |
| 10 | 100bestschools.netlenglschoolsl     |
| 11 | www.therichest.coml...lthe-top-10-  |
| 12 | mtechadm.iitm.ac.inladmproc.php     |
| 13 | www.ieor.iitb.ac.inlfaqImtech-      |
| 14 | www.cse.iitk.ac.inlacadladm mt.html |
| 60 | http:ltestfunda.comlCAT             |

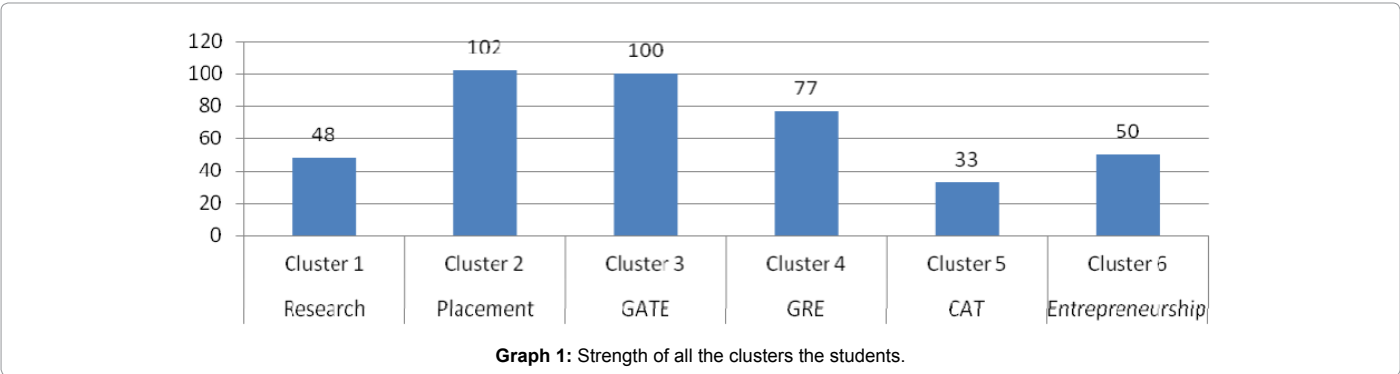
**Table 1:** Shows the list of distinct URL's , with their corresponding id's.

| User | Sequence    |
|------|-------------|
| 1    | 1,1         |
| 2    | 2           |
| 3    | 3,2,4,2,3,3 |
| 4    | 2,9,9,12    |
| 5    | 1,2,11,15,8 |
| 6    | 1,12,12,8   |

**Table 2:** A sample of user sequences.

| User, Page | P1 | P2 | P3... | P60 |
|------------|----|----|-------|-----|
| 1          | 1  | 0  | 0     | 0   |
| 2          | 1  | 0  | 0     | 0   |
| 3          | 0  | 1  | 1     | 1   |
| 4          | 0  | 1  | 0     | 0   |
| 5          | 1  | 1  | 0     | 0   |
| 6          | 1  | 0  | 0     | 0   |

**Table 3:** Sample user sequences represented by their visit to a page.



we identified that the students of six different clusters concentrate on six different goals as their future endeavors. The Graph 1 below shows the strength of all the clusters the students belong to:

|           |                  |
|-----------|------------------|
| Cluster 1 | Research         |
| Cluster 2 | Placement        |
| Cluster 3 | GATE             |
| Cluster 4 | GRE              |
| Cluster 5 | CAT              |
| Cluster 6 | Entrepreneurship |

## Classification

The objective of classification in this context is to assign the student to a particular cluster that describes the student behavior more relatively with his similar group.

### Adaptive gaussian mixture model

AGMM is an improvised version of GMM in which the probability density is a function of input vector  $x$ , mean  $\mu$ , standard deviation  $\sigma$  as equivalent to GMM [8] and with two additional parameters  $n$  and  $N$ . Where  $N$  is total number of samples present in the data and  $n$  is number of samples in each cluster.

The probability density function of Adaptive GMM is given by:

$$f(x, \mu, \sigma, n, N) = \frac{1}{\sqrt{2\pi}\sigma} \left( \frac{\mu}{n} \left( \frac{e^{-(x-\mu)^{n+1}}}{2\sigma^2} + \frac{\mu}{N} \sum_{i=1}^n \left( \frac{x-\mu}{\sigma} \right)^{\frac{\mu}{\sigma}} \right) \right) \quad (2)$$

Assume that each sample  $x$  is a  $d$  dimensional vector. Let  $x = [x_1, x_2, \dots, x_i, \dots, x_d]$ . As the features are independent, the mean and standard deviation are also calculated independently. For a cluster with  $n$  samples, the mean  $\mu$  and standard deviation  $\sigma$  of each feature  $x_i$  is calculated by taking the  $x_i$ 's of all the samples in that particular cluster. So the mean and the standard deviation are given by the equations 2 and 3.

$$\mu = \sum_{j=1}^n \frac{x_{ij}}{n} \quad (3)$$

$$\sigma = \sqrt{\sum_{j=1}^n \frac{(x_{ij} - \mu)^2}{n-1}} \quad (4)$$

After classification the new student with access sequence is assigned to one of the above mentioned clusters.

## Regression

A data mining (machine learning) technique used to fit an equation to a dataset is called Regression. This is a statistical model for estimating relationships among variables. Regression analysis with a single explanatory variable is termed simple regression or linear regression. Linear regression, uses the formula of a straight line ( $y=mx + b$ ) and determines the appropriate values for  $m$  and  $b$  to predict the value of  $y$  based upon a given value of  $x$ . Multiple regression, allow the use of more than one input variable and allow for the fitting of more complex models, such as a quadratic equation [4].

In this article the concept of regression analysis is used to estimate the regression pattern of  $X$  (Independent variables, various navigation patterns of the student) and  $Y$  (Dependent variables rating and experience of the faculty). In our work basing on the regression analysis we made an attempt to identify the regression of  $X$  and  $Y$  and thereby identify the significance and inclination of a faculty experience and rating on student browsing behavior. Considering the multiple linear regression analysis by taking independent variables (1 to 48) and

considering the rating and experience as dependent variables, we have the regression lines as

$$y = 8.04 + 28.0 X_2 + 193 X_3 - 177 X_4 - 318 X_5 + 317 X_6 - 17.0 X_7 + 12.2 X_9 - 12.3 X_{10} - 16.9 X_{11} + 2.1 X_{13} - 3.7 X_{15} - 114 X_{17} + 196 X_{18} + 40 X_{19} - 57 X_{20} - 81.2 X_{21} - 38.4 X_{23} + 6.6 X_{25} - 10.9 X_{26} + 0.1 X_{27} + 3.8 X_{28} + 9.7 X_{29} - 1.2 X_{31} - 65.1 X_{33} + 61.4 X_{34} + 41.4 X_{35} - 57.9 X_{36} + 5.7 X_{37} + 68.5 X_{39} + 294 X_{41} - 314 X_{42} - 301 X_{43} + 318 X_{44} - 11.3 X_{45} + 3.9 X_{47} + 6.56 X_{48}$$

and

$$y = 1.75 + 1.57 X_2 - 7.0 X_3 + 8.8 X_4 - 3.27 X_5 + 1.63 X_6 + 1.60 X_7 - 17.6 X_9 + 15.2 X_{10} - 4.18 X_{11} + 5.78 X_{13} - 12.0 X_{15} - 0.24 X_{17} + 1.29 X_{18} - 4.5 X_{19} + 7.0 X_{20} + 0.54 X_{21} + 1.94 X_{23} + 1.61 X_{25} + 0.42 X_{26} + 15.7 X_{27} - 15.1 X_{28} + 1.45 X_{29} + 0.08 X_{31} + 2.47 X_{33} + 0.06 X_{34} - 0.55 X_{35} + 3.36 X_{36} + 0.32 X_{37} + 0.78 X_{39} + 6.53 X_{41} - 9.80 X_{42} - 9.77 X_{43} + 17.1 X_{44} - 4.03 X_{45} - 3.28 X_{47} - 3.27 X_{48}$$

The corresponding results are presented in Tables 4 and 5. The results obtained are presented in Figures 1 and 2. From the above

| Predictor constant | Coef    | SE Coef | T     | P     |
|--------------------|---------|---------|-------|-------|
| <b>X1</b>          | 1.7468  | 0.9541  | 1.83  | 0.092 |
| <b>X2</b>          | 1.572   | 1.777   | 0.88  | 0.394 |
| <b>X3</b>          | -6.95   | 13.29   | -0.52 | 0.61  |
| <b>X4</b>          | 8.84    | 13.43   | 0.66  | 0.523 |
| <b>X5</b>          | -3.271  | 8.157   | -0.4  | 0.695 |
| <b>X6</b>          | 1.629   | 2.086   | 0.78  | 0.45  |
| <b>X7</b>          | 1.6     | 1.521   | 1.05  | 0.313 |
| <b>X9</b>          | -17.646 | 7.302   | -2.42 | 0.033 |
| <b>X10</b>         | 15.226  | 6.97    | 2.18  | 0.049 |
| <b>X11</b>         | -4.176  | 3.918   | -1.07 | 0.308 |
| <b>X13</b>         | 5.778   | 4.284   | 1.35  | 0.202 |
| <b>X15</b>         | -12.019 | 5.665   | -2.12 | 0.055 |
| <b>X17</b>         | -0.244  | 2.74    | -0.09 | 0.93  |
| <b>X18</b>         | 1.289   | 2.479   | 0.52  | 0.612 |
| <b>X19</b>         | -4.46   | 13.08   | -0.34 | 0.739 |
| <b>X20</b>         | 6.96    | 12.79   | 0.54  | 0.596 |
| <b>X21</b>         | 0.54    | 1.476   | 0.37  | 0.721 |
| <b>X23</b>         | 1.944   | 1.761   | 1.1   | 0.291 |
| <b>X25</b>         | 1.615   | 2.707   | 0.6   | 0.562 |
| <b>X26</b>         | 0.417   | 2.663   | 0.16  | 0.878 |
| <b>X27</b>         | 15.67   | 10.82   | 1.45  | 0.173 |
| <b>X28</b>         | -15.06  | 10.62   | -1.42 | 0.182 |
| <b>X29</b>         | 1.454   | 1.955   | 0.74  | 0.471 |
| <b>X31</b>         | 0.085   | 1.801   | 0.05  | 0.963 |
| <b>X33</b>         | 2.466   | 3.467   | 0.71  | 0.491 |
| <b>X34</b>         | 0.06    | 4.193   | 0.01  | 0.989 |
| <b>X35</b>         | -0.545  | 4.858   | -0.11 | 0.912 |
| <b>X36</b>         | 3.363   | 3.666   | 0.92  | 0.377 |
| <b>X37</b>         | 0.318   | 2.587   | 0.12  | 0.904 |
| <b>X39</b>         | 0.784   | 2.751   | 0.29  | 0.78  |
| <b>X41</b>         | 6.529   | 4.967   | 1.31  | 0.213 |
| <b>X42</b>         | -9.805  | 7.785   | -1.26 | 0.232 |
| <b>X43</b>         | -9.77   | 7.994   | -1.22 | 0.245 |
| <b>X44</b>         | 17.09   | 12.05   | 1.42  | 0.182 |
| <b>X45</b>         | -4.032  | 3.453   | -1.17 | 0.266 |
| <b>X47</b>         | -3.276  | 3.526   | -0.93 | 0.371 |
| <b>X48</b>         | -3.271  | 8.157   | -0.4  | 0.695 |

**Table 4:** Results obtained after considering experience as dependent variable.

| Predictor constant | Coef   | SE Coef | T     | P     |
|--------------------|--------|---------|-------|-------|
| X1                 | 8.036  | 2.347   | 3.42  | 0.005 |
| X2                 | 28.05  | 57.41   | 0.49  | 0.634 |
| X3                 | 193.4  | 147.4   | 1.31  | 0.214 |
| X4                 | -176.8 | 141.8   | -1.25 | 0.236 |
| X5                 | -318.4 | 294.9   | -1.08 | 0.302 |
| X6                 | 317.1  | 278.2   | 1.14  | 0.277 |
| X7                 | -17.01 | 58.48   | -0.29 | 0.776 |
| X9                 | 12.22  | 32.22   | 0.38  | 0.711 |
| X10                | -12.31 | 40.23   | -0.31 | 0.765 |
| X11                | -16.88 | 23.31   | -0.72 | 0.483 |
| X13                | 2.12   | 23.75   | 0.09  | 0.93  |
| X15                | -3.73  | 27.18   | -0.14 | 0.893 |
| X17                | -114.2 | 150.5   | -0.76 | 0.462 |
| X18                | 196.3  | 156.5   | 1.25  | 0.234 |
| X19                | 40.1   | 159.7   | 0.25  | 0.806 |
| X20                | -56.5  | 168.9   | -0.33 | 0.744 |
| X21                | -81.25 | 44.12   | -1.84 | 0.09  |
| X23                | -38.44 | 63.45   | -0.61 | 0.556 |
| X25                | 6.56   | 22.45   | 0.29  | 0.775 |
| X26                | -10.87 | 22.49   | -0.48 | 0.638 |
| X27                | 0.08   | 28.59   | 0     | 0.998 |
| X28                | 3.82   | 31.6    | 0.12  | 0.906 |
| X29                | 9.68   | 18.36   | 0.53  | 0.608 |
| X31                | -1.23  | 14.74   | -0.08 | 0.935 |
| X33                | -65.1  | 77.05   | -0.84 | 0.415 |
| X34                | 61.37  | 94.83   | 0.65  | 0.53  |
| X35                | 41.39  | 71.57   | 0.58  | 0.574 |
| X36                | -57.88 | 78.75   | -0.74 | 0.476 |
| X37                | 5.75   | 38.86   | 0.15  | 0.885 |
| X39                | 68.49  | 66.74   | 1.03  | 0.325 |
| X41                | 293.6  | 268     | 1.1   | 0.295 |
| X42                | -314.3 | 267.3   | -1.18 | 0.262 |
| X43                | -301.4 | 261     | -1.15 | 0.271 |
| X44                | 318.1  | 269.8   | 1.18  | 0.261 |
| X45                | -11.27 | 33.94   | -0.33 | 0.746 |
| X47                | 3.89   | 33.27   | 0.12  | 0.909 |
| X48                | 6.56   | 22.45   | 0.29  | 0.775 |

**Table 5:** Results obtained after considering rating as dependent variable.

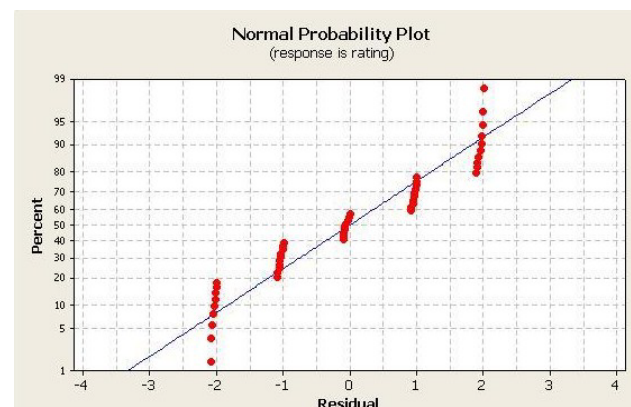
Figures 1 and 2, it can be clearly seen that there is a considerable impact of experience on the students' behavioral pattern rather than the impact of the faculty having rated as good for a particular semester or a course.

## Conclusion

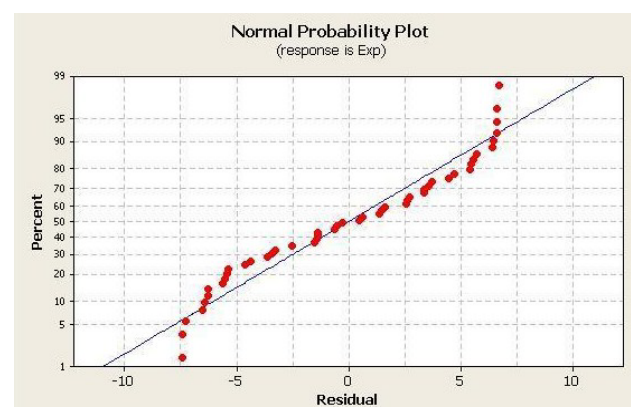
This paper presents the process of identifying different student clusters in a pool of students concentrating on different goals for their future, dynamically classifying a new student to one of the predefined clusters based on his behavioral pattern and then using regression to retrieve the effect of faculty experience and rating on students browsing behavior. Using regression analysis we concluded that the experience of the faculty impacts more on student's behavior.

## References

1. El-Halees A (2008) Mining students data to analyze learning behavior: A case study. Department of Computer Science, Islamic University of Gaza, Gaza, Palestine.
2. Steinberg DM (2010) Towards autonomous habitat classification using



**Figure 1:** The result on the multiple linear regression analysis by taking independent variables (1 to 48) and considering the rating and experience as dependent variables, we have the regression lines as and are represented below.



**Figure 2:** The result on the multiple linear regression analysis by taking independent variables (1 to 48) and considering the experience as dependent variables, we have the regression lines as and are represented below.

Gaussian Mixture Models. IEEE/RSJ International Conference on Intelligent Robots and Systems. Taipei, Taiwan.

3. Thury EM (1998) Analysis of student web browsing behavior: Implications for designing and evaluating web sites. SIGDOC 265-270.
4. Kularbphetong K, Tongsiri C (2014) Mining educational data to support students' major selection. International Journal of Computer, Information, Systems and Control Engineering 8: 1.
5. Talavera L, Gaudioso E (2004) Mining student data to characterize similar behavior groups in unstructured collaboration spaces. In Proc European Conf AI 17-33.
6. Abbas OA (2008) Comparisons between Data Clustering Algorithms. International Arab Journal of Information Technology 5: 320-325.
7. Srimani PK, Patil MM (2013) Linear Regression Model for Edu-mining in TES. Int J Conceptions on Elec and Electronics Engineering 1: 45-49.
8. Torres SD (2014) Analysis of search and browsing behavior of young users on the web. ACM Transactions on Embedded Computing Systems 8: 7.