

# Exploring Web Mining for Auditing Social Networking: A Clustering Approach

V.S. Kumbhar<sup>1</sup>, K.S. Oza<sup>2</sup> and R.K. Kamat<sup>3</sup>  
<sup>1&2</sup> Assistant Professor and <sup>3</sup>Professor

Department of Computer Science, Shivaji University, Kolhapur

<sup>1</sup>vsk811@gmail.com

<sup>2</sup>skavita.oza@gmail.com

<sup>3</sup>rkk\_eln@unishivaji.ac.in

## Abstract:

In the quest for excellence, the institutes of higher learning are revitalizing their institutional websites so as to persuade the stakeholders. With the growth of social networking, the aforesaid task however is not as easy as said and done. Auditing the web pages in the wake of their popularity on the well known social networking platforms such as facebook could be an appropriate methodology for continuous improvement. In light of this the paper put forth the metrics for analysis followed by carving the database using a web based online tool. Results are indicative enough to showcase usefulness of the approach adopted herein for improving the websites.

## 1. Introduction

With the upsurge of internet and web there has been sea change as regards to its applications in various domains. Social networks the most innovative arm of the web is all poised to occupy the societal life of the entire globe. Social network sites (SNSs) are increasingly attracting the attention of academic and industry researchers intrigued by their affordances and reach[1]. While still a relatively new online phenomenon, social networking has already attracted attention from major Internet corporations. Microsoft, Google, Yahoo! and AOL are among companies offering online community services [6]. Apart from the for profit players, the social networking is also in demand for the nonprofit organizations. Many institutes of higher learning are making use of this innovative concept to get their stakeholders united and disseminate their intended information to their stakeholders.

In one of the papers [4] have done extensive analysis based on the user behavior and interviews presents the main motivation behind using the social networking. It is revealed that the professionals use internal social networking to build stronger bonds with their weak ties and to reach out to employees they do not know. Their motivations in doing this include connecting on a personal level with coworkers, advancing their career with the company,

and campaigning for their projects. As far as higher education is concerned, the social media has already made its place in this ever emerging paradigm of knowledge. Indeed, many technological commentators have now moved on to enthuse about the next set of 'next big things' within the world of digital technology—such as the semantic web, cloud computing and the internet of things. In this sense, there is now room for the higher education community itself to assume a greater role in shaping the development of social media on the ground in higher education settings. After all, social media technology is something that is supposed to be created by its users—higher education institutions and educators included [2].

Maintaining a social media presence however can be a full-time job, which makes it kind of difficult if your full-time job requires your time and attention be spent elsewhere. Social media profiles can fall into disrepair quickly when left alone. An audit can help get things back on track [5]. Since social networking sites, such as MySpace and Facebook, began allowing organizations to create profiles and become active members, organizations have started incorporating these strategies into their public relations programming [7]. For all the above said instances an extensive audit of these sites using Clustering theory is need of an hour. Taking this thread for further exploration, the present paper has developed a notion of application of Clustering theory for auditing the facebook pages of the institutes of higher learning.

The paper is structured in several sections. After introducing the concept, the test bed development has been put forth. This is followed by application of the Clustering theory followed by results and conclusions.

## 2. Test bed Development

This paper has taken the test bed developed through the social networking analysis tool likealyzer [3]. Data from the educational domains has been taken to form the data base which is the basic subject for clustering. The database is shown in table 1.

**Table 1: Data base generated through the Likealyzer.**

#	Page	Likes	PTAT	ER	LikeRank
1	Onewbut Education	422	6	1.42%	38
2	Kinesiology Training Education	6,618	333	5.03%	80
3	Cal Poly Parents Education	3,569	3	0.08%	19
4	Pilot Flyskole Education	8,875	266	3.0%	81
5	Formaterapia Lda Education	16,172	431	2.67%	61
6	Root Education	1,432	282	19.69%	77
7	MEFA Befestigungs- und Montagesysteme GmbH Education	101	1	0.99%	25
8	Nevermore - Conectando Conhecimento Education	4,462	354	7.93%	81
9	Unidad de Bienestar Universitario Education	892	108	12.11%	78
10	Instituto CEA Education	31,946	229	0.72%	66
11	UNAQ Education	2,896	327	11.29%	77
12	OC&C Escuela de Marketing Digital Education	611	18	2.95%	64
13	Aula CM Education	3,467	230	6.63%	78
14	Waynflote School Education	1,393	78	5.6%	64
15	Escuela Europea de Gerencia Education	1,235	0	0.0%	27
16	Pi Creation Education	1,355	279	20.59%	86
17	Nido La Casa Amarilla Education	12,376	683	5.52%	86
18	Nido Kangurito Education	8,679	478	5.51%	74

#	Page	Likes	PTAT	ER	LikeRank
19	Bee Bee Cuna Jardín Education	3,278	129	3.94%	65
20	Orange training center Education	41,872	2,346	5.6%	74
21	University of Leicester Education	25,952	244	0.94%	55
22	Mokyklų tobulinimo centras Education	1,138	16	1.41%	42
23	PONTIFICIA UNIVERSIDAD CATOLICA DEL PERU Education	340,472	4,980	1.46%	67
24	Nido Zig Zag Education	5,531	853	15.42%	81
25	Renkuosi mokyti Education	2,183	138	6.32%	81
26	Premier Sport - Suffolk Education	326	12	3.68%	57
27	Aranäs gymnasiet Kungsbacka Education	1,552	47	3.03%	68
28	Instituto Politecnico Santa Cruz Education	4,020	20	0.5%	42
29	Studio de Danzas Aswan Education	1,130	30	2.65%	75
30	Tadpole Training Ltd Education	679	77	11.34%	68
31	AEPS - Aviation English Private School Education	152	3	1.97%	36
32	Centro Studi Raffaello Education	1,759	5	0.28%	34
33	Waterford English and Training Academy Education	1,464	31	2.12%	45
34	Teknikspränget Education	5,050	32	0.63%	60
35	Berghs School of Communication Education	10,338	324	3.13%	73
36	EF Finland - Kielikurssit Education	9,063	101	1.11%	49
37	Explorius Finland Education	1,911	51	2.67%	65
38	STS vaihto-oppilaat Education	1,155	25	2.16%	57
39	ESCP Europe Paris Campus Education	4,943	155	3.14%	63

#	Page	Likes	PTAT	ER	LikeRank
40	ISC Paris Education	8,316	66	0.79%	48

Clustering theory has been applied to this Education institute facebook page dataset. Same has been depicted in the following section.

**3. Clustering theory for website audit**

The main intent of the clustered dataset is to enable auditing the web pages shine in table 1 for further improvement. The data set is result of likealyser tool applied on facebook pages. Metric used in the dataset is related to the facebook page of the respective education institute.

Parameters used in the metric are:

**Likes:** It gives the count of people Clicking Like below a post on Facebook and its is an easy way to let people know that others have enjoyed it without leaving a comment

**PTAT (People Talking About This):** It measures three types of user actions: Likes, shares and comments (but does not include clicking on the post). It is to count number of facebook fans who did something to show engagement to their friends.

**ER (Engagement Ratio):**It is the average percentage of users who have seen the post and then either liked, commented, shared or clicked on the post.

**LikeRank:** It's a number between 1 to 100, which represents how likeable the brand is. Here it is educational institute. Better the performance (Facebook page performance) higher will be the LikeRank.

Above dataset is clustered into three categories ie. Cluster1, Cluster2, and Cluster3. An open source tool for mining, Weka is used for analysis of data.

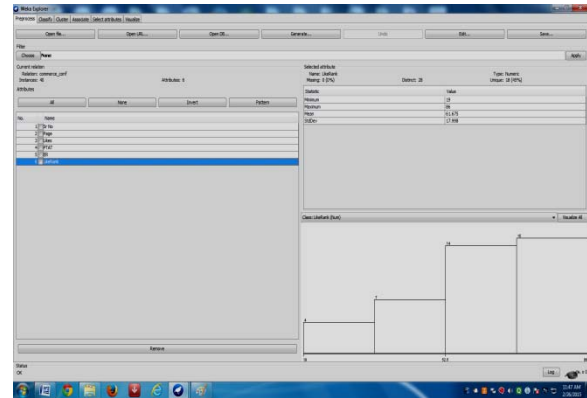


Fig. 1 Screenshot showing Weka interface.

Data is analyzed by clustering using simple k-means algorithm.

Three distinct clusters are generated as shown in table 2.

**Table 2: Clustering of data**

Cluster_no	No. of instance
C1	18
C2	11
C3	11

Output of cluster algorithm applied on dataset using Weka is shown in fig. 2.

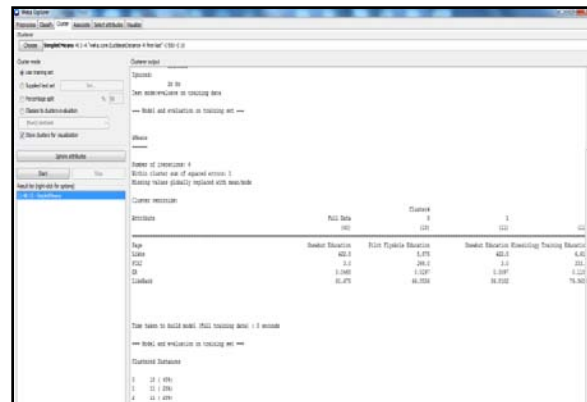


Fig. 2 Screenshot revealing cluster output

#### 4. Summary and Conclusion

Clustering the facebook pages using K-means enables their categorization in three classes viz. C3: Excellent web pages, C2: Poor webpages, C1: Moderately good web pages. The like ranks for C3, C2 and C1 are 68 to 86, 19-49 and 55-81 respectively. Another metrics used in the auditing process is number of likes which ranges from 679 – 12376, 1012 – 9063 and 326 – 3404 for C3, C2 and C1 respectively. PTAT ranges from 77 – 853, 0 – 101 and 12 – 478 for C3, C2 and C1 respectively. Yet another parameter ER ranges from 0.0503 – 0.2059, 0 – 0.212 and 0.0063 – 0.056 for C3, C2 and C1 respectively. All the values agree well to the clustered categories. Through this auditing approach the institutes of higher learning can follow the best practices adopted by those performing well and improve the institute websites which now a days is serving as an trait of quality.

#### 5. References

- 1) Boyd, D., & Ellison, N. (2010). Social network sites: definition, history, and scholarship. *IEEE Engineering Management Review*, 38(3), 16-31. doi:10.1109/emr.2010.5559139
- 2) Fenwick, T. (2014). Social media, professionalism and higher education: a sociomaterial consideration. *Studies In Higher Education*, 1-14. doi:10.1080/03075079.2014.942275
- 3) LikeAlyzer,. (2015). *Analyze your Facebook page - LikeAlyzer*. Retrieved 25 February 2015, from <http://likealyzer.com/>
- 4) Omar, A., Rashid, W., & Majid, A. (2014). Motivations Using Social Networking Sites on Quality Work Life. *Procedia - Social And Behavioral Sciences*, 130, 524-531. doi:10.1016/j.sbspro.2014.04.061
- 5) Social,. (2014). *The 15-Minute Social Media Audit Everyone Can Do*. Retrieved 25 February 2015, from <https://blog.bufferapp.com/social-media-audit>
- 6) Trusov, M., Bucklin, R., & Pauwels, K. Effects of Word-of-Mouth versus Traditional Marketing: Findings from an Internet Social Networking Site. *SSRN Journal*. doi:10.2139/ssrn.1129351
- 7) Waters, R., Burnett, E., Lamm, A., & Lucas, J. (2009). Engaging stakeholders through social networking: How nonprofit organizations are using Facebook. *Public Relations Review*, 35(2), 102-106. doi:10.1016/j.pubrev.2009.01.006