International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 4 Issue: 4
ISSN: 2454-4248
831 – 834

Predictive Analysis of Facebook using WEKA

Raghvendra Singh
Saroj Institute of Technology and Management, Lucknow
raghvendra711@gmail.com,

Veenu Yadav
Saroj Institute of Technology and Management, Lucknow
yadavveenu402@gmail.com

Dilip Kumar
Kamla Nehru Institute of Technology, Sultanpur
dilip1987kumar@gmail.com

Abstract : Web-based social networks have become more prevalent as a medium for connecting like-minded people. The public accessibility of such social networks with the capability to share information, thoughts, opinions, and experience offers great potential to mankind and organizations. Social network has gained amazing attention in the last decade. Accessing social network sites such as Facebook, Google+, Twitter and LinkedIn through the internet and the web 2.0 technologies has become more affordable. Facebook is social networking service on which after registering on the the site, users can create their profile, add other users as friends, interchange messages, post status updates, photos, and share videos etc. People are more interested in and relying on Facebook for information, news and opinion of other users on various subject matters. Based on the data available for the facebook, the number of profiles has increasing expressively but with the fast growth of users, fake profiles/users have also grown.
The WEKA data mining tool was used by performing adjustments of the attributes in order to come up with a decisive output. This paper presents the comprehensive review of social network and the trustworthiness of social networks.

Keywords: *Social Networking, Facebook, WEKA.*

*****

I. INTRODUCTION

Social network is a term used to depict online administrations that enable people to make a public/semi-public profile inside a domain to such an extent that they can informatively associate with different users inside the network [1]. Social network has enhanced the idea and innovation of Web 2.0, by empowering the arrangement and trade of User-Generated Content [2]. Basically, Social network is a graph comprising of nodes and links used to speak to social relations on Social network destinations [3].

Since social networks are sorted out around the general population who utilize them, believing the content which is engendered in them is exclusively reliant on the assurance capacity of the clients. In the event that the clients don't believe the data then he/she won't spread it.

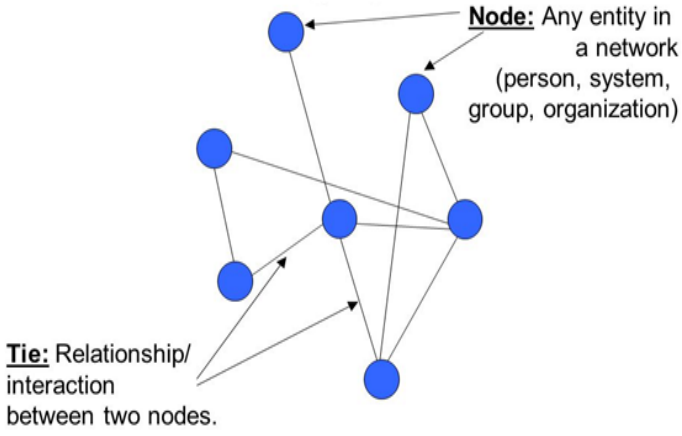


Fig.1 Social Networking concept with links and nodes

There many social networking sites available like Twitter, Facebook, Google +, Youtube, LinkedIn .

II. LITERATURE SURVEY

As it is unmistakably portrayed in the last sentence of the past passage, a great many people trust others since they had encountered trustworthiness from them in their prior

 ISSN: 2454-4248
831 – 834

connection. So we can utilize this factor for displaying of trust in this study, on the grounds that in social network sites this factor affect confiding in a substance which is shared by individuals who have as of now get a believably due to their past posts quality. In social network sites the most imperative elements for building trust are notoriety and impact. When we say notoriety in web-based social networking it implies the way you are seen by others exclusively in view of your posts.

A trust demonstrates a constructive confidence in someone else, or content in this specific case. Standard clients will probably confide in individuals who share data which is exclusively in light of established truths, as by connecting the connections identified with the substance they share, which will doubtlessly expand the validity of the data they share. Such as “Propagation Models for Trust and Distrust in Social Networks” by Cai-Nicolas Ziegler and Georg Lausen [4],recommends a model for both trust and distrust in social networks.

III. DATA MINING Techniques

A. Classification

This research utilizes classification techniques for foreseeing trust. The three sorts of classification methods that were utilized to build prediction models are Decision Tree, J48 and Bayesian(Naïve Bayes) Classifiers.

i. Decision Trees

A decision tree is an data mining system that creates a graphical representation and investigation of the model it produces.

decision trees are generally utilized for classification purpose; they can be utilized likewise for various types of regression analysis.

ii. J48 Classifier Algorithm

J48 is an execution of the notable C4.5 calculation for delivering either pruned or unpruned C4.5 tree. The C4.5 algorithm was assembled in view of the idea of data getting or entropy lessening to choose the most proficient split.

iii. Naive Bayes

Naïve Bayes classifier works under the presumption of that the nearness of a particular feature of a class have no relationship to the nearness of some other constituent.

B. Clustering

This clustering technique is used to dividing data items into collection/groups based on some resemblance called clusters like we are having several type of data in our one folder, we are arranging them in several folder based on resemblance like audio in audio folder and text data saved in text folder, etc with the intention of user can easily access the data rendering to their need.

IV. SOURCE OF DATA

The wellspring of data for this research is my own data set, which is acquired by utilizing a questionnaire and center gathering to gather data. The questionnaire was gathered data since it makes it is less demanding to appropriate to the same number of individuals as you need, be that as it may, it is very hard to get a detailed analysis by utilizing only the data which is gathered by questionnaire . Therefore, we chose to utilize the focus group method to supplement the data we get from the questionnaire by talking about with individuals who have information technology back ground and really great specialized know-how of the exploration research. Before beginning to compose the inquiries which were utilized as a part of the questionnaire we made broad research by perusing articles identified with the point of our task, specifically about "trust".

V. IMPLEMENTATION

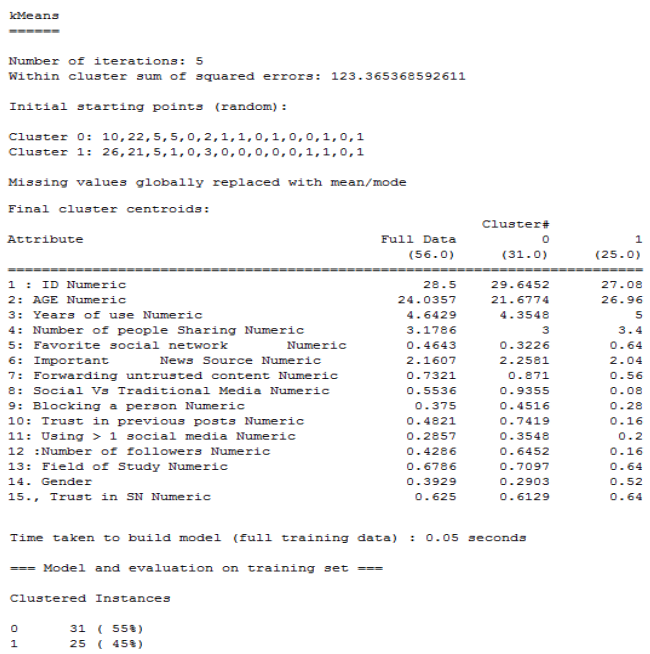
This research presents data mining techniques can be utilized effectively to demonstrate and predict trust. The result of this research can be utilized to assist individuals with making more reliable expectation of trust to social media content.

Eleven tests were done altogether for this examination.

This experiment was performed for K=2, with default estimations of seed and distance function. Each one of the last picked 14 attributes and 56 records were utilized as a part of this trail. To cluster the records as per their qualities this model was prepared by utilizing the default estimations of the K-Means algorithm. The table underneath demonstrates the result of the trail and cluster distribution of the data set.

Fig.2 Run Data-Instances and Attributes

 ISSN: 2454-4248
831 – 834



```
kMeans
======

Number of iterations: 5
Within cluster sum of squared errors: 123.365368592611

Initial starting points (random):

Cluster 0: 10,22,5,5,0,2,1,1,0,1,0,0,1,0,1
Cluster 1: 26,21,5,1,0,3,0,0,0,0,0,1,1,0,1

Missing values globally replaced with mean/mode

Final cluster centroids:
                                                 Cluster#
Attribute                                Full Data       0          1
                                          (56.0)    (31.0)     (25.0)
=============================================================================
1 : ID Numeric                              28.5   29.6452    27.08
2: AGE Numeric                           24.0357   21.6774    26.96
3: Years of use Numeric                   4.6429    4.3548        5
4: Number of people Sharing Numeric       3.1786         3      3.4
5: Favorite social network       Numeric  0.4643    0.3226     0.64
6: Important      News Source Numeric     2.1607    2.2581     2.04
7: Forwarding untrusted content Numeric   0.7321     0.871     0.56
8: Social Vs Traditional Media Numeric    0.5536    0.9355     0.08
9: Blocking a person Numeric               0.375    0.4516     0.28
10: Trust in previous posts Numeric       0.4821    0.7419     0.16
11: Using > 1 social media Numeric        0.2857    0.3548      0.2
12 :Number of followers Numeric           0.4286    0.6452     0.16
13: Field of Study Numeric                0.6786    0.7097     0.64
14. Gender                                0.3929    0.2903     0.52
15., Trust in SN Numeric                   0.625    0.6129     0.64


Time taken to build model (full training data) : 0.05 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       31 ( 55%)
1       25 ( 45%)
```

Fig.3 Clustering output experiment

TABLE 1

The values of parameters used for experiment

Cluster output/result				
Distance Function	Seed Value	K	Cluster Distribution	
Euclidean Distance	10	2	C0	C1
			31(55%)	25(45%)

As per the above *Table 1*, we observe that the first trail/experiment was performed with default values of the algorithm (Euclidean distance, K = 2 and Seed Value= 10).

The result in Fig. 4 demonstrates to us the togetherness of the clusters, "1" implies every one of them in that cluster share precisely the same of one, and a "0" implies every one of them in that cluster has an estimation of zero for that specific attribute. Alternate numbers are generally the normal incentive inside in the cluster. Singular cluster shows a kind of conduct in our members, in view of which we can begin to reach a few conclusion.

VI. CONCLUSION

The data set utilized as a part of this research was accumulated from our own particular survey, which was arranged exclusively to collect data that can be utilized as a part of this study. After the data was gathered, it was preprocessed and arranged in a route reasonable for the data mining tasks. At that point the objective was completed in two sub stages, first the cluster modeling which at that point took after by classification modeling.

REFERENCES

[1] Chen, Z. S., Kalashnikov, D. V. and Mehrotra, S. Exploiting context analysis for combining multiple entity resolution systems. In Proceedings of the 2009 ACM International Conference on Management of Data (SIGMOD'09), 2009.

[2] Kaplan, A.M. and Haenlein, M.: Users of the world unite! The challenges and opportunities of social media. Science direct, 53, 59-68, 2010.

[3] Borgatti, S P.: "2-Mode concepts in social network analysis." Encyclopedia of Complexity and System Science, 8279-8291, 2009.

[4] Cai-Nicolas Ziegler and Georg Lausen (2005) "PropagationModelsforTrustandDistrusting SocialNetworks

[5] Carrington P. J., Scott J., and Wasserman S., 2005. "Models nd Methods in Social Network Analysis." Cambridge University Press, New York, 2005.

[6] Liu Y., Yau S., Peng D., and Yin Y., 2008. "A Flexible Trust Model for Distributed Service Infrastructures." In Proceedings of the 2008 11th IEEE Symposium on Object Oriented RealTime Distributed Computing, Orlando, USA, 108-115.

[7] AzraShamim, VimalaBalakrishnan, MadihaKazmi, and ZunairaSattar, "Intelligent Data Mining in Autonomous Heterogeneous Distributed and Dynamic Data Sources", 2nd International Conference on Innovations in Engineering and Technology (ICCET'2014) Sept. 19-20, 2014.

[8] Rumi Ghosh, SitaramAsur, "Mining Information from Heterogeneous Sources: A Topic Modeling Approach" ACM 978-1-4503-2321,2013.

[9] Prakash R.Andhale1 , S.M.Rokade2, "A Decisive Mining for Heterogeneous Data", International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 12,pp. 43-437, December 2015.

[10] Amir Ahmad, Lipika De, "A k-mean clustering algorithm for mixed numeric and categorical data" Data & Knowledge Engineering Elsevier,pp. 503-527,2007.

[11] Dr. Goutam Chakra Borty,Murali Krishna Pagolu, Analysis of Unstrucured Data: Application of Text Analytics and Sentiment Mining,2014.

[12] http://www.definitions.net/definition/trust

[13] http://changingminds.org/explanations/trust/what_is_trust.htm

[14] Ming-Syan Chen, Jiawei Han, and Philip S.Yu, "Data Mining – An Overview from Database Perspective",Knowledge and Data Engineering, IEEE

 ISSN: 2454-4248
831 – 834

Transactions on ,Volume 8 , No.6 , pp 866-883,Dec 1996.

[15] Nicholas J Belkin and W Bruce Croft,“Retrieval techniques”, Annual Review of Information Science and Technology,Volume 22,pp 109-45,Information Today,1987.

[16] Romero, Cristobal, Sebastián Ventura, and Paul De Bra, "Knowledge discovery with genetic programming for providing feedback to courseware authors." Volume 14,Issue 5, pp 425-464, 2004.

[17] Al Jarullah, A.A., "Decision tree discovery for the diagnosis of type II diabetes," Innovations in Information Technology (IIT), 2011 International Conference on , vol., no., pp.303,307, 25-27 April 2011

[18] Quinlan J (1993) Programs for Machine Learning Morgan Kaufmann Sait

[19] Bhargavi, P, &Jyothi, S. (2009). Applying Naive Bayes data mining technique for classification of agricultural land soils. International journal of computer science and network security, 9(8), 117-122.