

Mining on Social Media Data: To Determine the Personality of Unrevealed Person



Deepak. V, M. Bhargava Teja, V. Sravani, K. Venkata Raghava

Abstract: *The unavoidable utilization of online networking like Facebook is giving exceptional measures of social information. Information mining methods have been broadly used to separate learning from such information. The character of the person is predicted whether he is good or not by using data mining techniques from user self-made data. Mining methods are being broadly using to separate learning from such information, main examples for them are network discovery and slant investigation. Notwithstanding, there is still a lot of room to investigate as far as the occasion information (i.e., occasions with timestamps, for example, posting an inquiry, altering an article in Wikipedia, and remarking on a tweet. These occasions react users' personal conduct standards and working forms in the social media websites.*

Keywords: *Data mining, social media, Online networking.*

I. INTRODUCTION

In the recent times, the world is changing into an internet-dependent world where unlimited raw data is generated and stored in the unstructured format. All those types of data can be analysed using the data mining and text mining and decision-making process is made easier. Text mining includes a lot of tasks such as document clustering, document classification by their types, text summarization, sentiment analysis, social network analysis and decision making, topic detection according to category, web page classification on different domains such as education based, social media based, scientific based, identification of author or the creator of the content, plagiarism detection, spam/malware analysis, patent analysis, financial decision making, sports related data mining etc. The basic challenge in any kind of mining is all kind raw data will be already be available, but all these data will be in structured format. The challenge is how it must be converted from structured data to structured information. Online networking sites is defined as a gathering of Internet-put together applications that work with respect to the ideological and innovative establishments of Web 2.0. It enables individuals to make, share, and additionally share data and thoughts [1][7][10].

Some outstanding social media websites are Facebook, Instagram, SnapChat, Wikipedia and Twitter. The ascent of web-based social networking has created extraordinary measures on the social information. Figure 1 explains that how much information are made in one moment in various internet-based life stages. For example, Facebook, which is the most dynamic of informal communities with over a billion months to month dynamic clients, makes the most measure of social information: clients post over 4.1 million for every moment. Instagram, with 300 million months to month clients in 2017, and secondly of photographs over 1.8 million likes every minute. The reason to choose the prediction of person character among all the social media websites because Facebook is the world best dominating site as shown in the figure 1[2][5][6].

The rest of the paper is structured as follow :

Section 2 : Data pre-processing on the dataset

Section 3 : Implement methodology

Section 4 : Analysis on the result

Section 5 : Conclusion

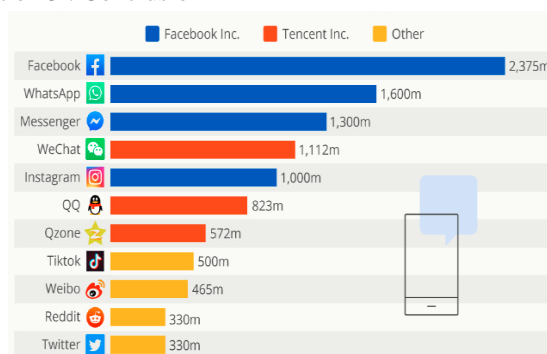


Figure 1: Facebook Dominates Social Media Landscape.

II. DATA PRE-PROCESSING

A. PRE-PROCESSING

The main objective of the here is to explore the power of data mining to extract the most relevant knowledge. The initial stage is to classify or cluster the most relevant data together and commonly name it so that it can be referred in the operations, in this case it be may grouping of similar users undertaking into the consideration of various aspects. The pre-processing stages include the document decay and document representation. In the digital world the pre-processing plays an important role because the huge amount raw data may not be useful until it is processed. According a study it can also be done using robotic data.[11]

Manuscript published on November 30, 2019.

* Correspondence Author

Deepak. V*, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh – 522502, India.

M. Bhargava Teja, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh – 522502, India.

V. Sravani, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh – 522502, India.

K. Venkata Raghava, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh – 522502, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Mining on Social Media Data: To Determine the Personality of Unrevealed Person

Data Pre-processing is a method that is utilized to change over the raw data into a data set (.csv). As such, at whatever point the information is assembled from various sources it is gathered in raw format we cannot import directly to the data mining so first we need to analyse the data [5][8]. Pre-processing include several techniques like cleaning, integration, transformation, fill in missing values(empty record with no specified value in it), smooth noisy data(to find the main function among the records), identify or remove outliers which affects the accuracy of the result, and resolve inconsistencies and reduction. This paper shows a detailed description of data pre-processing techniques which are used for data mining.

B. FACEBOOK USER SELF MADE DATA SET

J	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	userid	age	dob_year	gender	tenure	mutual_friends	friend_count	friendships_initiated	likes	likes_received	video_likes	photo_likes	videos	photos	Result
2	1697267	7	2012	female	574	0	70	50	19	23	19	4	1	86	Bad
3	1587428	23	1996	male	167	0	69	52	19	47	19	28	1	48	Bad
4	1975229	19	2000	male	269	3	70	42	20	5	3	2	13	27	Good
5	1736475	20	1999	male	495	4	70	46	20	14	3	11	26	86	Good
6	1726327	20	1999	female	707	0	69	28	20	68	10	18	23	14	Bad
7	1675796	18	2001	male	22	0	70	68	20	21	20	12	26	27	Bad
8	1623209	17	2002	female	286	7	70	59	20	36	20	16	25	83	Good
9	1895042	16	2003	female	50	9	69	46	20	43	20	13	22	5	Good
10	2028152	23	1996	male	303	20	70	34	20	27	20	7	10	43	Good
11	1426973	12	2007	male	919	15	70	44	20	56	20	36	2	37	Good
12	1996752	16	2003	female	84	14	69	47	21	44	21	23	26	65	Good
13	1664395	8	2011	male	186	11	69	33	22	29	22	7	16	23	Good
14	2148754	9	2010	female	484	0	70	22	22	139	44	95	16	56	Bad
15	1719388	9	2010	male	450	0	70	45	22	26	19	7	8	5	Bad
16	1138929	15	2004	male	478	3	69	55	22	25	20	5	4	30	Good
17	1748854	20	1999	male	1930	7	69	25	22	30	21	9	31	9	Good
18	1194071	22	1997	male	91	8	69	44	22	24	22	2	16	13	Good
19	1137596	19	2000	male	264	7	69	66	22	23	22	1	16	74	Good
20	1109825	12	2007	male	420	5	70	51	22	24	22	2	30	99	Good
21	1162760	12	2007	male	276	0	69	50	22	80	22	16	18	19	Bad
22	1838267	16	2003	male	619	0	69	34	23	22	12	10	41	80	Bad
23	1975785	15	2004	female	100	0	70	55	23	27	22	5	29	82	Bad
24	1906630	15	2004	female	568	0	70	29	23	25	23	2	32	20	Bad
25	2157700	14	2005	male	995	0	69	26	23	24	23	1	29	45	Bad

Figure 2: Sample data of 25 users of Facebook

We had made an informational collection of Facebook users of 344 examples in that we have given an example picture as appeared in figure1. There are 15 properties in figure 2 with userid, age, dob_year, gender, tenure, mutual_friends, friend_count, friendship_initiated, likes, likes_received, video_likes, photo_likes, videos, photos lastly result as shown in the figure 3[2][10]. For the making of this informational index, we have made utilizing Microsoft excel(paid) and we can likewise utilize WPS spreadsheets which are open-source applications [4]. Constraints for the informational collection are user id(it is a consistent substance used to distinguish a user on a product, framework, site or inside any conventional IT condition) ought to be a 7 digit number, date of birth of the user ought to be in the scope of 1900-2019, by utilizing year of birth we had determined the age(the period of time that an individual has lived or a thing has existed) of user, sexual orientation of the user of will be of two decisions male or female, mutual_friends (the individuals who are Facebook friends with both you and the individual whose profile you're seeing) can be of any number yet ought to be not exactly friend_count and who have zero common companions then the information digging is occurred for example, in the event that you're friends with Chris, and Mark is friends with Chris, at that point Chris will appear as a mutual friend when you're survey Mark's profile., friend_count means absolute number of companions for the user, friends_initiated implies the number of companions whom you have sent companion demand, likes intends to

what number of posts do you preferred for different users and clicking like below a post on Facebook is an approach to tell individuals that you appreciate it without leaving a comment, likes_received means all out preferences from the posts of you, for videos(the recording, repeating, or broadcasting of moving visual pictures) and photos(It is utilized as a prefix is in the word photograph, which is a picture taken with the assistance of light) independently there are 2 segments videos_likes and photo_likes and in conclusion at long last outcome to anticipate the user is good/bad.

III. IMPLEMENTED METHODOLOGY

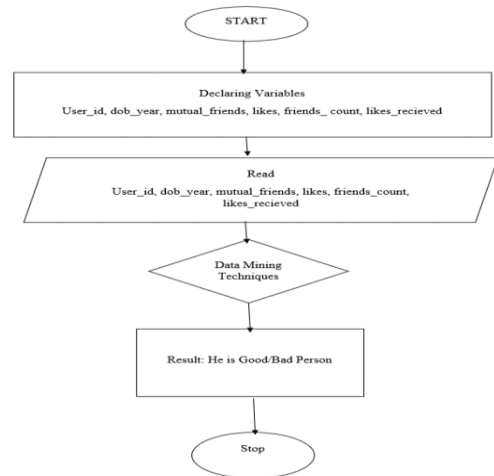


Figure 3: Flow Chart for the Character prediction of Unknown person

In this area we examine an informational collection produced by Facebook and illustrate some found fascinating users' personal conduct standards. Figure 4 shows some common place occasions are present in the form of "Action log" in Facebook. Even more accurately, the figure 4 indicates the kinds of occasions while the circular icons exhibit reactions for the posts of him or his friend. For example, "Posts" cor-reacts to an occasion type and the thing "Bhargava Teja reacted for the post" (the subsequent row in the blue square is also same) relates to occasion of the "Posts". For every occasion it has a date showing when the occasion occurs. For case, the date in the white square shows that the occasion "Bhargava Teja has accepted the Friend Request with Praveen Kumar, Raviteja Chowdary and Yaswanth Krishna" on Sept 2019, Oct 2019.

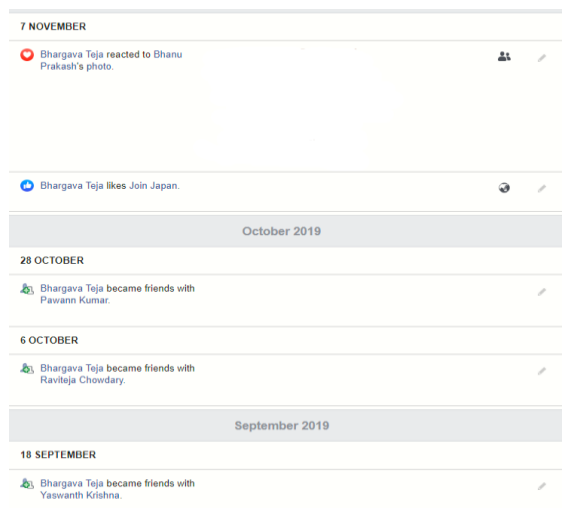


Figure 4: A Sample Example on a person’s Activity Log in Facebook.

Table 1: Recorded files of the person in Activity Log.

Name in Activity Log	Example (with reference to Bhargava Teja)
Others_People_posts_reactions	Reacted for the post of Join Japan and Bhanu prakash photo.
Accepted Friend Requests	Became friends with Ravi chowdary, pavan kumar and yashwanth krishna.

It is conceivable to download "Activity log" from Facebook which comprises of a lot of HTML records. Table1 presents two model records. For e.g., posts which you kept in Facebook records in form a module name called "Your_Posts", for e.g., commenting on a post, sharing/tagging a connection among friends and adding a latest user’s photo with another photos[1][3].An "Example of document content" segment shows one case of the substance of each and every document. Occasions of the user can have a separate module from these documents, since date and time are remembered for every event of the posts. For example, a "send demand" occasion occurred at Sept 2019 can be separated dependent on the substance model in the "view_sent_friend_requests" record above the "friend_suggestions". In view of an occasion log removed from the information appeared in table 1[9], a data mining technique can be used to determine the posts of the user as appeared in Figure 4. This model depicts about intriguing conduct design, which is represented as pursues. Most unwanted, unusual, very strange which implies to an individual that he is not the friend of the user, however, he may be the friend of his friend of the user. Later the outsider posted in his account and user’s companion remarks like comment or a reaction on it (demonstrated as shown in figure k1), the posted comment of his companion will be shown in the user’s "News Feed". Suppose the user has seen the post, that person can answer it (as showed in k2). Subsequent to answering about post, the user will be on long run sends a companion suggestion to an outsider and he/she may have a chance of being companions (shown by k3). The founded example can be separated of a user’s movement log

(which is appeared in Figure 4), that is nothing but whether this user will be in general companion demands to the individuals whose information is remarked by the user’s present companions.

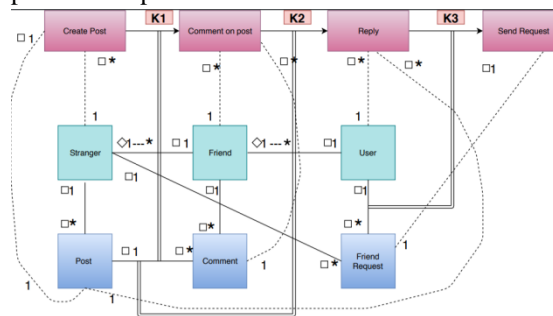


Figure 5: Behaviour Pattern of the User while making friends. [1]

This personal conduct standard may be important for Facebook. For example, assume that there is motivation to associate this user to any of his/her companion's companions, at that point Facebook can beginning indicating in his/her news feed of user’s posts. Also,companions are present or not for the user presently and, as result, Facebook turns into a progressively dynamic online networking stage. In addition, the companions filling in as extensions (interfacing the user to outsiders) affect the FIGURE 5. Two other standards of conduct found from Facebook. user, possibly because they have indistinguishable side interests from the user. In this manner, user needs to additionally need to receive notices focused by these individuals. Figure 5 presents 2 other found standards of conduct. The number in bracket on each social requirement demonstrates the degree the imperative is bolstered in the information. In Figure 5, k1 demonstrates that 80% companions are commented on the users posts while k2 implies that 92% are remarks for the user’s answers. As per this design, user’s posts are very fascinating also which is derived from the post’s, alluring to his/her companions[1]. Furthermore, the user appreciates companions for their remarks by nearly answering to eachone. User’s posts,user’s liveness and lastly shared posts are determined from the two examples in Facebook.

IV. ANALYSIS ON THE RESULT

Mining algorithms given the exact predicted result in the tested data. In 344 records of sample data the 240 records are trained (70 % of the sample data) the remaining data is tested as result is shown in the figure 6.

		Reference	
Prediction	Bad	Good	
Bad	12	6	
Good	7	77	

Accuracy : 0.8725
 95% CI : (0.7919, 0.9304)
 No Information Rate : 0.8137
 P-value [Acc > NIR] : 0.07637
 Kappa : 0.5709
 McNemar's Test P-value : 1.00000
 Sensitivity : 0.6316
 Specificity : 0.9277
 Pos Pred Value : 0.6667
 Neg Pred Value : 0.9167
 Prevalence : 0.1863
 Detection Rate : 0.1176
 Detection Prevalence : 0.1765
 Balanced Accuracy : 0.7796
 'Positive' Class : Bad

Figure 6: Statistics on the User's Data.

From figure 6 the negativity prediction value is around 91% so that it has highest prediction rate from the sample data and the accuracy between the users whether he/she is good or not from the sample data for every 10 records the accuracy within 104 records(which is test data) is identified the prediction is minimum around 0.82(as the probability of the result should be of range between 0 to 1).The Accuracy among the tune lengths the 0.88 is the average accuracy of all the tune lengths.[1][8]

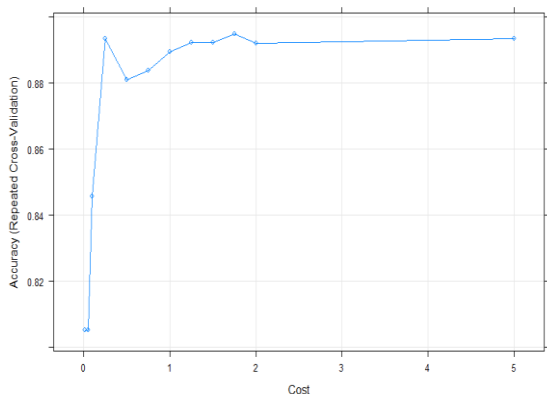


Figure 7: It shows the distribution of the behaviour of Likes and friends_count to predict the Unknown Person Character.

V. CONCLUSION

In this paper, we had determined whether the user is good person or a bad person who had send the friend request by using his data comparing with other members by using data mining techniques from likes, posts, mutual friend's and friends. In future, we will use procedure mining methods and information mining methods to tackle issues which is impossible just with both [1]. For example, how the comment for the post is given by another user who is unknown to him (e.g., to what degree the substance of the appropriate responses co-ordinate the comparing questions) found by information (content) mining.

REFERENCES

- Guang Ming Li and Renata Medeiros De Carvalho, "Process Mining in Social Media: Applying Object-Centric Behavioural Constraint Models", vol.7, Apr.2019.
- G. Barbier and H. Liu, "Data mining in social media," Social Netw. Data Anal., vol. 17, pp. 327–352, Mar. 2011.
- K. Cai, S. Spangler, Y. Chen, and L. Zhang, "Leveraging sentiment analysis for topic detection," Web Intell. Agent Syst., An Int. J., vol. 8, no. 3, pp. 291–302, Jan. 2010.
- L. Cheng, B. F. van Dongen, and W. M. P. van der Aalst, "Efficient event correlation over distributed systems," in Proc. 17th IEEE/ACM Int. Symp. Cluster, Cloud Grid Comput., May 2017, pp. 1–10.
- E. Baatarjav, S. Phithakkitnukoon, and R. Dantu, "Group recommendation system for Facebook," in Proc. OTM Confederated Int. Conf. Move Meaningful Internet Syst., Nov. 2008, pp. 211–219.
- D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins, "Information diffusion through blogspace," in Proc. 13th Int. Conf. World Wide Web, May 2004, pp. 491–501.
- P. Gundecha and H. Liu, "Mining social media: A brief introduction," in Proc. New Directions Inform., Optim., Logistics, Prod., Sep. 2012, pp. 1–17.
- H. Yin, B. Cui, H. Lu, Y. Huang, and J. Yao, "A unified model for stable and temporal topic detection from social media data," in Proc. IEEE 29th Int. Conf. Data Eng. (ICDE), Apr. 2013, pp. 661–672.
- A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media," Bus. Horizons, vol. 53, no. 1, pp. 59–68, 2010.
- G. Li and W. M. P. van der Aalst, "A framework for detecting deviations in complex event logs," Intell. Data Anal., vol. 21, no. 4, pp. 759–779, Jan. 2017. 84372.
- Jeejo Vetharaj, J., Selvanayaki, S., Deepak, V., "User hand gesture recognition in robotics," International Journal of Control Theory and Applications, Volume 9, Issue 34, 2016, Pages 639-642.

AUTHORS PROFILE



DEEPAK V – Currently working as assistant professor, Koneru Lakshmaiah Education Foundation, Vijayawada, and an active researcher scholar in the same university, mainly learning and working in data mining, knowledge engineering and also in other domains
Email – v.d@live.in



M.BHARGAVA TEJA - Currently Pursuing my B.Tech final year in the Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vijayawada, a software developer, an active member in the group, started my research career in data mining.
Email - maddulabhargavateja@gmail.com



K. VENKAT RAGHAVA – A B.Tech final year student in Department of computer science and engineering, Koneru Lakshmaiah Education Foundation, Vijayawada, and an actively participating in all the research work carried on in the department. Proposed this project as a final year project and actively looking forward to implementing in the real time scenario in the future work.
Email - raghava.venkat.143@gmail.com



SRAVANI V - A B.Tech final year student in Department of computer science and engineering, Koneru Lakshmaiah Education Foundation, Vijayawada, and an actively participating in all the research work carried on in the department. Proposed this project as a final year project and actively looking forward to implementing in the real time scenario in the future work.
Email - vankayalapatiravani12@gmail.com