To Identify the Criminal tendency of a Person by analysis of their social media profiles (Facebook)

Sajjad Khan
MS Computing,
Shaheed Zulfikar Ali Bhutto Institute of Science and Technology
90 and 100 Clifton
Karachi -75600
sniazi@hotmail.com

Zohaib Jan
Assistant Professor, Department of Computer Science
Shaheed Zulfikar Ali Bhutto Institute of Science and Technology
90 and 100 Clifton
Karachi -75600
zohaib.jan@szabist.edu.pk

Abstract— **The social networks become the important part of the life. During the last decade, use of social networks has increased to a great extent. People publish their content on the web they use to convey their idea by posting statuses, links, videos, images, comments(Facebook). The use of social networks rapidly increased during the last decade and there is huge amount of data publish on these networks People share their personal information and insights of their lives. In this paper we thoroughly studied five big personality models and come to the point that neuroticism reflects the criminal mind of person, we are using the text mining algorithm, which will be used to extract necessary information from the text massages (statuses, comments etc.) containing neuroticism words e.g. anger, anxiety stress, depression etc. and by using these words prediction can be made. The main aim of this paper is to find the crime tendency in person from the information they share on social networking site (FACEBOOK).**

Keywords—*Data mining, Psychology Big five personality traits, Facebook, K-NN, Crime Tendency.*

I. INTRODUCTION

Social media is the place where people share their personal information to the world. . A survey had been conducted in January 2005 and estimated that there were approximately more than 115 Million members on social networks [6].In last 5 years the FACEBOOK alone has around 500 million members. People reveal a lot of information about themselves during creation of their profile and even after the use by sharing or commenting. People share statuses, upload images, and interests so that will be useful in determining their personality. In this paper we are going to use the information people share on their FACEBOOK profile through commenting or status updates and then by using the text mining technique we will predict the crime tendency of a person. We will search for the anger or negative words people used to publish on the web as these words will reflect their state of mind. In the previous research it was shown that information people give at the time of creation of their profile is idealized version not the actual reflective of their personality. We will use the information people post on the web not the information they wrote in their profile.

We are actually using the language features of the users FACEBOOK profiles. We put the linguistic feature of the profiles that is status updates, Comments and the post containing text into a single string for analysis.

A. Big Five Personality traits

According to the psychology there are five big personality traits through which the personality can be judge and these five includes openness, neuroticism, agreeableness, extraversion, and,conscientiousness.

A.   *Openness:* It is related to the person inner feelings and absorption of environment, creativity and also the political views etc. the person with high openness can be artistic in nature.

B.   *Extraversion:* People with high extraversion are talkative and are socially very active they tend to make new friends and want to attend big social gatherings. Examples are

C.   *Agreeableness:* People having high agreeableness tend to believe people more easily and they think that people around them are honest and trust worthy.

D.   *Conscientiousness:* People with high conscientiousness are usually very organized and discipline they are very careful and they want themselves to be perfect in their lives so they care a lot before doing anything.

E.   *Neuroticism:* People with high neuroticism are more depress and remain angry over little things. In other words these personality traits define the negativity of person. Person with high neuroticism area has angry faces, mental disorders, anxiety and also they have more jealousy factor.

B. Text Mining

It is the process of extracting the useful information from text. It is the discovery of text by the computer from previously unknown information by mining it from different written documents. It includes tokenization, stemming, parsing and N-gram of structuring text. Our research is based on the prediction and classification. There are lots of different problems that can be solved by using text mining procedures.

C. Text Mining Applications

Text mining is mainly used for the extraction of information from the text. This field also covers lot of other applications and problems, e.g. retrieval of data, storage, email support, spam filtering, recommendations and suggestions(e.g. Amazon) and labeling of document(Automatically)..

D. Information Retrieval

Information retrieval mostly belongs to the online documents. In this we retrieve the information from the unstructured text mostly the textual document by using the query which might itself be unstructured. We provide attribute to the written or online document to extract the desire values from that document. In short the concept is to retrieve the similar things among the document.

E. Classification of Document

Document classification is the process of categorizing unlabeled articles or documents. The sample training data is provided with labels for example "cricket", "Information technology" and "News" and from these information the classifier should be able to accurately predict the newly unseen document to correct class.

II.   CLUSTERING OF DOCUMENT.

Clustering document is not as much power as text classification. It is basically the assignment of label which is required for text categorization. Clustering document involves descriptors and extraction of descriptors. It is useful in the companies who want to know which domain problem is severe then other.

A. Information Extraction

It is the process of automatically extracting the information from the semi structured or un-structured text. In most cases it processing is done through human language text through natural language.

B. Prediction and Evaluation

Prediction and evaluation in general belongs to the predication. In this we train the program through generalized rules from the sample document and it will give relatively correct answer based on generalized rules. Basic concept is to find out similarity between the documents.

III.   SOCIAL NETWORKING SITE (FACEBOOK)

Facebook is the most famous social networking site. User has a list of friends, can join groups, likes pages of relative interest. User can update their status and comment on the other posts or their own post, similarly they can also like each other activity and share photos, images, links etc., User share their emotions and state of mind through different status updates and by commenting on the pages, groups. User can share each other status and post which are public or they have access. Facebook also has the options for chat and can send inbox messages to each other.

Recently Facebook also included the voice and video calling service and also the graph API which is given randomly given to the users through graph API users can search for people or any other thing directly by typing in the name.   In last 5 years Facebook alone have roughly 500 million users. So we can expect how huge amount of data is being shared on the Facebook. Users share a lot of their personal information on FACEBOOK which can be used to analyze their personality.

IV.   LITERATURE REVIEW.

FACEBOOK becomes one the most popular social network where lot of user share their information and hence a lot of research has been made in detecting the personality of the users by using their personal information.. Golbeck, and Roblesr shown that from the public information users share on their Facebook Profiles their Big Five personality traits can be predicted [1].   Bachrach and Stillwell Show that personality traits are correlated with patterns of social network use, as reflected by features of Facebook profile[2] , In previous research Back and. Stopfer, stated that Facebook profiles does not reflect the idealized version but it reflects the users actual personality [3]. It has been shown in[4] that extroversion and conscientiousness positively correlate with the perceived ease of use of social media websites.. Hodgkinson and Ford concluded that people's personality can be successfully judged by the others based on their Facebook profiles [5].

V.   DISCUSSION OF RELATED WORK.

Personality prediction can be done by using the personal information of the user profile. There are big five personality traits which are now used worldwide by the psychologist known as big five personality model. Researchers used this personality model in order to detect the personality of a person through Facebook. These five personality traits are Openness, Conscientiousness, Extroversion, Agreeableness and Neuroticism. Many researchers found that personality can affect job performance and other issues. Personality can also be important for marriages and most of the people first look into the social network profile before going to meet someone and marry. In previous researches look for the language features and personal information interest hobbies to judge the personality

of users by searching for the words within their profile and to then define where these words life in the big five personality traits researcher look for the swear, anger, anxiety and other words and also the last name plus the about me in their profile and prediction is then made upon these results [1]. Researchers also uses the other information in order to predict the personality of the user by looking at the number of status updates, likes, comments, photo sharing and no of friends[2].in above researches data is collected through the Facebook API which save the time and also the cost spend on questionnaires. Some researchers also think that the information user give when creating profile is an idealized version not the actual representation of their personality. But the research also shows that human somehow share their emotion somewhere on social networks and as the time goes on the use of social network is keep on increasing which make it quite easy to find the personality. After all these literature reviews and research we come to the point of detecting the criminal tendency of a person by using their social profiles as at some point of time people will show their reaction and state on mind on social networks hence we gather the information based on anger words used by the users in their post to find out criminal tendency.

VI. ANALYSIS OF RELATED WORK.

In this research IS we have worked on criminal tendency of person through their social networking profile. We gather the data from different profiles and apply algorithm on it to predict the criminal tendency of person. We have used supervised approach on it and trained the sample dataset by marking them in two attributes tendency and non-tendency and afterwards we apply algorithm on non-trained data to get the tendency of person through this process

VII. EXPERIMENTAL WORK

In this research paper we get the Facebook data by hours and hours of crawling by using the Facebook graph API. The database we use is MySQL and RAPIDMINOR is used for text mining algorithm. We used supervised approach and mark them into two attributes tendency and non-tendency. Supervision is done on the text which we get from the profiles and rather than the profile itself.

We export the trained example set data in the CSV format from the Database. We import the CSV files in the RapidMinor and perform the necessary steps for the output.

In RapidMinor Main process we read the CSV file by the CSV Read CSV operator. We then process the document and generated word vectors from String attributes.

we define the set the role and also sets the label for target attribute for learning and applied the K-NN algorithm for text mining. We applied the model operator to our example dataset

for the prediction. In the last we used the read dataset operator for reading the example dataset from the database.

Plotter view is used to get the graph used for showing our results. The Results we obtained are discussed in the Results section and also the plotter graphs are displayed.

VIII. RESULTS

The results obtained by applying the algorithm are discussed in this section. We first apply the algorithm on single profile to find that the crime tendency.
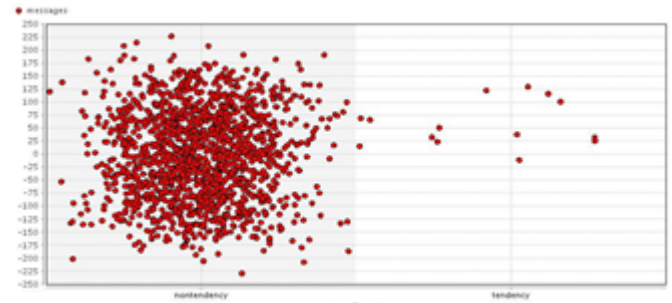

Fig 1: Results of single profile

In the above graph there is relation between attribute and messages where we can clearly see the relation among them. This profile user uses more keywords which are non-tendency and less crime tendency words so we can conclude that this profile has very little tendency of crimes.

Then we compare two different profiles to find out that which one has more crime tendency relative to each other.
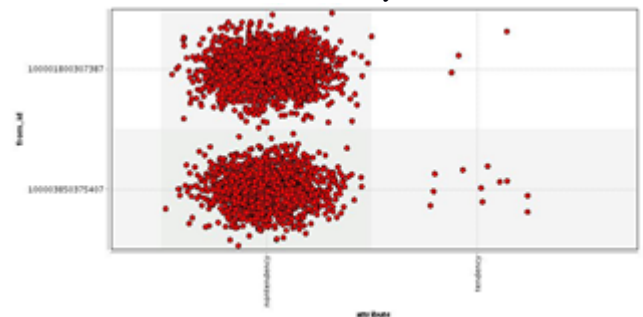

Fig 2: Comparison of two profiles.

We can clearly see in the above graph that first profile uses more words with crime tendency and second profile uses less crime tendency keywords so first profile has more tendency towards crime compared to the second.
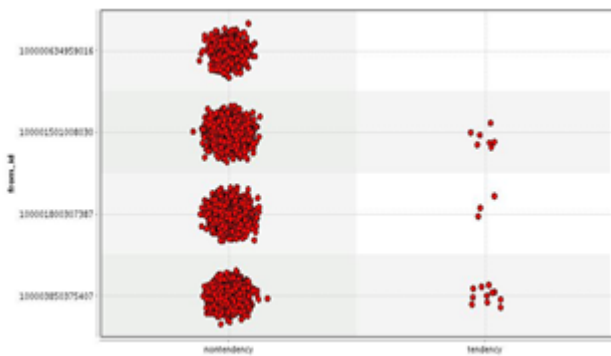
Fig 3: Comparison of Four different profiles.

Then we implemented the same algorithm on four different profiles and find out the result written above the fourth profile id do not have any word related to crime tendency so we can say that fourth profile has zero crime tendency and the first profile has the more crime tendency in contrast to others .we can also see that the fourth profile although uses the criminal tendency words but it has not crime tendency as shown above so it shows uses the keywords doesn't reflect the criminal tendency of the person but the way person use these words in their sentence show the criminal tendency.
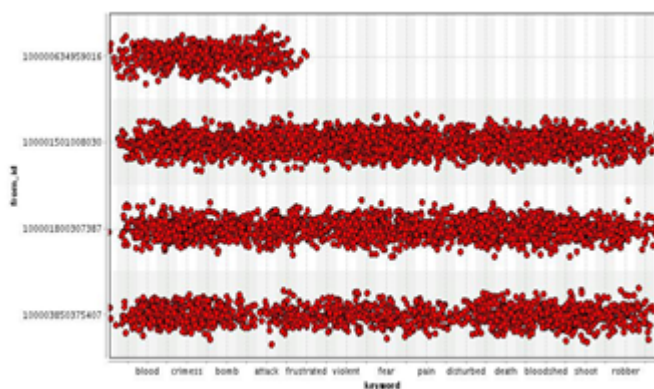

Fig 4: Keywords used by four Profiles

IX.   CONCLUSION & FUTURE WORK

We give our approach to find out the crime tendency on person by using their social networking site (Facebook) dataset. It is quite possible to predict the tendency of a crime in personality. But there are few concerns which effect the crime prediction.

- Our result might do not reflect the accurate results as we are only using the public data.

- There are very huge people who use these sorts of words in friendly mood or just to irritate some friends.

- People with criminal mind share less things and socially not very active.

Above listed issue are challenging and need a lot of further work and need to be research.

It is an emerging topic detecting criminal tendency through social media profiles and by using this we can get aware of the people with criminal mind with low cost and efficiency.

Acknowledgment

References

[1].   Bernardo J. Golbeck, C. Robles, and K. Turner, "Predicting personality with social media," in Proc. of the 2011 annual conference extended abstracts on Human factors in computing systems, ser. CHI EA '11,2011, pp. 253–262.

[2].   Yoram Bachrach, David Stillwell and Michael Kosinski. Personality and Patterns of Facebook usage, ACM Web Science Conference 2012.

[3].   M. Back, J. Stopfer, S. Vazire, S. Gaddis, S. Schmukle,B. Egloff, and S. Gosling. Facebook Profiles Reflect Actual Personality, Not Self-Idealization. Psychological Science, 21(3):372, 2010.

[4].   The Impact of the Big Five Personality Traits on the Acceptance of Social Networking Website. AMCIS 2008 Proceedings, page 274, 2008.

[5].   G. Hodgkinson and J. Ford. International Review of Industrial and Organizational Psychology, 2008.Wiley-Interscience, 2008.

[6].   J. Golbeck. Computing and Applying Trust in Web-based Social Networks. PhD thesis, University of Maryland, College Park, MD, USA, April 2005.

[7].   F. Mairesse, M. Walker, M. Mehl, and R. Moore. Using linguistic cues for the automatic recognition of personality in conversation and text. Journal of Artificial Intelligence Research, 30(1):457–500, 2007.

[8].   J. Pennebaker and L. King. Linguistic styles: Language use as an individual difference. Journal of personality and social psychology, 77(6):1296–1312, 1999.

[9].   M. Barrick and M. Mount. The Big Five personality dimensions and job performance: A meta- analysis. Personnel psychology, 44(1):1–26, 1991.

[10]. G. Hodgkinson and J. Ford. International Review of Industrial and Organizational Psychology, 2008.Wiley-Interscience, 2008.

[11]. J. Jost, T. West, and S. Gosling. Personality and ideology as determinants of candidate preferences and Obama conversion in the 2008 US presidential election. Du Bois Review: Social Science Research on Race, 6(01):103–124, 2009.

[12]. P. Rosen and D. Kluemper. The Impact of the Big Five Personality Traits on the Acceptance of Social Networking Website. AMCIS 2008 Proceedings, page 274, 2008.

[13]. M. Selfhout, W. Burk, S. Branje, J. Denissen, M. van Aken, and W. Meeus. Emerging Late Adolescent Friendship Networks and Big Five Personality Traits: A Social Network Approach. Journal of personality, 78(2):509–538, 2010.

[14]. J. Salgado. The Big Five personality dimensions and counterproductive behaviors. International Journal of Selection and Assessment, 10(1&2):117–125, 2002.

[15]. Y. Amichai-Hamburger and G. Vinitzky. Social network use and personality. Computers in Human Behavior, 26(6):1289–1295, 2010.

[16]. T. Ryan and S. Xenos. Who uses Facebook? an investigation into the relationship between the big five, shyness, narcissism, loneliness, and Facebook usage. Computers in Human Behavior, 2011.

[17]. S.D. Gosling, S. Gaddis, and S. Vazire. Personality impressions based on Facebook profiles. 2007.

[18]. S. Zhao, S. Grasmuck, and J. Martin. Identity construction on Facebook: Digital empowerment in anchored relationships. Computers in Human Behavior, 24(5):1816–1836, 2008.

[19]. B. Zhong, M. Hardin, and T. Sun. Less effortful thinking leads to more social networking? the associations between the use of social network sites and personality traits. Computers in Human Behavior, 2011.

[20]. D. Peabody and B. De Raad. The Substantive Nature of Psychological Personality Factors: A Comparison across Languages. Journal of Personality and Social Psychology, 83(4):983–997, 2002.