Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

An Analysis for the Prediction of Human Behaviour & observation level on social media Using Machine Learning Approaches

Luxmi Sapra[1], Rahul Bhatt[2], Gesu Thakur[3]

[1, 3]Associate Professor, Computer Science & Engineering, School of Computer Science & Engineering, Dev Bhoomi Uttarakhand University, Chakrata Road, Manduwala, Naugaon, Uttarakhand 248007

[2]Assistant Professor, Computer Science & Engineering, School of Computer Science & Engineering, Dev Bhoomi Uttarakhand University, Chakrata Road, Manduwala, Naugaon, Uttarakhand 248007

[1]socse.luxmisapra@dbuu.ac.in, [2]socse.rahul@dbuu.ac.in, [3]head.ca@dbuu.ac.in

Article Info
Page Number: 2606-2620
Publication Issue:
Vol. 71 No. 4 (2022)

Article History
Article Received: 25 March 2022
Revised: 30 April 2022
Accepted: 15 June 2022
Publication: 19 August 2022

Abstract

Usage over Internet has been significantly increased during last few decades. Peoples sparing more time on social media services. Social media is a place where users present themselves to the world, revealing personal details and insights into their lives. We are starting to comprehend how some of this information might be employed to better the users' experiences with interfaces and with one another. In this research proposal, we are interested to predict the personality of users by evaluating their tweets. In the past, users were required to complete a personality test before their characteristics could be adequately analyzed. Because of this, doing personality analysis in many different aspects of social media became impracticable. In this research proposal, we apply neural networks by which a user's personality can be accurately predicted through the publicly available information on their Twitter profile. We will present the sort of data obtained, our methods of analysis, and the machine learning approaches that enable us to correctly predict personality. It is essential for companies to have this information in order to target potentially interested customers or to get customer feedback in the event that diversification is pursued as a business strategy. Thus, this work analyzes social media data to predict significant personality traits, i.e. qualities or characteristics specific to an individual. The widespread use of social media sites results in an increase in both the quantity and amount of data. The quantity of data that is submitted to these social networking platforms is expanding day by day. Because of this, there is a significant need to investigate the very variable behavior of consumers in relation to these services. This is a preliminary work to model the user patterns and to study the effectiveness of machine learning active modeling approaches on leading social networking service Facebook. We created a model of the user comment patterns found on Facebook Pages and used it to make a prediction about the number of comments each post is likely to receive within the next 24 hours.

Keywords: Machine Learning, Natural Language Processing, Online Social Network, Personality Test, Profiling, Sentimental Analysis, and Twitter are some of the keywords that might be associated with this article.

1 INTRODUCTION

People have begun to spend a significant amount of their time on websites where anybody may edit and add content to. As a result, in order to satisfy this need, several web technologies that enable users to collaborate and contribute in an interactive manner have been implemented. Blogs, wiki pages, portals, and social networking sites are examples of these types of technology. The term "Web 2.0 Technologies" has been given to refer to these developments. Users are able to contribute and share material via the use of these technologies, and it is not necessary for them to have any prior experience or understanding in web development. People are now able to connect with individuals who share their interests via the use of these various technologies. The last few decades have seen the launch of a number of social networking sites, several of which have gone on to achieve enormous levels of popularity throughout the globe. Facebook, Twitter, YouTube, LinkedIn, Instagram, Academia, and other social networking sites are included in this category. Each one of these initiatives seeks to achieve a unique goal of encouraging people to open up and talk about meaningful events, ideas, or experiences from their lives. Facebook users have access to a communication network that is made up of their friends, family members, and other individuals that they already know via their real-world interactions with other people. Twitter gives users the ability to broadcast their thoughts and get quick feedback from other users, many of whom may already be acquainted with one another in real life. LinkedIn is a social networking site that focuses on the professional world and offers business professionals a platform for business networking, allowing them to communicate with one another, follow one another, and improve their ability to recruit new employees through enhanced searching facilities based on their professions. These online social networking communities have an impact on our everyday lives. On such platforms, there are a lot of people who are well organized. Twitter, for example, has emerged as a significant alternative to traditional forms of media due to the fact that it facilitates the dissemination of news at a quicker rate and provides more latitude for expression.

Users of social media platforms present themselves to the rest of the world by sharing intimate information and providing glimpses into their lives on these platforms. We are getting a better understanding of how some of this information might be used to enhance the experiences of users both with the interfaces they interact with and with one another.

Within the context of our study project, the personalities of users are of particular interest to us. Personality has been demonstrated to be effective in predicting work satisfaction, the success of professional and romantic relationships, and even preference for various types of user interfaces. This is due to the fact that personality is important to many different kinds of interactions. In the past, users were required to complete a personality test before their characteristics could be adequately analyzed. Because of this, doing personality analysis in many different aspects of social media became impracticable. In this research proposal, we describe a system that may reliably predict a user's personality based on the information that is publicly accessible on their Twitter profile. This information can be gleaned from the user's public profile. In the next section, we will discuss the many types of data that were gathered,

our methods of analysis, and the machine learning approaches that enable us to correctly predict personality. After that, we go on to a discussion of the repercussions this has for the design of interfaces, social media platforms, and other areas.

2. PROBLEM STATEMENT

Sharing one's experiences, thoughts, interests, and memories with other members of one's community satisfies a fundamental human desire to communicate with other members of one's community. They like to communicate over social media platforms such as Twitter, Facebook, personal blogs, and wikis the majority of the time. There are a lot of individuals that contribute to social media platforms on a regular basis by writing about their personal experiences and posting images and status updates. The vast bulk of the stuff that is being shared is private information. The Big Five personality traits—openness, agreeableness, conscientiousness, extraversion, and neuroticism—can be predicted from the material that users of social media platforms post with one another, according to research that has been published in academic journals. These studies often make use of linguistic characteristics, information from social networks, and the frequency with which individuals engage with the platform in question, such as the amount of status updates, images, videos, and likes that are submitted. The purpose of this study is to determine whether characteristics of the material that users publish on Facebook are connected with the Big 5 Personality Traits of other users and then build a model for predicting personality characteristics based on these characteristics. The contribution of this research can be broken down into two parts. First, we demonstrate that the current methods for predicting a person's Big 5 Personality perform more well when there is adequate data in the form of a high number of posts in their social media profile. Second, we demonstrate that the incorporation of information on users' friends, such as information regarding their Big 5 Personality, enhances the accuracy in comparison to other approaches that have been presented in the existing body of research.

Traditionally, a user's profile information, status updates, messages made by the user, and other such things are used to make predictions about their personality. This research presents a method for predicting a person's personality based not on a single line of text but on a collection of tweets rather than on a single tweet alone. This particular system is made up of three separate parts. The first module gathers all of the tweets' meta-data into a database known as a "meta-base." It does not take into account any information about the user's profile. The multi-label problem is converted into five different binary classification problems by the second module. The final step in determining personality characteristics is to use a classification algorithm that makes use of a multilayer neural network. This method employs a lot of labeled data in order to properly categorize the unlabeled data. This is necessary due to the increased velocity at which social media data is being created. In order to illustrate how well this system works, a collection of Tweets that are specific to a person is received through the Twitter API and then applied to the system. As a method of classification, the machine learning technique known as Multilayer Perception Neural Network is used. The system achieves favorable results and properly predicts the personalities of 'Tweeters' without taking into account the information that is included in their

profiles. This method is distinct from others that have been examined in the research literature.

Since the beginning of the last decade and a half, the leading trends toward social networking services have attracted a significant amount of public attention. The convergence of computers and the physical world has made it possible to transform commonplace items into information appliances. These services are functioning as a multi-tool with a variety of ordinary uses, such as communication, commenting, news, ads, banking, and marketing, amongst others. Every day, these services bring forth new and exciting innovations, and many more are on the way. All of these services produce large amounts of content on a daily basis, and Hadoop clusters are the most likely place for this content to be stored. As an example, Facebook receives over 500 terabytes of fresh data every single day, has over 100 petabytes of disk capacity in one of its biggest Hadoop (HDFS) clusters, and shares over 2.5 billion pieces of content every single day (status updates, wall postings, images, videos, and comments). In 2007, Twitter saw an average of 5,000 tweets per day; in 2013, that number increased to 500,000,000 tweets per day. As of January 31, 2011, Flickr hosted 5.5 billion images, and between 3,000 and 5,000 new images are being uploaded to the site every minute.

This research proposal, in addition, makes use of the most popular social networking service Facebook, more specifically 'Facebook Pages' (one of the products offered by Facebook), for the purpose of performing an automatic analysis of user trends and patterns. As a result of this, in order to complete this job, we produced a software prototype that is comprised of a crawler, information extractor, information processor, and knowledge discovery module. Our line of inquiry is focused on the comment volume projection (CVP) of the responses to a document that are anticipated to be made during the following H hours.

3. PERSONALITY MODEL

In light of the fact that the Five Factor Model (FFM, also known as the Big Five Model) is currently the model of personality that has received the greatest amount of attention and is the one that is considered to be the most accurate, we came to the conclusion that it would be best to use it in this investigation. [5, 6] This model has been proven to subsume the most well-known personality characteristics, and it offers a nomenclature as well as a conceptual framework that integrates a significant amount of the research results in the field of psychology dealing with individual variations and personality.

The Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism factors are the five dimensions that make up a person's personality, according to the Five Factor Model (OCEAN). Each dimension has its own set of features that best describe it.

A person's willingness to try new things, be curious, seek out new experiences, and be interested in culture, ideas, and aesthetics can all be gauged by their openness to experience.

The degree to which a person is organized, diligent, and scrupulous may be inferred from

their level of conscientiousness.

The inclination of a person to look for stimulation in the outside world, to seek the company of other people, and to show happy feelings is what's meant by extraversion.

The level to which a person is focused on maintaining pleasant social ties may be measured using the agreeableness scale. This scale reflects a person's disposition to be trusting, empathetic, and cooperative.

The propensity to suffer mood swings and unpleasant feelings like guilt, anger, anxiety, and despair are hallmarks of neuroticism, a personality trait that is also sometimes referred to as emotional instability. The five characteristics have been shown to be inherited genetically, to be constant throughout time, and to be consistent across genders, cultural backgrounds, and racial groups. [7].

Table 1 summarizes the big five personality traits along with their representative descriptive terms for both low and high scorers.

Table 1: Overview of the Big Five personality model

Trait	**Description**
Openness	Openness is related to imagination, creativity, curiosity, tolerance, political liberalism, and appreciation for culture. People scoring high on Openness like change, appreciate new and unusual ideas and have a good sense of aesthetics
Conscientiousness	Conscientiousness measures preference for an organized approach to life in contrast to a spontaneous one. People scoring high on Conscientiousness are more likely to be well organized, reliable, and consistent. They enjoy planning, seek achievements, and pursue long-term goals. Non-conscientious individuals are generally more easy-

	going, spontaneous, and creative. They tend to be more tolerant and less bound by rules and plans.
Extroversion	Extroversion measures a tendency to seek stimulation in the external world, the company of others, and to express positive emotions. People scoring high on Extroversion tend to be more outgoing, friendly, and socially active. They are usually energetic and talkative; they do not mind being at the center of attention, and make new friends more easily. Introverts are more likely to be solitary or reserved and seek environments characterized by lower levels of external stimulation
Agreeableness	Neuroticism Emotional Stability, reversely referred to as Neuroticism, measures the tendency to experience mood swings and emotions such as guilt, anger, anxiety, and depression. People scoring low on Emotional Stability are more likely to expensive stress & nervousness.

4. Background and related work

Previous studies have shown that a person's personality can effectively explain a significant amount of the variability that exists in human preferences and behavior across a variety of different domains, such as their preferences regarding media and cultural aspects [8, 9] and their use of social networking websites [10]. The degree of enjoyment that an individual

receives from the stimulation they get from the outside world is said to be contingent on the individual's optimum or ideal level of arousal. It is believed that one's information processing capabilities and emotional orientations have a role in determining which item they like more [11]. Because of this, researchers have concluded that an individual's personality is an important factor in comprehending their level of enthusiasm for the arts, such as paintings and music [8, 11]. Recent studies have revealed that some personality traits might be regarded to be key mediators of preferences for the content of various media. Kraaykamp and Eijck [12] investigated the effect of the Big Five personality traits on individuals' choices for media (namely, television shows) and their involvement in cultural activities (book reading and attending museums and concerts). According to what they discovered, openness certainly fosters an interest in intricate and fascinating forms of leisure activity. The traits of conscientiousness and friendliness (agreeableness), on the other hand, have a tendency to have a negative influence on activities that are either challenging or unorthodox, whereas the trait of emotional stability tends to have a negative influence on more predictable means of escape from day-to-day life. According to the research presented in [13], preferences for websites are impacted by personality factors, much as preferences for products in the real world. According to the authors' findings, the viewers of websites often have different personality profiles, and the association between personality and preferences connected to websites and website categories has psychologically relevant implications. Websites dedicated to social media, such as Facebook and Twitter, have recently become more popular as a primary medium through which individuals can connect with one another and share their own perspectives. Researchers have recently taken an interest in the ways in which users' personalities influence their interactions on social networking networks. According to the research presented in [14], extroverts have a tendency to perceive social networking sites to be intuitive and helpful. Users are more likely to pick contacts who have comparable personality qualities, and they have a general tendency to favor those who are high in Agreeableness [15]. Current research interests have been increasingly focused on the relationships between personality and users' use behaviors (such as the number of posts and likes) and profiles (such as the number of friends/followings/followers, age, and gender) on social networking websites [10, 16]. In addition to this, there has been an increase in focus placed on the ability to predict personality characteristic scores based on publicly accessible information on a person's behavior and profile [16, 17]. According to the findings of Golbeck et al. [17], people with varying personalities have a tendency to utilize a variety of terms in their posts and descriptions. According to research carried out by Quercia et al. [16], who analyzed the behavior of Twitter users, the most popular and influential individuals are extroverts who are also emotionally stable. They went on to uncover that users who are popular tend to be "imaginative" (high in Openness), but users who are influential tend to be "organized" (high in Conscientiousness). On a separate social networking site, Facebook, Quercia et al. studied the association between sociometric popularity (number of Facebook contacts) and personality qualities in their study [10]. They came to the conclusion that popular individuals on Facebook tend to have personalities that are similar to those who are popular in the real world. In a similar vein, [18] showed that there is a substantial correlation between a person's personality qualities and different aspects of their Facebook accounts. To

the best of our knowledge, very few research have been conducted on the implications of users' personalities on their behaviors when it comes to modeling user preferences. Within the scope of this study proposal, our primary objective is to provide a response to the following key research question: to what degree do personality factors influence rating behaviors?

The user-generated material on Twitter (such as tweets, for example) is another useful source of information that may be used to infer the personality qualities of Twitter users. One of the Twitter datasets that is often utilized in the research is the one that was gathered as part of the Personality project. Only a few hundred people out of the thousands of participants who took part in the my Personality study actually provided links to their Twitter accounts. The information that is included in this dataset comes from those individuals. Both the goal of automatically predicting the personalities of the users and the work of conducting studies of user behavior have been accomplished with the assistance of this data set [16, 17, 19]. Extroverts and those who are emotionally stable tend to be the most popular and influential users on Twitter, according to research conducted by Quercia et al. [16]. Additionally, it was shown that popular users have a vivid imaginations, but powerful individuals on Twitter are more well-organized. When Golbeck et al. [17] trained machine learning algorithms to predict scores on each of the five personality characteristics, they utilized profile information from the data seats' attributes. The scores were accurate to within 11–18% of the actual value for each of the five personality qualities.

On the other hand, Hughes et al. [19] acquired a distinct dataset from Twitter by posting an advertisement on both Twitter and Facebook. This allowed them to collect data from both platforms. The results of their research indicated that there is a distinct difference in the association between actions on Twitter and Facebook. It was also shown that individuals who have a preference for Facebook or Twitter had personality distinctions, which suggests that different people use the same sites for different reasons. This was found to be the case despite the fact that both sites are popular.

5. Methodology to predict personality on social media platforms

This work provides a personality prediction system for social network data analysis. The flow diagram of the system is given in Fig. 2. The system consists of four modules: Data collection, pre-processing, transformation and classification.

Fig. 2 Flow chart of Personality trait prediction system

The next paragraphs will elaborate on these modules.

1) Data Collection: In order to successfully demonstrate the system, we need tweets that have been posted by individuals (s). Tweets are received via the use of the Twitter API for this purpose. The Twitter Application Programming Interface [20] gives users access to Twitter data like as information about people, tweets produced by users, search results on Twitter, and other relevant data. JSON is the format used for the Tweet object.

2) The pre-processing module begins by obtaining the tweets from the object referred to as the tweet. The algorithm then pulls meta-attributes from the tweets that have been sent. The information that was gleaned may be categorized as social behavior information and linguistic information respectively. The information on the grammar comprises the average length of the text, the average amount of positive and negative words, and the average number of special characters such as commas, question marks, and so on. The information on social behavior consists of things like the average number of links, the average number of hash tags, the average number of mentions, and so on. To arrive at the average, just divide the sum of a category's attributes that pertain to grammatical and social behavior by the total number of characters in all of the tweets. After the meta-attributes have been extracted, they are sent to the transformation module to be processed.

The "multi-label problem" is transformed into "binary classification issues" by the third module, the transformation module. This module is the recipient of the meta characteristics that were retrieved from the module that came before it. A feature vector is built by using these information as building blocks. The vector is organized in such a way that each location corresponds to a meta-attribute. After that, the classification module receives this vector, and its output will either be a "1" (indicating "yes") or a "0" (indicating "no") (no). Because the multilayer sensory neural network that is employed as a classifier can only recognize numbers, this change is essential.

4) Classification module:

For the purpose of categorization, a Multilayer Perception (MLP) Neural Network is used. One neural network (classifier) is devoted to each of the five different aspects of personality. The classification module is given a test set in addition to a training set that has already been converted. Every neural network analyzes the test feature vector in comparison to the training feature vectors it was given. The output of each classifier is either a "1" (meaning "yes") or a "0" (meaning "no") based on whether the vectors match or not, which further infers that the individual has the personality feature or not.

6. A METHOD TO GUESS THE NUMBER OF COMMENTS THAT WILL BE LEFT ON SOCIAL MEDIA PLATFORMS

Our primary emphasis was placed on more nuanced forms of predictive modeling. We approach this issue as a regression problem so that we may make predictions at a finer granular scale. We modelled the situation using certain articles that had already been published and for which the desired numbers (number of comments received) were already known. The objective of this job is to estimate the number of responses to a post that will likely be received during the following hour. In order to do this, we browsed the Facebook

sites in search of raw data, then we preprocessed the data, and last, we produced a temporal split of the data in order to have the training and testing sets ready. After that, this training set is utilized to train the repressor, and the performance of the repressor is then calculated using testing data (whose, goal value is Checkinled) in conjunction with certain assessment measures. Figure 1 illustrates the whole procedure, and the following text provides further information.

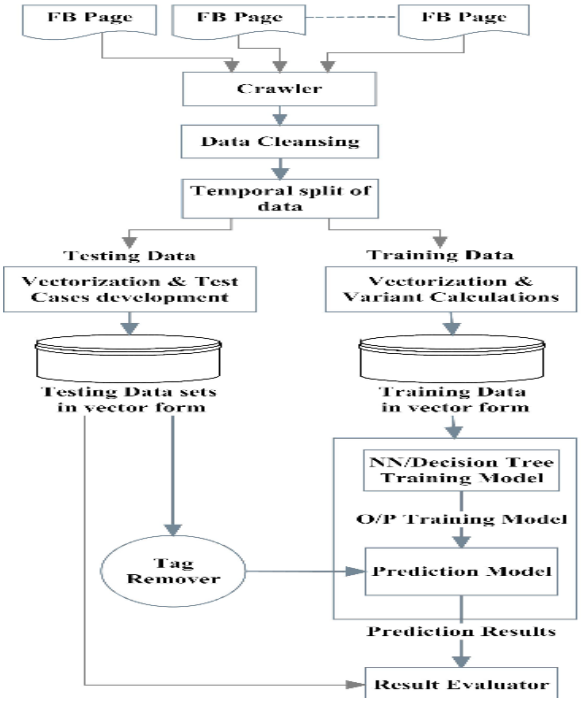

Figure 1: Comment Volume Prediction Process Demonstration

(Facebook set used for this work

We had identified 69 features and 1 as target value for each post and categorized these features as:

1) *Page Features:* We identified 4 features of this category that includes features that define the popularity/Likes, category, checkin's and talking about of source of document.
Page likes: It is a feature that defines users support for specific comments, pictures, wall posts, statuses, or pages.
Page Category: This defined the category of source of document eg: Local business or place, brand or product, company or institution, artist, band, entertainment, community etc.

Page Checkin's: It is an act of showing presence at particular place and under the category of place, institution pages only.

Page Talking About: This is the actual count of users who are 'engaged' and interacting with that Face book Page. The users who actually come back to the page, after liking the page. This includes activities such as comments, likes to a post, and

shares by visitors to the page.

2) *Essential Features:* This includes the pattern of comment on the post in various time intervals w.r.t to the randomly selected base date/time demonstrated in Figure 2, named as C1 to C5.

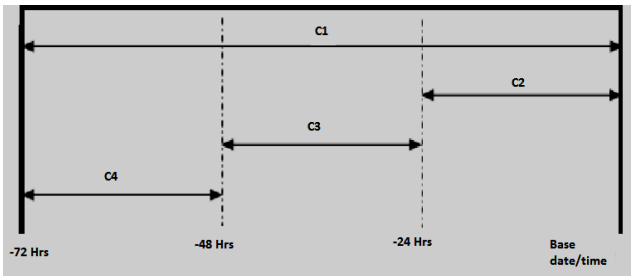

Figure 2. Demonstrating the essential feature details.

C1: Total comment count before selected base date/time.

C2: Comment count in last 24 hrs w.r.t to selected base date/time.

C3: Comment count is last 48hrs to last 24 hrs w.r.t to base date/time.

C4: Comment count in first 24 hrs after publishing the document, but before the selected base date/time.

C5: The difference between C2 and C3. Furthermore, we aggregated these features by source and developed some derived features by calculating min, max, average, median and Standard deviation of 5 above mentioned features. So, adding up the 5 essential features and 25 derived essential features, we got 30 features of this category.

3) *Weekday Features:* Binary indicators (0, 1) are used to represent the day on which the post was published and the day on selected base date/time. 25 features of this type are identified.

4) *Other basic Features:* This include some document related features like length of document, time gap between selected base date/time and document published date/time ranges from (0, 71), document promotion status values (0, 1) and post share count. 5 features of this category are identified.

(ii) Crawling

The data originates from Facebook pages. The raw data is crawled using crawler that is designed for this research work. This crawler is designed using JAVA and Facebook Query Language (FQL). The raw data is crawled by crawler and cleaned on basis of following criteria:

• We considered, only those comments that was published in last three days w.r.t to 1base date/time

as it is expected that the older posts usually don’t receive any more attention.

• We omitted posts whose comments or any other necessary details are missing. This way we produced the cleaned data for analysis.

(iii) Pre-processing

The crawled data cannot be used directly for analysis. So, it is carried out through many processes like split and vectorization. We made *temporal split* on this corpus to obtain training and testing data-set as we can use the past data(Training data) to train the model to make predictions for the future data(Testing data)[21, 22]. This is done by selecting a threshold time and divides the whole corpus in two parts. Then this data is subjected to *vectorization*. To use the data for computations it is required to transform that data into vector form. For this transformation, we identified some features as already discussed in this section, on which comment volume depends and transformed the available data to vector form for computations. The process of vectorization is different in training and testing set:

1) Training set vectorization: Under the training set, the vectorization process goes in parallel with the variant generation process. *A variant* is defined as, how many instances of the final training set are derived from a single instance/post of the training set. This is done by selecting different base date/times for same post at random and processing them individually as described in Figure 2. Variant - X, defines that, X instances are derived form single training instance as described in the example of Facebook official page id: 107684304856555with post id: 219515841718793, posted on TueFeb

19 08:14:19 IST 2018, post crawled on SatFeb2414:51:48 IST 2018. It received total of 515 comments sat time of crawling as shown in Figure 3.

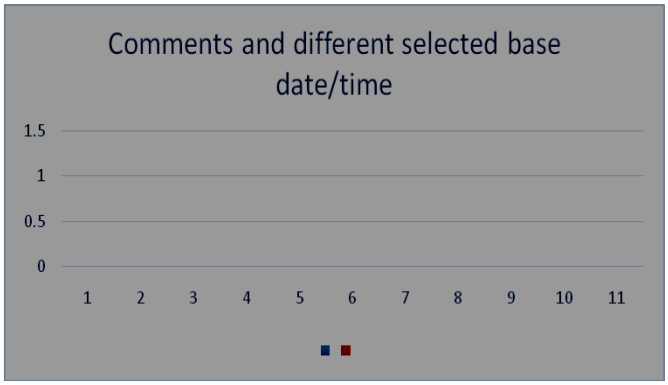

Figure 3. Collective Comments and different selected base date/time.

Know, by selecting different base date/time at random for single post, different variants are obtained for above example shown in Figure 4.

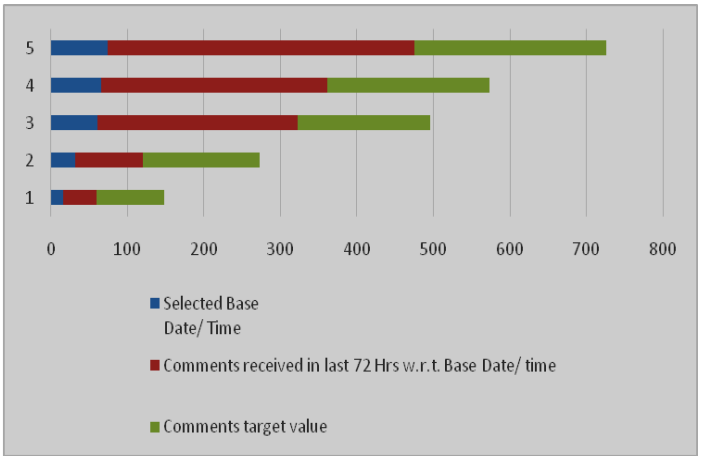

Figure 4

2) Testing set vectorization: Out of the testing set, 10test cases are developed at random with 100 instances each for evaluation and then they are transformed to vectors.

(iv) Predictive Modeling

For the fine-grained evaluation, we have used the Decision Trees(REP Tree[23] and M5P Tree[24]) and Neural Networks(Multi-Layer Preceptron[25]) predictive modeling techniques.

(v) Evaluation Metrics

The models and training set variants are evaluated under the light of Hits@10, AUC@10, M.A.E and Evaluation Time as evaluation metrics:

1) Hits@10: For each test case, we considered top 10posts that were predicted to have largest number of comments, we counted that how many of these posts are among the top ten posts that had received largest number of comments in actual. We call this evaluation measure *Hits@10* and we averaged Hits@10 for all cases of testing data [26].

2) AUC@10: For the AUC [27], i.e., area under the receiver-operator curve, we considered as positive the 10 blog pages receiving the highest number of feedbacks in the reality. Then, we ranked the pages according to their predicted number of feedback and calculated AUC.

Conclusion:

The model demonstrated in section 6 can further be practically evaluated and experimental results may have been obtained. In this further research proposal, we try to explore it practically and more precise results may satisfy our proposed model.

References:

[1] A. Kamilaris, A. Pitsillides, Social networking of the smart home,in: Personal Indoor and Mobile Radio Communications (PIMRC),2010 IEEE 21st International Symposium on, 2010, pp. 2632–2637. doi:10.1109/PIMRC.2010.5671783.

[2] K. Shvachko, H. Kuang, S. Radia, R. Chansler, The hadoop distributedfile system, in: Mass Storage Systems and Technologies(MSST), 2010 IEEE 26th Symposium on, 2010,

pp. 1–10.doi:10.1109/MSST.2010.5496972.

[3] I. Polato, R. R´e, A. Goldman, F. Kon, A comprehensive view ofhadoopresearcha systematic literature review, Journal of Network andComputer Applications 46 (2014) 1–25.

[4] T. Reuter, P. Cimiano, L. Drumond, K. Buza, L. Schmidt-Thieme,Scalable event-based clustering of social media via record linkagetechniques., in: ICWSM, 2011.

[5] Costa, P.T., Jr., McCrae, R.R.: NEO PI-R Professional Manual. Psychological AssessmentResources, Odessa, FL (1992)

[6] John, O., Srivastava, S.: The Big Five trait taxonomy: History, measurement, andtheoretical perspectives. In: Pervin, L., John, O. (eds.) Handbook of Personality: Theoryand Research, pp. 102-138. Guilford Press (1999)

[7] John, O.P., Robins, R.W., Pervin, L.A.: Handbook of Personality: Theory and Research(3rd Edition). Guilford Press, New York (2008)

[8] Rentfrow, P.J., Gosling, S.D.: The do re mi's of everyday life: the structure and personality correlate of music preferences. J PersSocPsychol 84, 1236-1256 (2003)

[9] Kraaykamp, G., Eijck, K.: Personality, media preferences, and cultural participation.Personality and Individual Differences 38, 1675-1688 (2005)

[10] Quercia, D., Lambiotte, R., Stillwell, D., Kosinski, M., Crowcroft, J.: The personality of popular facebook users. Proceedings of the ACM 2012 conference on ComputerSupported Cooperative Work, pp. 955-964. ACM, Seattle, Washington, USA (2012)

[11] Ganzeboom, H.B.G.: Explaining differential participation in high-cultural activities: Aconfrontation of information-processing and status-seeking theories. In: Raub, W. (ed.)Theoretical models and empirical analyses: Contributions to the explanations of individual actions and collective phenomina. E.S. Publications, Utrecht (1982)

[12] Zuckerman, M., Ulrich, R.S., McLaughlin, J.: Sensation seeking and reactions to naturepaintings. Personality and Individual Differences 15, 563-576 (1993)

[13] Kosinski, M., Stillwell, D., Kohli, P., Bachrach, Y., Graepel, T.: Personality and WebsiteChoice. ACM Conference on Web Sciences, (2012)

[14] Rosen, P., Kluemper, D.H.: The impact of the big five personality traits on the acceptanceof social networking website. AMCIS 2008 Proceedings 274 (2008)

[15] Schrammel, J., Köffel, C., Tscheligi, M.: Personality traits, usage patterns and informationdisclosure in online communities. Proceedings of the 23rd British HCI Group AnnualConference on People and Computers: Celebrating People and Technology, pp. 169-174.British Computer Society, Cambridge, United Kingdom (2009)

[16] Quercia, D., Kosinski, M., Stillwell, D., Crowcroft, J.: Our twitter profiles, our selves:Predicting personality with twitter. Privacy, security, risk and trust (passat), 2011 ieeethird international conference on and 2011 ieee third international conference on socialcomputing (socialcom), pp. 180-185 (2011)

[17] Golbeck, J., Robles, C., Turner, K.: Predicting personality with social media. CHI '11Extended Abstracts on Human Factors in Computing Systems, pp. 253-262. ACM,Vancouver, BC, Canada (2011)

[18] Bachrach, Y., Kosinski, M., Graepel, T., Kohli, P., Stillwell, D.: Personality and patterns

of facebook usage. ACM Conference on Web Sciences, (2012)

[19] Hughes, D.J., Rowe, M., Batey, M., Lee, A.: A tale of two sites: Twitter vs. Facebook and the personalitypredictors of social media usage. Comput. Hum. Behav. **28**(2), 561–569 (2012)

[20] Twitter API available on *https://apps.Twitter.com/.*

[21] Cambria, B. Schuller, B. Liu, H. Wang & C. Havasi "*Knowledge-based approaches to concept-level sentiment analysis*", IEEE Intelligent Systems, 28(2), pp.12–14, 2013.

[22] E. Tupes and R. Christal, "*Recurrent personality factors based on trait ratings*", Journal of Personality, vol. 60, no. 2,pp. 225–251, 1992.

[23] R. McCrae and O. John, "*An introduction to the five-factor model and its applications*", Journal of personality, vol. 60, no. 2, pp. 175–215, 1992.

[24] J. Digman, "*Personality structure: Emergence of the five-factor model*", Annual review of psychology, vol. 41, no. 1, pp. 417–440, 1990.

[25] J. Han, M. Kamber, & J. Pei, "*Data mining: concepts and techniques*", (3rded.), Morgan Kaufmann, 2011.

[26] R. Wald, T. M. Khoshgoftaar, A. Napolitano, C. Sumner, "*Using Twitter content to predict psychopathy*", in proceedings of the 2012 11th international conference on machine learning and applications—Volume 02, pp. 394–401, Washington, DC, USA.

[27] R. Wald, T. Khoshgoftaar& C. Sumner, "*Machine prediction of personality from Facebook profiles*", in 2012 IEEE 13th international conference on Information Reuse and Integration (IRI), pp. 109–115.