

Personality Classification using Social Media

Prof. Prathmesh Bajare, Supriya Shete, Shrutika Chavan,
Megha Kharade, Soumya Raj

B.E student, Dept. of Computer Engineering,

NBN Sinhgad School of Engineering, Pune,
Maharashtra, India



ABSTRACT

Personality is one of feature that defines how people interact with the other peoples. Personality can be determined as an important element of a person's behaviour. The way in which people interact with world determines their personality. This paper covers the topic of Personality Prediction Using Social Media; This system analyses the personality based on social media comments. In this paper, we proposed which analyses the personality of an applicant. This system will be very helpful for organization and also other agencies who would be recruiting applicants based on their personality rather than their technical knowledge. The personality prediction results are based on users comments using machine learning algorithms like Naive Bayes Algorithm and Support Vector Machine.

KEYWORDS: Support Vector Machine Algorithm, Personality Prediction, Data Mining, Feature Extraction. .

ARTICLE INFO

Article History

Received: 12th February 2022

Received in revised form :

12th February 2022

Accepted: 15th February 2022

Published online :

19th February 2022

I. INTRODUCTION

Social media platforms such as Facebook, Twitter, Google, and Instagram has gained popularity due to ease access throughout the world and user-friendly interfaces to start communicating with others within a short period of time. Each user in these social networking sites (SNSs) is considered as an entity, and each entity is connected with other entities as friends, connections, or followers. While using these SNS's, users are facilitated by many activities, such as posting statuses/tweets, sharing others' posts/retweets, liking others' posts, commenting on others' posts, chatting directly with the friends, and playing online games with the friends. It is evident that from the activities performed by users, online behaviour could be depicted. Understanding users' behaviour may help to identify personality traits. Predicting customers' personalities from digital footprints of social media is a hard challenge as the context of figuring out personality tendencies in social media isn't always trivial. advantageous or negative opinions/opinions about a political celebration. those styles of statuses might also have contextual traits, as other pals of the users can also be worried in posting comparable statuses. thinking about the fashion, user may also publish his/her

political beliefs. customers behave otherwise in social media and real lifestyles.

Conventional techniques of assessment for PP are based on "self-report inventory" [4], in which someone is run a character check (PT). Psychologists frame the questions for these techniques are acknowledged to appropriately return the character rating of a player however be afflicted by the following deficiencies:

- 1) Inefficient - while the experimentation is on a large scale, guide techniques tend to be high-priced in phrases of manpower and material.
- 2) Guide computation is further had to arrive at the outcomes, once the answers of the inventories are lower back.
- 3) A long term is wanted to complete the stock survey. In quick, traditional strategies take time to reach at the final outcomes i.e., the anticipated persona and are much less realistic. as an alternative, on-line questionnaires administered via websites may be used for personality assessments. The user nonetheless wishes to answer diverse questions. latest research undertaken have affirmed that persona tendencies may be gotten automatically

from the text that consumer's wrote. Use of online environments (such as micro blogging sites) ,turned into therefore the point of interest of several researcher's. advancements in IT and unheard of use of social media web sites for social communication has paved a new way to conduct studies on personality and behavior – automatic prediction of character from social media information.

II. LITERATURE SURVEY

1.Paper Name: Social Media Text - A Source for Personality Prediction

Author: P.S.Dandannavar

Abstract ::Social media usage has been on an ever increasing exponential upward push. utilization of social media websites, consisting of Twitter and facebook, for social interaction has additionally turn out to be a famous fashion. it's far anticipated that on an average, round 6,000 tweets are tweeted on Twitter each second. With people spending on an average 35 mins on fb every day, it is also predicted that there are approximately 317,000 status updates on fb in line with minute. those vast volumes of information have powerful facts locked inside them. This information can be analyzed and several purposes. the use of such social media records for predicting user persona is common. Prediction fashions have been effectively constructed that can are expecting several user attributes - age, gender, personality traits, career, political orientation and many others. requirements in character fashions inclusive of the huge 5 model, DISC and the MyersBriggs type Indicator had been the premise for all such persona prediction. A user's social media information can thus be used to expect his/her persona. the primary objective of this paintings is to check the paintings finished for person a prediction using social media data.

2.Paper Name:Personality Prediction of Social Network Users

Author name: Chaowei Li, Jiale Wan, Bo Wang

Abstract : Through weibo users, we extract social data and questionnaire, and focus on how to use the user text information to predict their personality characteristics. We use the correlation analysis and principal component analysis to select the user information, and then use the multiple regression model, the gray prediction model and the multitasking model to predict and analyze the results. It is found that MAE values of the gray prediction are better than the multiple regression model Multitask model, the overall effect of the prediction between 0.8 and 0.9, the overall accuracy of good prediction. This shows that gray

prediction in the user's personality prediction shows a good generalization and non-linear ability.

3.Paper Name: Personality Classification System using Data Mining Author: Sandhya Katiyar

Abstract: Personality is one feature that determines how people have interaction with the outdoor world. character may be defined as a essential detail of a person's conduct. The manner human beings engage with other human beings determines their character. This paper covers the subject of computerized personality category – a machine that analyses the persona of a person based on certain features using facts Mining Algorithms. in this paper, a system is proposed which analyses the persona of an applicant. This device might be beneficial for agencies in addition to different organizations who might be recruiting applicants primarily based on their personality instead of their technical know-how. The character prediction consequences are based totally on big 5 character traits and the classification is achieved the usage of Naive Bayes algorithm and support Vector device.

4. Paper Name: Clustering based Personality Prediction on Turkish Tweets

Author name: Tutaysalgir

Abstract: In this paper, we gave a framework for predicting the person a trends by reading tweets written in Turkish. The prediction model is built with a clustering based method. since the version is primarily based on linguistic capabilities, it's miles language unique. The prediction version uses functions applicable to Turkish language and associated with writing style of Turkish Twitter customers. Our method uses anonymous BIG5 questionnaire scores of volunteer individuals because the floor truth that allows you to generate persona model from Twitter posts. test consequences display that constructed version can predict personality tendencies of Turkish Twitter customers with especially small errors.

5.Paper name: Comparative Analysis of Feature

Selection Algorithms for Computational Personality Prediction From Social Media Author: Ahmed Al Marouf , Md. Kamrul Hasan, and Hasan Mahmud

Abstract:—With the speedy growth of social media, customers are becoming involved in digital socialism, generating a big quantity of textual and picture contents. thinking about the contents together with status updates/tweets and shared posts/retweets, liking other posts is reflecting the web behaviour of the customers. Predicting character of a consumer from these virtual footprints has end up a computationally hard hassle. In a profile-primarily based method, making use of the user generated textual contents might be beneficial to mirror the personality in

social media. using massive range of capabilities of various classes, which includes conventional linguistic features (character-degree, word-degree, structural, and so forth), psycholinguistic capabilities (emotional affects, perceptions, social relationships, and so on) or social community capabilities (community size, betweenness, and so on) can be beneficial to predict persona tendencies from social media. in step with a widely famous personality version, particularly, Big-five-factor model (BFFM), the 5 elements are openness-to revel in, conscientiousness, extraversion, agreeableness, and neuroticism. Predicting persona is redefined as predicting each of those traits one after the other from the extracted functions. historically, it takes massive number of features to get better accuracy on any prediction assignment even though applying feature choice algorithms may additionally improve the performance of the version.

III. PROPOSED TECHNIQUES

1] MACHINE LEARNING APPROACH

Machine learning approach uses supervised, or unsupervised learning, semi-supervised to construct a model from a large training corpus. Machine Learning algorithms are highly efficient in recognizing patterns in datasets that humans cannot even perceive. The use of these ML models can lead to good, more objective, and automated personality assessments. People interact and express their likes, thoughts, feelings, and opinions on social media like Twitter, Facebook capturing their personality traits.

2] SUPPORT VECTOR MACHINE

Help of Vector machine (SVM) is a classifier is defined by using developing a separating hyper plane. it is used for linearly separable via binary set. The principle purpose of assist vector gadget is to categorise the training vectors into classes and to lay out a hyper plane. it is also used for multi-dimensional datasets and this data points are represented as vectors. The hyper plane which offers maximum margin is considered because the high quality hyper plane. Minimising the above time period will maximize the separability and this offers the largest margin right here. SVM may be very powerful in excessive dimensional areas, and we've got specific kernel capabilities for diverse selection functions.

PROPOSED METHODOLOGY:

The operating of the device is widely divided into following sections:

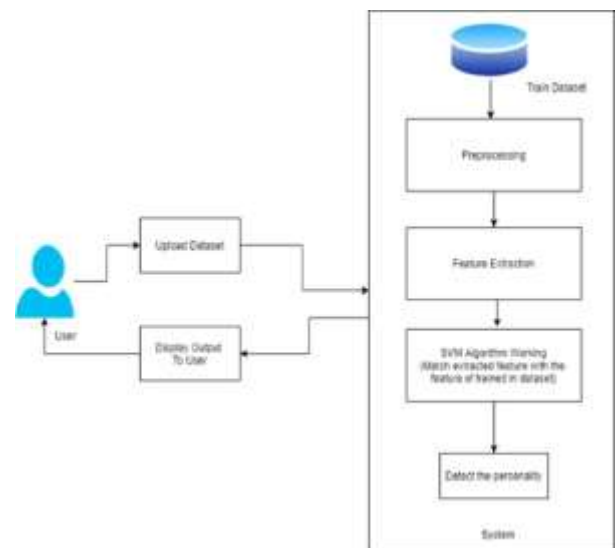


Fig.1. System Architecture

1. Data Extraction

This section entails figuring out and amassing facts units deemed appropriate for the utility being evolved. Twitter has been the most important source for social media information. records sets for education and testing are accumulated from Twitter using suitable API's. different resources of records consist of the publicly to be had datasets viz, my persona and essays.

2. Data Pre-processing

That is a very critical segment inside the pipeline as it decides the efficiency of the alternative steps down in line. Pre-processing entails the usual steps of Case Conversion, stop-phrases elimination, Punctuation removal, Stemming, Lemmatization, POS Tagging.

3. Feature Extraction & Selection

After the applicable tweets are fetched and preprocessed, capabilities relevant for prediction are extracted and selected. person's profile data is the maximum common function set utilized by most researchers. among different capabilities used are – range of followers, number of following, twitter posts, linguistic capabilities and so forth. The capabilities extracted are depending on the prediction to be made.

4. Classification

This segment includes modules for version education, checking out and prediction. After the selected model has been skilled and examined using distinctive statistics units , an unseen (new) dataset is presented to the educated model for prediction, so that it will classify and are expecting the persona trait of recent consumer(s). several well-known algorithms have been utilized by authors like help Vector machine, k Nearest Neighbour, Multinomial Naive Bayes, Naive Bayes, multi challenge regression, incremental regression set of rules and many others.

UML DIAGRAM:

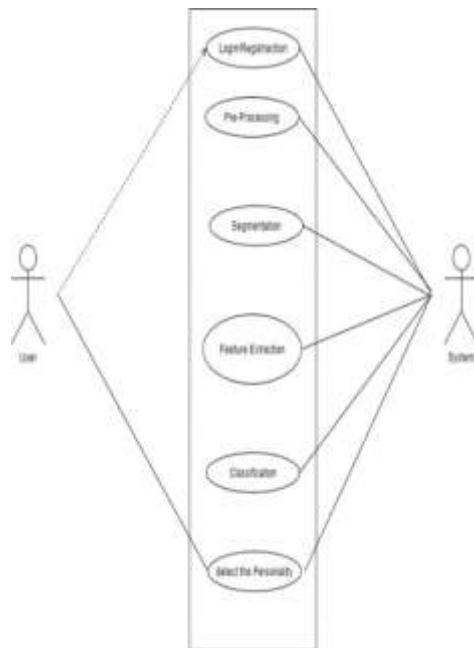


Fig 2:Use Case Diagram

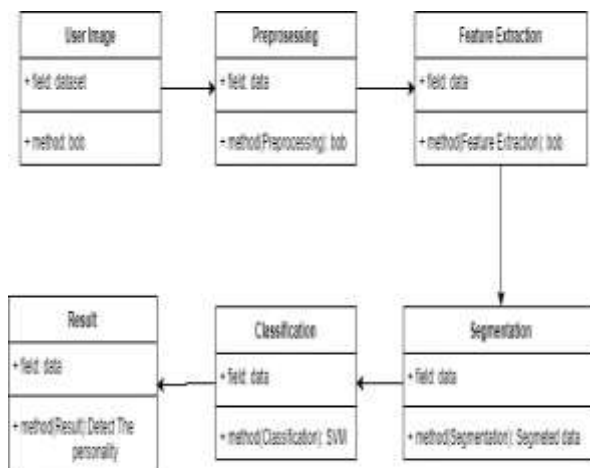


Fig 3: Class Diagram

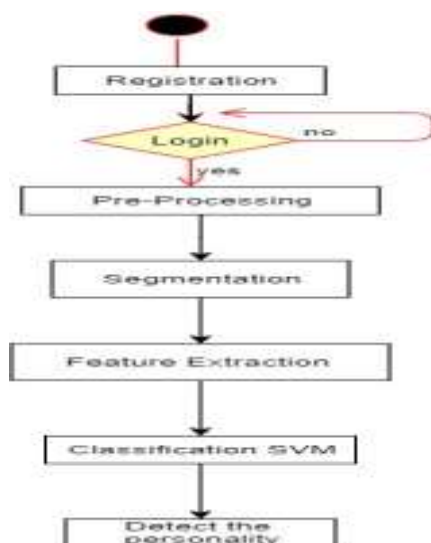


Fig 4 : Activity Diagram

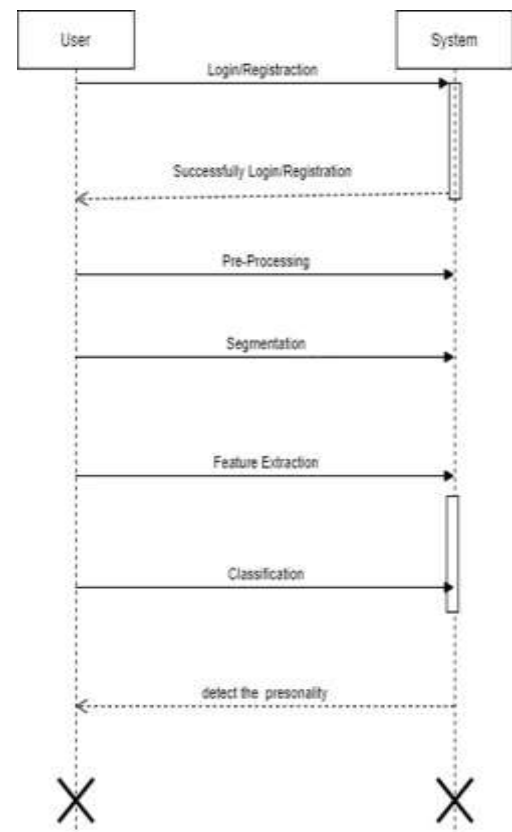


Fig 5: Sequence Diagram

IV. FUTURE SCOPE

One of the test areas that requires a lot of determination is to predict personality based on communication data. A lot of work has been done but it still needs to be done to increase the accuracy of the prediction. Therefore, it needs to be upgraded to various system components such as Algorithms, Features Extraction, and Data set etc. To get the right mental health status, offline personal evaluation should be done for users to be given proper care. The retrospective algorithm designed to predict personality can be used to increase accuracy by providing inputs as semantic features. As the amount of data increases it becomes more difficult or difficult to record data, thus unchecked reading can be used to predict personality by using external information and thus compiling text. A big data set can help improve accuracy and help recommend users for services, movies, music etc. Personality varies over time, so the data collected should cover all of the previous posts.

V. CONCLUSION

As we know, the base of social media is basically a graph. The connections between the nodes and impact on them due to social media interactions could be reflected through the SN features. We have designed and performed experiments utilizing the linguistic features as well as SN features. To the best of our knowledge, we have used the most number of

features to compare the performance of the feature selection algorithms to predict personality traits.

REFERENCES

- 1] M. R. Barrick and M. K. Mount, "The big five personality dimensions and job performance: A meta-analysis," *Personnel Psychology*, vol. 44, no. 1, pp. 1–26, 1991.
- 2] S. Rothmann and E. P. Coetzer, "The big five personality dimensions and job performance," *SA Journal of Industrial Psychology*, vol. 29, no. 1, pp. 68–74, 2003.
- 3] J. F. Salgado, "The big five personality dimensions and counterproductive behaviors," *International Journal of Selection and Assessment*, vol. 10, no. 1, pp. 117–125, 2002.
- 4] J.W. Lounsbury, R. P. Steel, L.W. Gibson, and A.W. Drost, "Personality traits and career satisfaction of human resource professionals," *Human Resource Development International*, vol. 11, no. 4, pp. 351–366.
- 5] J. W. Lounsbury, N. Foster, H. Patel, P. Carmody, L. W. Gibson, and D. R. Stairs, "An investigation of the personality traits of scientists versus nonscientists and their relationship with career satisfaction," *RD Management*.
- 6] V. Ariyabuddhiphongs and S. Marican, "Big five personality traits and turnover intention among thai hotel employees," *International Journal of Hospitality Tourism Administration*, vol. 16, no. 4, pp.