AUTOMATED PERSONALITY CLASSIFICATION BASED ON DATA MINING TECHNIQUES – A CONCEPT SURVEY

[1] PROF. MANDAR S.JOSHI, [2]MS.SANIYA F.FADNAIK , [3]MS. AISHWARYA S.SHETYE ,[4]MR. JYOTI S.NACHANKAR
[1]Assistant Professor, [2]Student, [3]Student, [4]Student
[1,2,3,4]Department of Information Technology,
[1,2,3,4]Finloex Academy of Management and Technology, Ratnagiri, Maharashtra, India

Abstract : *As Understanding personality type can help you to understand your preferences and the preferences of other people and how or why these might be different. Personality types are useful for recognising how people lead, influence, communicate, collaborate, negotiate business and manage stress. This project is helpful where we have data related to personal Behaviour. This personal behaviour data can be useful for identifying person based on his personality traits. The personality characteristics will be already stored in database. Later when user enters his personality characteristics his personality is examined in database and system will detect the personality of user. This system will use Naïve Bayes and SVM algorithms. This system is helpful for organization or companies to hire their employees by observing behaviour of employees and also in various e-Commerce sites to identify the customer behaviour.*

Index Terms – Personality, Behavior.

I. INTRODUCTION

Personality is unique combination of patterns that influences behaviour, thoughts, Motivation and emption in human being. The personality of an individual can be defined as a set of features that induces a tendency on the behaviour of the individual; this tendency is stable through time and situations. This system will identify personality of a given person which will provides hints about how he would probably react when facing different situations. Identifying a user's personality can contribute to know, for example, his potential needs in different contexts. Therefore, this automated personality classification system may benefit from having models of users personality to adapt their behaviour accordingly. There is a wide variety of domains in which this can be useful, i.e., assistive technologies, e-learning, e-commerce, health care or recommender systems, among others. Methods like survey, interview are very time consuming and as people are prone to prejudices which affect the accuracy and cannot provide us result as correct result as this personality classification system.[4]

II. LITERATURE SURVEY

1. Understanding Personality through Social Media

The relationship between human language on Twitter and personality traits is studied. Specifically, we want to know how linguistic features correlate with each personality trait and to what extent can we predict personality traits from language. Here they gather personality data from Myers-Briggs personality test which contains thinking, feeling, sensation, intuition, introversion, extroversion, judging and perceiving. Also, they collect 200 most recent tweets from users with personality values. This system design three categories of feature, namely bag of n-grams, Twitter POS tags, and word vectors. Analysis of these features provide insights of language use for different personalities. [3]

2. Predicting personality types from user comments

This project explores the significance of personality types in determining the use of language, particularly online comments in forums. The machine learning task is to predict personality type of a user based on comments on an online forum. The project has far-reaching implications if you consider the ever-increasing availability of user data on online forums and social media platforms. The words that the feature selection ranked as significant are greatly insightful to the nature of what it means to be of a "Thinking (T)" or "Feeling (F)" type. It is amazing to see that the following words, which are less explicitly related to the "feeling" dichotomy but still one of its characteristics, are predictors of the dichotomy: love, beautiful, really, heart, thank, hope, haha and song. For the "thinking" dichotomy, the results are less interesting because the difference in the frequencies for each output class are not as significant.

3. Mining Facebook Data for Predictive Personality Modeling

Initial investigation in personality modelling based on Facebook data are encouraging evidence that by selecting the most indicative features the precision of the classifiers could be improved. Extracting qualitative knowledge from the large quantities of data is just the beginning of our search for meaning and plausible explanation of personality-determined social network activities. This paper explores the feasibility of modeling user personality based on a proposed set of features extracted from the Facebook data. [2]

4. Personality Classification Based on Twitter Text

Social media is a place where users express themselves to the world. Posts made by users of social media can be analyzed to obtain their personal information. This experiment uses text classification to predict personality based on text written by Twitter users. The languages used are English and Indonesian. Classification methods implemented are Naive Bayes, K-Nearest Neighbors and Support Vector Machine. Testing results showed Naive Bayes slightly outperformed the other methods. [1]

III. EXISTING SYSTEM

In Existing system, Facebook or Instagram profiles, Twitter language or text is used to identify/judge the personality of user which may not give the correct result. Judging the personality on social media text cannot give us the accurate result. However the person can be misinterpreted or his actual personality can be ignored. Methods like survey, interview are very time consuming and as people are prone to prejudices which affect the accuracy and cannot provide us with accurate result. While taking interview the interviewer may not have correctly understood a word or phrase you used because of your pronunciation, your accent, or inflection—which may have led them to (mostly negatively) misinterpret your meaning. Therefore, such ways results in misunderstanding the personality of person.

IV. BACKGROUND OF DATA MINING SYSTEMS

There are different types of data mining methods like classification, Clustering etc. As this system works on classification process we dont use clustering methods. The input data to a specific categoryis mapped by classifier. Naive Bayes and Support Vector Machine are the classifiers we are using in this system.

1. Naive Bayes Classifier.

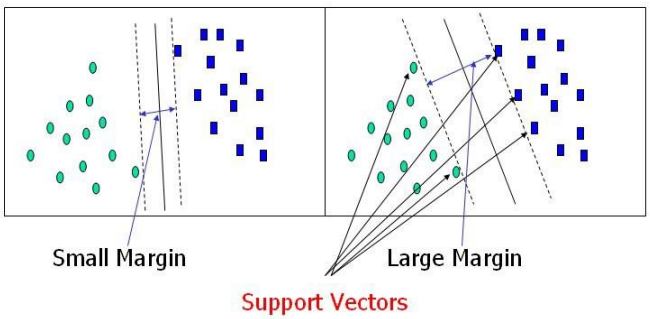
Naive bayes classifier is based on Bayes Theorem. Naive Bayes is a collection of classification algorithm.Bayes' Theorem Bayes' Theorem finds the probability of an event occurring given the probability of another event that has already occurred.

Bayes' theorem is stated mathematically as the following equation:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

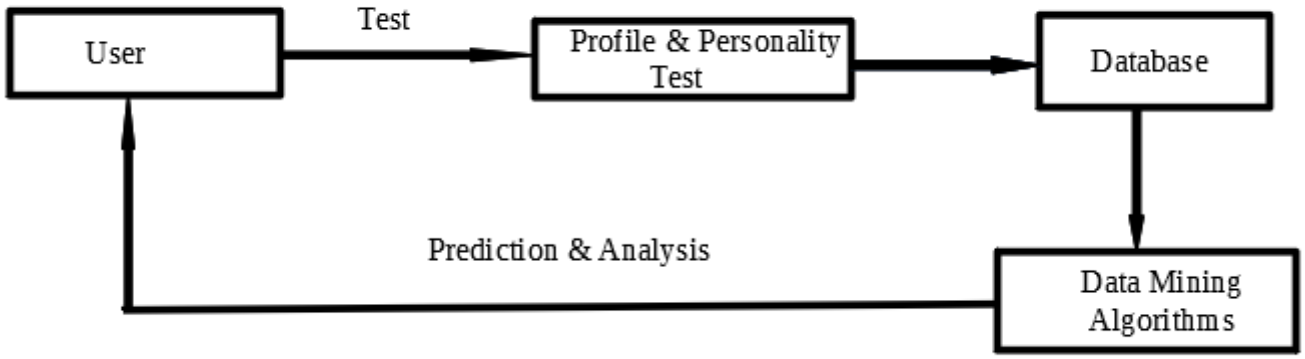
2.Support Vector Machine

Support vector machine algorithm finds a hyperplane in an N-dimensional space that distinctly classifies the data points. Support Vector Machine map the input data to some high dimensional space, where the data can be linearly separated, thus providing great classification. One of the bottlenecks of the SVM is the large number of support vectors used from the training set to perform classification operation.



V. PROPOSED SYSTEM

Personality is the product of social interaction in group life**.** People have different types of personalities because individuals are not alike**.** It refers to the habits, attitudes as well as physical traits of a person which are not same but have vary from group to group and society to society, everyone has personality, which may be good or bad, impressive or unimpressive. Personality classification is determined by the patterns of **thinking** and behaviour that develop over time. Five characteristics of different individuals commonly known as big five characteristics namely, **openness, neuroticism, conscientiousness, agreeableness and extraversion** are stored in a dataset and used for training. Based on this training, the personality of individuals are predicted using data mining concepts. The personality characteristics will be stored in database. Later, when user enter his personality characteristics his personality is examined in large pre-existing database and system will detect the personality of user. The system will detect users personality based on previous data stored in database. This system will examine personality of user based on personality traits mentioned by the user. And will provide user with various features relevant to his personality. The relation between the personality and user behaviour is tested. This system will help advertisement people to market their product based on personality of the user which in-turn provides income to the firm who is using this system. This system can be embedded with social sites, as many user can buy and sell their products using these social network



.

VI. CONCLUSION

In this paper we study the relationship between user and his personality. This personality classification system is very helpful in other social networks where personality plays a vital rote like matrimonial sites, E-commerce sites. The techniques employed for detection of personality is machine learning based approach like naïve Bayes and SVM. The available personality-data dataset is discussed. Thus the personality of person is automatically classified by system.

Further improvement can be done by using more accurate dataset to improve the accuracy also this can be helpful for career guidance module where, if user has good speaking and convincing skills he can approach towards marketing.

VII. REFERENCES

[1]. BAYU YUDHA PRATAMA,RIYANARTO SARNO, "Personality classification based on twitter text," *International Conference on Data and Software Testing ,2015*.

[2]. DEJAN MARKOVIKJ, SONJA GIEVSKA, MICHALKOSINSKI, DAVID STILLWELL, "Mining Facebook data for Predictive personality Modeling," *Association for the Advancement of Artificial Intelligence,2013*.

[3]. YILUN WANG, "Understanding Personality through social media," *International of computer Science standford University,2015*.

[4]. MANASI OMBHASE, PRAJAKTA GOGATE, TEJAS APTIL, KARAN NAIR, PROF. GAYATRI HEGDE, "Automated Personality classification using Data Mining Techniques," *International Conference on Data and Software Testing , April 2017*.

[5]. MAHESH KINI, SAROJA DEVI, PRASHANT G DESAI, NIRANJAN CHIPLUNKAR,"Text Mining approach to classify Technical Research Document using naive bayes", *International journal of Advanced Research in computer and communication engineering,Vol $,issue 7, July 2015*.

[6]. DURGESH K.SRIVASTAVA, LEKHA BHAMBHU, "DATA Classification using Support Vector Machine," *Journal of Theoretical and Applied Information Technology,2005*.