Personality Prediction

Mansi Pable
Department of Computer Engineering
Ramrao Adik Institute of Technology
Navi Mumbai,India
mansipable123@gmail.com

Shital Nehete
Department of Computer Engineering
Ramrao Adik Institute of Technology
Navi Mumbai,India
shital.nehete@gmail.com

Jeet Choudhari
Department of Computer Engineering
Ramrao Adik Institute of Technology
Navi Mumbai,India
jeetchoudhari99@gmail.com

***Abstract*— In recent years, the problems of mental health of people are increasing day by day due to increase of the stress levels. Employment of less skilled people results is reducing of the outcome of the organization. To solve these problems, the personality of the person plays a major role. So, this project aims at predicting the personality of the person in Big-Five traits. The prediction can be done by 3 processes i.e. text, questionnaire and social media. The machine learning model used to predict the personality uses myPersonality dataset. It uses tf-idf vectorization to convert words to vectors and Random Forest Algorithm for all the prediction. Social media also helps to compare the personality of the person with the friends it has on it. Radar graph is used to show and compare the personality division on the web app. Web app provides a proper interface for the processes and display of the personality for easy access.**

Keywords—Personality Prediction, myPersonality,OCEAN

I. INTRODUCTION

The traits, set of etiquettes, emotional patterns define one's personality which plays an important role in taking major life decisions. The capability of personality features to forecast the life upshots have been interrogated due to the presumed repercussion of personality. Predicting personality with the help of OCEAN model can be obliging in order to screen a particular person or diagnose metal health, or to improve his/her own personality. As when we acknowledge our own personality, we can question our requirements also reform our behavior.

Social media is the social interlinkage among people in which there is creation, sharing or exchange of information and ideas in virtual communities and networks. Nowadays, people use more and more time in social media, every time there is a flood of information creation and dissemination. This information includes user's social behaviour, user generated texts and language habit, to some extent reflects user's personality. This project tries to use user generated information on social network which is easy to get to predict user's personality. This Big 5 Personality Traits, also known as OCEAN stands for Openness, Conscientiousness, Extroversion, Agreeableness, Neuroticism. According to analysts, every single person is born with these personality traits. Individuals who overlay high in the first trait are often eccentric. The more organized you are, more you rate in Conscientiousness. "I am the life of the party", can be the best example for Extroversion. Low Agreeableness shows logical behaviour. Stress, mood comes under Neuroticism. Thus, with the help of these five personality traits listed out under OCEAN model we can easily predict individual's personality.

II. LITERATURE SURVEY

Personality measure is done for better hiring decisions within the selection process. Data from over 46,000 job applicants who completed a web chat interview that also included a personality questionnaire was collected. Using NLP and ml a regression model was built to infer HEXACO trait values from textual content.[2]

Facebook's User information can be used to predict the human personality and behavioural learning. Facebook has the largest users reaching 1.8 billion users and 800 millions spending 40 minutes per day. The status updates on Facebook are important and the dataset (250 dataset from myPersonality) was made according to the updates and scores was calculated. The BIG Five Model traits is used for predicting the personality. The old machine learning models as well as the deep learning architectures are used for the prediction and the performances are compared. [9]

Researches were done for textual data for predicting personality. Data was converted into numbers using Word2Vec model and non-neural sentences are filtered. Then, given to deep CNN for prediction which was trained on essay dataset of 2,468 people. [4]

III. PROPOSED WORK

We propose a Personality Prediction web app which predicts personality based on the BIG 5 i.e. OCEAN model and rates the personality for his Openness, Conscientiousness, Extroversion, Agreeableness, Neuroticism of that person. This app consists of 3 sections namely Text Predictor, Questionnaire, Facebook.

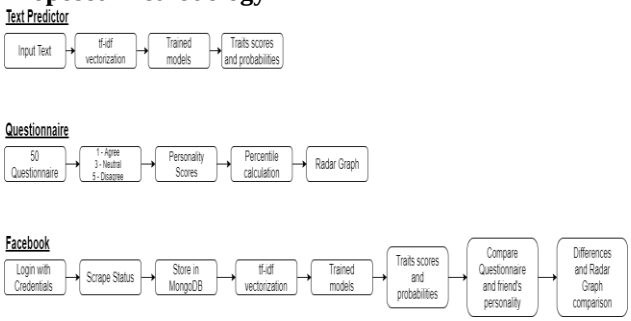
In first section, user interface of the web app gives option to input a text which is provided as an input to the OCEAN model for predicting. This is followed by the scanning those text sentences in the myPersonality sample dataset (250 users, 9917 status updates).

In Questionnaire, User undergoes 50 questions regarding the O.C.E.A.N model, based on which radar graph is implemented. It can also be termed as Psychometric Assessment. The options given for the questions consists of Disagree, Neutral, Agree which correspond to 1, 3,5 respectively, based on which the personality score of a particular user is calculated. This can be further used for comparing the personality among social network.

For comparing the personality among the Facebook connections or any social media site network tab is helpful. Comparison between user's personality score evaluated from the Questionnaire section and the model based prediction is used in this section. With the help of Facebook Web Scrapper the statuses of user's Facebook network is extracted. For every connection personality is

predicted and compared with the user's personality based on the personality points.

Proposed Methodology



1. Text Predictor:

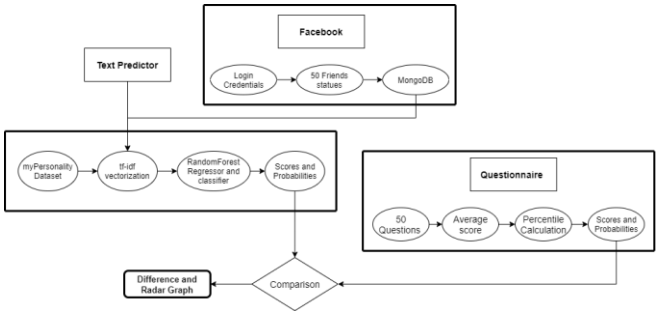
The Personality of a person can be predicted by inputting any text with the help of this tab. The user interface of the web app gives option to predict the personality based on the text. The text is converted into vectors using tf-idf vectorization then, five models for the Big-Five personality traits are trained using RandomForest Algorithm. The vectors are given input to the models and the score as well as the classification is obtained.

2. Questionnaire:

In this section, user undergoes 50 questions regarding the O.C.E.A.N model, 10 questions for each personality traits. The score is evaluated based on the answers to the questions and the radar graph for user's personality is implemented. It can also be termed as Psychometric Assessment.

3. Facebook:

For comparing the personality among the Facebook network or any other social media site network this tab is helpful. Comparison between user's personality score evaluated from the Questionnaire section and the same models trained for text prediction are used here for evaluation of the scores of the Big-Five personality traits. The comparison is shown in the radar graph too on the user interface.



The web app consists of 3 sections Text Predictor, Questionnaire and Facebook. The model is trained on myPersonality dataset. The Models produce a predicted personality score, using the regression model, and a probability of the binary class, using the classification model, for each personality trait. In the first section personality of a person can be predicted by inputting any text with the help of text predictor. This section is totally based on the models trained earlier. The user interface of the web app gives option to predict based on the text. The models trained will give a score on each of the 5 personality traits.

In the Questionnaire section user undergoes 50 questions Big 5 personality test regarding the O.C.E.A.N model, User will get options like agree, disagree and neutral on basis of that certain score will be generated, which then displays your corresponding percentile scores for each personality trait and based on which radar graph is implemented. It is like a psychometric assessment. The score generated in this section is used to compare personality obtained in the Facebook section.

Facebook is used for comparing the personality among the Facebook or any social media site network. Here users credentials are provided in a config file to access Facebook, through automation the statuses of user's friends are extracted and stored in database. The model which is trained earlier is used to give a prediction on the data which is stored in database. This data is also used for comparison between user's personality score evaluated from the Questionnaire section and the model-based prediction.

IV. RESULTS

1. Text Predictor:

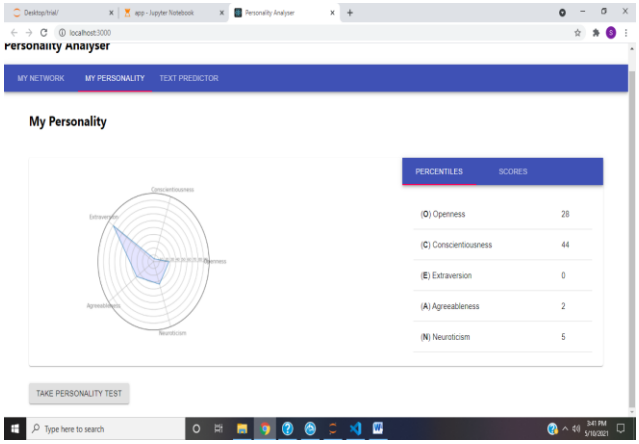
First, the text inputted by the user is converted into vectors using tf-idf vectorization. Second, the vectors of the text are provided as inputs to the five models trained on myPersonality dataset by RandomForest algorithm. Third, the scores and probabilities are evaluated from the models and are displayed on the screen according to the respective personality traits.

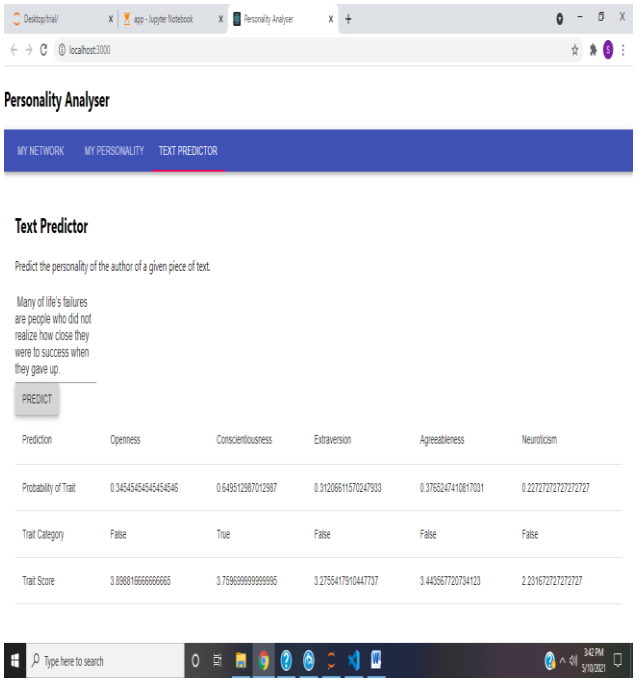
2. Questionnaire:

First, the questions are prepared according to the Big-Five personality traits, 10 questions for each personality trait. Thus, the user has to answer 50 questions for the evaluation of the personality. Second, the answer the questions are stored in the dictionary and the score is averaged according to the answer. Third, the percentile is also evaluated based on the user's answer and the dataset of values for each personality trait. Forth, the score and the percentile is displayed on the user interface.

3. Facebook:

First, the user provides credentials for the facebook account. Second, selenium is used for automation for extracting the facebook statues of 50 friends from the friend list provided in the Friends section on the Facebook account. Third, the extracted statues are stored in MongoDB. Forth, the statues are used to evaluate the personality of each friend according to the RandomForest model. Fifth, the calculated personality is used to compare with the user's personality in questionnaire section. Sixth, the visual comparison can be done user by clicking on the compare button on the user interface. The radar graph and the difference between the scores is clearly visible on the user interface.





The Selenium automated system will give user a better analysis for the personality as it compares the scores with others. If the user feels one way isn't sufficient for prediction, he can go for the other. This system consists of questionnaire as well as text prediction of personality and also, the user can compare his personality with his Facebook's friends. Combining these methods can help the user to get the result on only one platform instead of searching for another.

The comparison tab can provide self-analysis to the user for the field he/she is lacking with respect to another.

The listed algorithms and their performances for 5 traits are compared and observed that Random Forest outperforms the other 5 algorithms with accuracy of 80 to 85%.

V. CONCLUSION

Personality is a very important aspect nowadays. In this growing world, the person is identified by the personality he/she carries so, this project aims to decoding the personality in 5 essential traits. The score obtained from the prediction are the percentage of the traits present in the person and it can be used further for many purposes. One's personality prediction can show the job satisfaction and performance in the corporation also. It can be used in the health care sector too for psychological analysis. The recommender systems also rely on the personality of the person and help as per his interests. There are many more applications such has marketing and advertisement. Hence, this project has growing demands and many aspects together.

Future Work is necessary for more accurate results as the dataset is very small. The limitations of the dataset affect the accuracy of the model. For questionnaire, the number of questions are limited and hence, the personality is accurately predicted. For social networks sites, there are large number of fake accounts which affects the accuracy and prediction and the access to the accounts requires the credentials that may be a concern. For text, the text the person enters might be influenced by the fact that it I getting evaluated for personality so, the chances of the person lying or not writing what he/she feels are very high which will result is improper prediction.

VI. REFERENCES

[1] Majid Ramezani, Mohammad Reza Feizi Derakhshi, Mohammad-Ali Balafar, Meysam Asgari-Chenaghlu, "Automatic Personality Prediction; an Enhanced Method Using Ensemble Modeling", 2020

[2] Madhura Jayaratne, Buddhi Jayatilleke , "Predicting Personality Using Answers to open-ended interview questions", 2020

[3] Aditi V.Kunte, Suja Panicker , "Using textual data for Personality Prediction: A Machine Learning Approach", 2019

[4] Abdur Rahman, Asif Al Faisal, Tayeba Khanam, Mahfida Amjad, Saeed Siddik, "Personality Detection from Text using Convolutional Neural Network," IEEE, 2019.

[5] Guanqun Sun, Ao Guo, Jianhua Ma and Jianguo Wei, "Personal Trait Analysis Using Word2vec Based on User-generated Text," IEEE, 2019.

[6] Raad Bin Tareaf, Seyed Ali Alhosseini, Philipp Berger, Patrick Hennig, Christoph Meinel, "Towards Automatic Personality Prediction Using Facebook Likes Metadata," IEEE, 2019.

[7] Mariam Hassanein, Wedad Hussein, Sherine Rady Tarek, F. Gharib, "Predicting Personality Traits from Social Media using Text Semantics," IEEE, 2018.

[8] Raad Bin Tareaf, Philipp Berger, Patrick Hennig, Christoph Meinel, "Personality Exploration System for Online Social Networks: Facebook Brands As a Use Case", 2018

[9] Tommy Tandera, Hendro, Derwin Suhartono, Rini Wongso, and Yen Lina Prasetio , "Personality Prediction System from Facebook Users", 2017