**Title**

Personality Prediction

**Rationale and Gap Analysis**

Over the last few years, the use of social media is increasing day by day with the rapid advancement of technology. Face-to-face communication of people is also decreasing as people tend to prefer to communicate through either smartphones or online communication platforms. For the last year due to the pandemic caused by the novel Coronavirus, online communication is preferable to physical communication. Studies in the area of psychology have also shown the relation between the personality and behavior of a person. People are escalating day by day due to an increase of the stress levels, thus the problems of mental health of peoples especially the employees working in organizations result in reducing the outcome of the organization. Thus, difficult in predicting the personality of people at the workplace as other media platforms. To solve these problems, the personality of the person plays a vital role. Also, Social media is the best platform where most people share their thoughts regularly. So, this project aims at predicting the personality of the person in Big-Five traits with social media as well as a questionnaire. The advanced technology nowadays helps the user analysis and hence, reform the behavior or personality.

This project aims at combining three different approaches (questionnaire, text, Facebook statuses) for predicting the personality and giving the user flexibility for analyzing the personality. The user can also use all three methods and average the personality score. It gives the personality score for BIG 5 traits (Openness, Conscientiousness, Extroversion, Agreeableness, Neuroticism). It can help better analysis for the user and can give more precise results as it combines three different methods than the previous methods used to predict the personality. The user can get an overall judgment of the strengths and the weakness of his/her personality. Averaging the three different methods can give more precise outputs. The major goal of the system is to predict personality accurately.

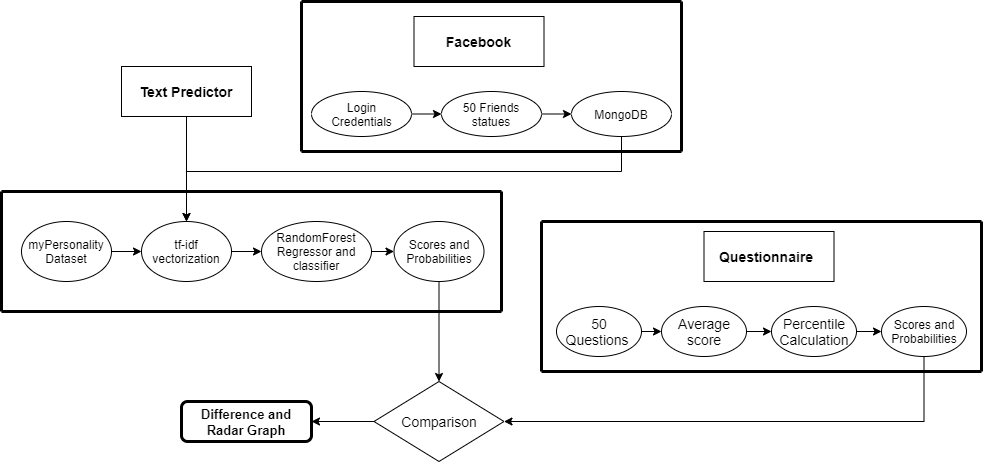
The systems used mainly two models for predicting the personality (BIG 5 and HEXACO- six-factor personality framework). The ways for predicting personality are vast. The similarity between the model vectors and the user inputs can be used to predict personality.[7] Similarly, artificial intelligence can analyze the trait scores for the personality models.[4]

There are existing techniques available such as facebook’s like metadata, [6] tweets, user-generated text to predict the personality of the user. Social networking sites like Facebook don’t directly provide needed data to researchers due to privacy reasons. The accuracy of the analysis can have a great impact due to the millions, trillions of fake users on Facebook. Dynamic changes in likes and number of friends can affect the accuracy [4][6]. The accuracy is highly suffered due to the quality and small database [1]. The aim can be to generate a large dataset as the accuracy for the models will get much better then. Apart from text, speech, images, and videos shall also be considered [2].In the majority of available systems, non-linguistic features and Semantic level analysis are not considered.

**Objectives**

1. The objective of this project is to predict personality based on the 5 features (BIG5 – O.C.E.A.N.) which stand for Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism.
2. To diagnose psychological problems.
3. Focusing on strengths, weaknesses, temperament, and style of leadership.
4. Predict personality at the workplace as well as for personal inventories.

**Research Design and Methods**

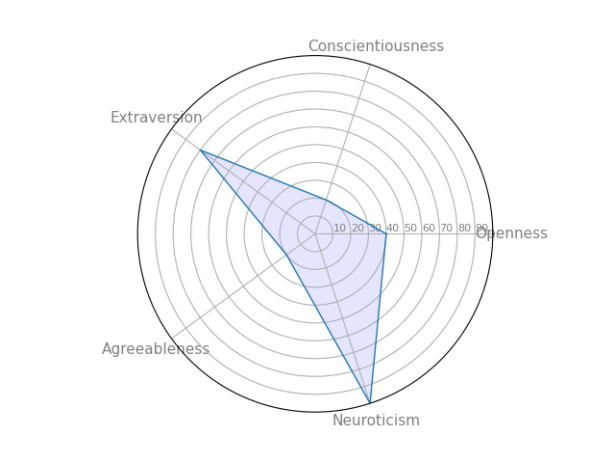
****

**Fig 1. Proposed Methodology**

We propose a Personality Prediction system that predicts personality based on the BIG 5 traits, i.e., OCEAN model, and rates the personality for his Openness, Conscientiousness, Extroversion, Agreeableness, Neuroticism of that person. The dataset named as myPersonality is constructed with 250 users 9917 Facebook’s status updates and labeled according to the BIG 5 traits.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Value | O | C | E | A | N |
| YES | 176 | 130 | 96 | 134 | 99 |
| NO | 74 | 120 | 154 | 116 | 151 |

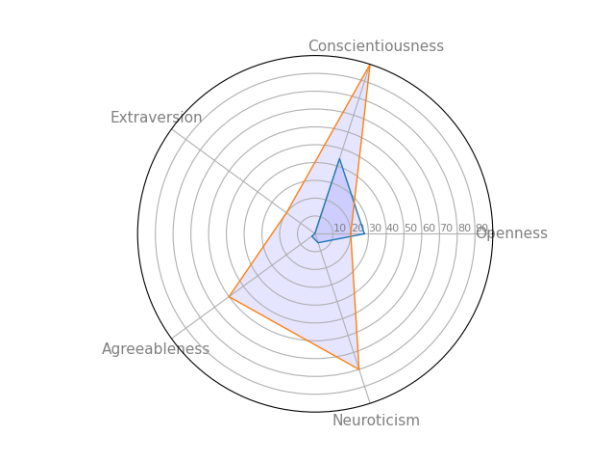
This project will have three ways for calculating and comparing the personality i.e., text, questionnaire, and Facebook. The user will input a text and will be converted into vectors by tf-idf vectorization. The model trained previously will be imported and each trait score will be calculated. Then, the scores can be further converted into percentiles and probabilities. These can also be represented in a radar graph for visualization shown in figure 2. The dataset is further used for training the Ensemble Model for bagging i.e. Random Forest algorithm. The statuses are pre-processed and converted into vectors by tf-idf vectorization and then, random subsets are created by bootstrap. There will be 200 decision trees for regression. The trees will have a minimum sample split of 2. The model for each personality trait will be saved as a pickle file for further prediction. The proposed methodology is shown in figure 1.



**Fig 2. Radar Graph**

The questionnaire, a set of 50 questions will be prepared for personality assessment, each trait will have 10 questions each. The score will be 1 for agree and 5 for disagree and the user will rate the question or the situation according to their preference. The scores entered by the users will be averaged and percentiles probabilities will be calculated. This is also called psychometric assessment. Here, the trained models will not be used because they are trained for textual data.

Facebook statuses can be used for determining the personality of the user. The credentials for the user’s account will be used for extracting the statuses. The statuses extracted will be converted into vectors and will be provided as input to the earlier trained models. The models will give the scores for each status and then, average the score. Similarly, the statuses of friends of the user will also be vectorized and given to the model for prediction. This will help the user compare their personality with the friends or anyone they want. The differences between the scores will be visible on the radar graph as shown in figure 3. This will help the user analyze the personality properly.



**Fig 3. Radar Graph with comparison**

**Preliminary Work / Survey**

This section summarizes the work done in the literature for predicting the personality of a person.

1. Five different basic methods in ensemble modeling are introduced which incorporate term frequency vector-based, ontology-based, enriched ontology-based, latent semantic analysis-based, and deep learning-based methods. Essays Dataset which is written by psychology students has employed altogether the methods. Further Hierarchical Attention Network is employed [1].
2. Data from over 46,000 job applicants who completed a webchat interview that contains 7 open-ended questions and 40 self-rating questions was collected. Using NLP and ML, a regression model was built to infer HEXACO trait values from the textual content. Frequency-inverse document frequency (TF-IDF) with Latent Dirichlet Allocation (LDA) used [2].
3. Machine learning algorithms had been used namely: Multinomial Naïve Bayes, AdaBoost, and LDA to predict personality based on features collected [3].
4. The personal Trait Matrix model is used for personality prediction. Words are extracted using LDA or lexicons and then converted into a vector using word2vec to form a matrix for different individuals. They scaled diversity on effect and social interaction with personal traits [5].
5. Facebook’s hierarchy to categorical pages as features and give it as input to ML model for predicting Big-Five personality traits scores. It established a relation between likes metadata and mapped it to feature to calculate the big-five scores.[6].
6. Measuring similarities between user’s text and vectors representing different personality traits. The score is presented to how much user’s text is similar to the personality vector using vector space model [7].

**Expected Outcomes**

This project will help users in analyzing the personality in three different ways. Hence, providing this platform to users will help them improve the mental health.

The text personality prediction will show the user at the time itself the weakness and strengths in the personality from the text they enter. It also helps in analyzing and improving the weakness. It will also help the psychologist to understand the patient’s health and diagnose properly.

The questionnaire can help companies analyze a candidate before the recruitment process. If they do not stand up for organization requirements, they will train them in a specific way so that they can have an easy way through dealing with and maximizing the candidates' efficiency. The margins for the scores will be decided by the organization. This will help them from in increasing the outcomes.

Facebook statuses prediction helps a user know the impact they are having on society. The user updating statuses on social media not only affects the user’s thoughts but also the people reading them. Comparing the personality with someone close to you or a well-known name can help you deal with your weakness or strength over them. This way you can know without help that what features you’re lacking and you need to improve. The comparison will help the user become a better personality by working on the weakness.

**Benefits to the society**

The main aim of the project is to predict the personality based on 5 traits i.e. O.C.E.A.N. The predicted personality can be used for recommender systems, job satisfaction and requirement details, educational purposes, psychological, and much more. So, this project aims to help the health care sector, educational sector, employment sector.

1. Diagnose mental health problems.

In case anyone is facing a mental health issue, this can be helpful in some way for the person giving treatment to the victim to know the exact issue. This can speed up the process of curing the issue.

1. Screen candidates based on predicted personality with the help of a questionnaire:

During the placement activities, to predict a candidate's personality, to check whether that candidate is suitable for that particular job role or not, this project can be used in which personality can be predicted based on five personality traits which are sufficient enough to predict one's personality. The personality test consists of 50 personality-based questions having scores associated with each option and based on the option which the user will be selecting, probabilities or the scores will be generated. Hence this can be very helpful in the hiring of any job role, especially for the HR department. A person can also take a personality prediction test to improve his performance to lead to good results.

1. To improve his own personality by comparing with others

In day-to-day life, the general-purpose use of this app can be comparing the user's personality with others like his friends, colleagues, etc. The My Network section given in this app can be very useful in this case where the user can compare his personality (taken from the questionnaire section) with his Facebook network. The icing on the cake is that the results are displayed in the form of a radar graph.

**Future Scope**

Future Work we can increase the dataset as well as the questionnaires so that the personality is accurately predicted. For social network sites, there is a large number of fake accounts which affects the accuracy and prediction.

**Limitations**

1. Smaller dataset - The myPersonality dataset is small as compared to the wide range of applications of the project, it contains data of (250 users, 9917 status updates).

2. Semantic level analysis (terms and topics) is not considered while designing the model. Other types of features, such as the use of parts of speech (POS), readability, formality, use of emojis can further increase the accuracy.

**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