

Github URL: https://github.com/nevilshah444/CMPE-257/tree/main/Assignment_5

Google Colab link:

https://colab.research.google.com/drive/13woQBGzTnymOWgLPqskGK438jiBW-0r0#scrollTo=E_3ljh8FtVUY

FRACTAL on Project

Everytime to find a solution we have to divide a problem in it's possible ways. The main theme here for our project is to detect the behavior of different personalities to make it easier for people to understand other people. It can be understood as an application that can be used by the HR department to know about people's mindset and nature and whether they can be a better fit for the company or not.

Problems Identifying on:

Business Task:

- To identify a person's behavior to predict an approach of theirs on a task.
- To identify a person's ability on handling the scenarios and pressures. In order to conclude the amount of assignments can be concentrated on.
- Can be utilized in various fields with a few customizations. So, to make it more precise on their mental health.

ML Task:

We will be classifying people on the basis of their response in particular personalities.

We will be clustering the similar kind of people under one kind of personality and lastly will be predicting what is the personality of any person on the basis of their provided data.

Approach:

Putting steps forward on simplifying things by dividing the solution to approach on five different personalities. We made this by dividing our dataset into 5 different categories by finding each cluster for every set and by with "silhouette elbow" we are ought to perform some comparison statistics to pick out the best cluster with greater accuracy to name it as the "Golden Cluster"

EDAV:

We will be carrying out numerical data which could be processed in kind of ratings on the scale from 0-5 to determine the effectiveness of the personality.

Data Representation:

Our thought of representation for this approach is to put it as simply as it to be. So, we opted for 2 of them:

- Numerical: In the form of ratings
- Categorical: In the form of text
- Graphical: Visualizing the results.

Datasets:

Currently we have decided upon only one dataset i.e. 5 Big Personality dataset, Top Personality Dataset and for third we will be scraping data online.

Ref: [Dataset1](#), [Dataset2](#)

Synthetic Minority Over-sampling Technique:

Commonly known as SMOTE process, the classification of our featured variable is somewhat uncommon. Hence, we have implemented 2 methods for SMOTE :

- Upsampling: A process to artificially generate data points consisting of minority class and injecting it to the existing dataset in order to generate better results is called upsampling. After executing this process, the number of records for each of the class in the featured variable are close to 7000
- Downsampling: A process which refers to removing records from majority classes in order to create a more balanced dataset is called Downsampling.

After completion of this process, the number of records for each of the class in the featured variable are close to 7000

Classification:

For classification problem, keeping the 5 personalities in mind, we are classifying the person in the 3 types of categories on the basis of answers to the 50 questions given by him:

- Excellent(Optimum)
- Average(Scope to improve)
- Low-Rated(Need to improve)

For classification, we have used a Muller Loop for classifying the person, which contains the following algorithms:

1. Nearest Neighbour
2. Linear SVM
3. MLPClassifier
4. SVC
5. Decision Tree

Visualized Dashboard:

By creating a dashboard, we are trying to derive an important conclusion for a person, who has to change or improve himself in whatever personality with respect to the other four personalities in terms of personalities . As compared to all the personalities, we have tried to determine the personality where he requires the most improvisation.