

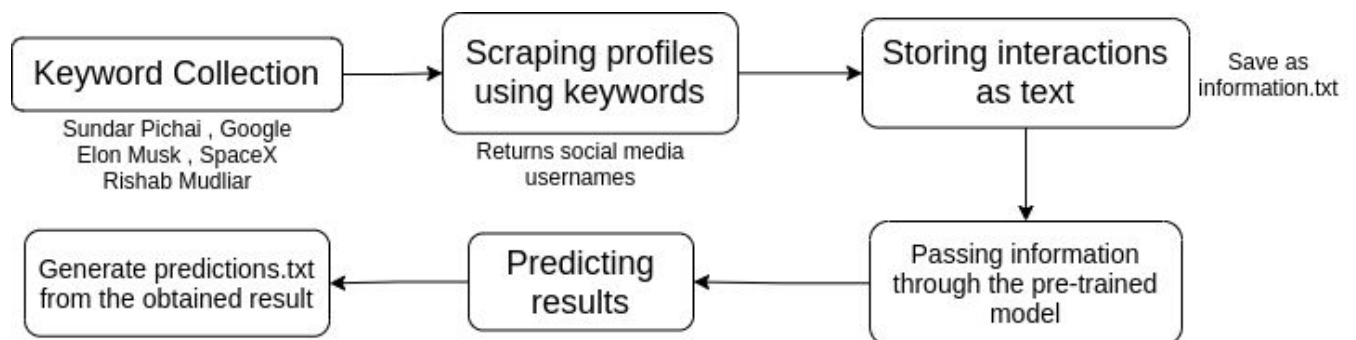
PROJECT REPORT

(Link to the repository - https://github.com/Mainakdeb/e_summit)

ABSTRACT:

The project focuses on predicting the personality of the subjects by analysing their interactions on the social platforms by fitting them to the popular BIG5 personality index model. The BIG-5 metrics responsible for the index score are Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism (**OCEAN metrics**)

ALGORITHM:



1. As a first step, the keywords which refer best to the subjects are collected and stored in a text file named **"names.txt"**.
2. The keywords listed in the file, are then passed iteratively to our profile scraper which returns the social handles of the subject.
3. The collected social handles are taken as a parameter for the platform's official scraping model iteratively which in-turn returns the summarised text of the subject's interactions which are then stored in a text file named **"information.txt"**.
4. On hunting for the data sources, there is much more MBTI data available which is scientifically less reliant, but there is only a very few data on the BIG FIVE traits. Machine Learning algorithms thrive on data, so we came up with an approach to combine MBTI and BIG Five data.
5. The data from three different sources was converted in mutual form and preprocessed according to the needs of the model.
6. The features from text were extracted to vectorize the data with bags of words and the GloVe approach was taken-in hand.
7. Several classification algorithms were employed to train the model on the collected dataset to predict the traits on untrained data.
8. Pre-trained models were generated to remove the need for running the huge model every time thereby saving the runtime of the user.

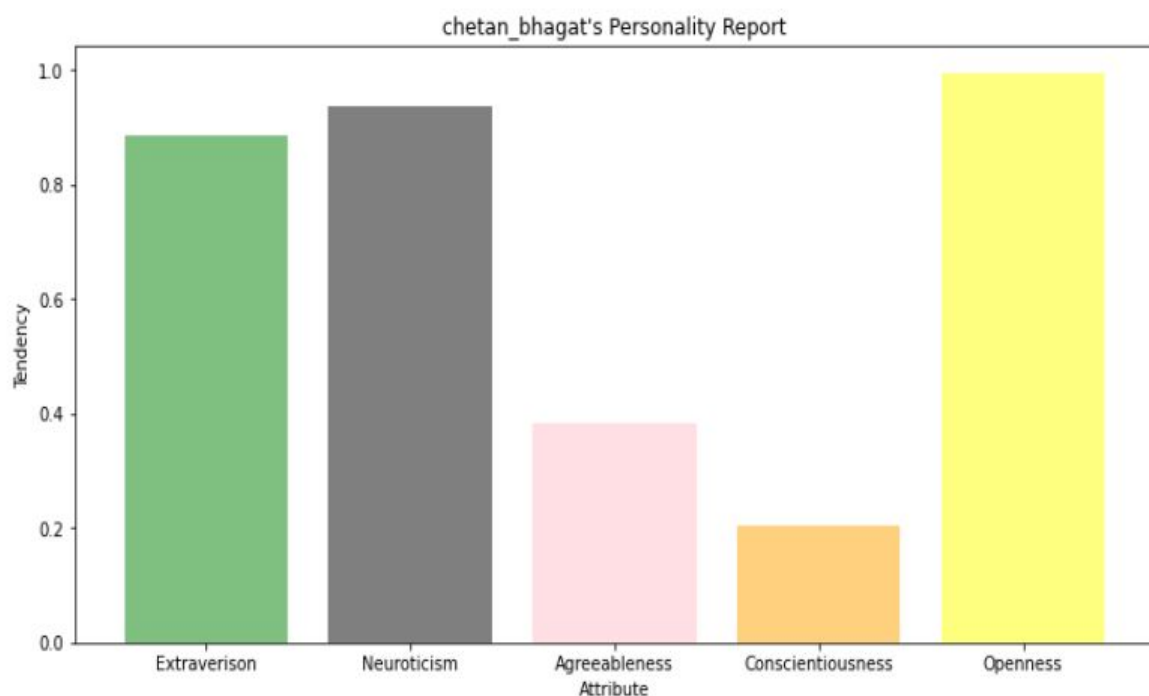
9. The collected summaries stored in “**information.txt**” are now iteratively fed to the pretrained models which in-turn returns the BIG5 index (in percentile).
10. The predictions obtained are then stored in a text file named “**predictions.txt**” along with the corresponding name of the subject.

FILE TEMPLATE:

names.txt	information.txt	predictions.txt
Chetan Bhagat, Writer	Summary of the interactions	Chetan Bhagat => [Scores]

VISUALISED REPORT (AVAILABLE IN OUR [COLAB NOTEBOOK](#))

Subject - Chetan Bhagat



HACKERS ON BOARD :

- Mainak Deb
- Rishab Mudliar
- Sanjay T
- Siddharth Cilamkoti