Personality Analysis using hash-tags from tweets

What is Personality Analysis?

- ***Personality*** is defined as the totality of character attributes and behavioral traits of a person.

- **Personality Analysis** is a methodology for categorizing the character and behavior of a person.

Concept of Analyzing

- The personality of an individual, and plotting the sentiments of tweets according to user/community location.
- Our application would consider the actual mentality of the individuals along with the situation and topic and assess their personality accordingly.

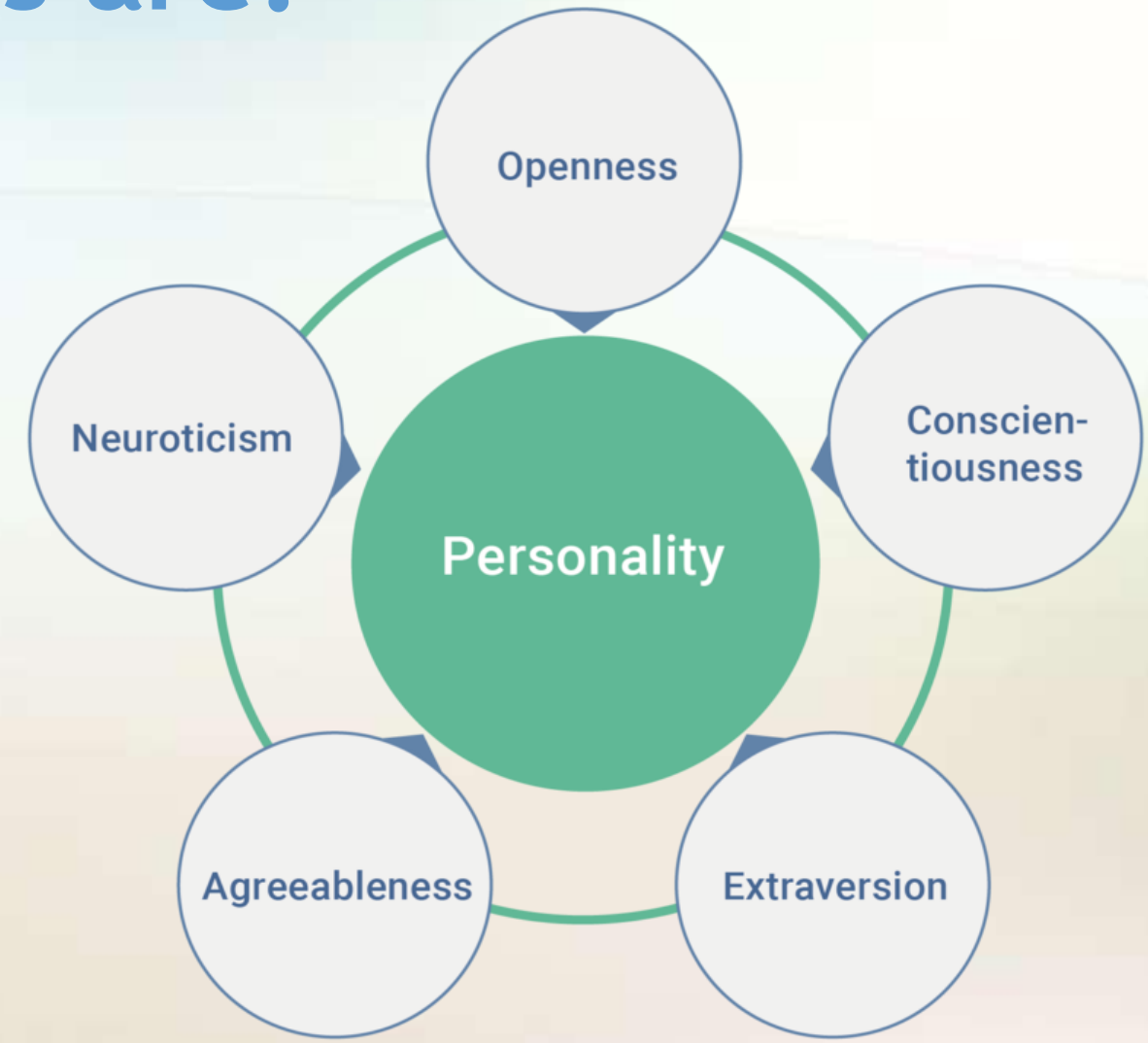
Big Five personality traits

- The Big Five personality traits, also known as the five factor model (FFM), is a taxonomy for personality traits.

- It is based on common language descriptors. When factor analysis (a statistical technique) is applied to personality survey data, some words used to describe aspects of personality are often applied to the same person.

The five factors are:

- Openness
- Conscientiousness
- Extraversion
- Agreeableness
- Neuroticism



Openness to Experience :

- Openness to experience has been described as the depth and complexity of an individual's mental life and experiences

- traits related to openness to experience include:
  - Imaginative
  - Wide variety of interests
  - Original
  - Daring
  - Clever , etc.

Conscientiousness :

- Conscientiousness is a trait that can be described as the tendency to control impulses and act in socially acceptable ways, behaviors that facilitate goal-directed behavior

- Traits within the conscientiousness factor include:
  - Persistent
  - Ambitious
  - Thorough
  - Self-disciplined
  - Consistent , etc.

Extroversion :

- This factor has two familiar ends of the spectrum: extroversion and introversion. It concerns where an individual draws their energy and how they interact with others.

- The traits associated with extroversion are:
  - Sociable
  - Assertive
  - Outgoing
  - Energetic
  - Talkative , etc.

Agreeableness :

- This factor concerns how well people get along with others. It is a construct that rests on how you generally interact with others.

- traits fall under the umbrella of agreeableness:
  - Altruistic
  - Trusting
  - Modest
  - Humble
  - Patient , etc.

Neuroticism :

- Neuroticism is the one Big Five factor in which a high score indicates more negative traits. Neuroticism is an ongoing emotional state defined by negative reactions and feelings.

- traits are commonly associated with neuroticism:
  - Awkward
  - Pessimistic
  - Moody
  - Testy
  - Fearful, etc.

Conceptual Framework

- The hashtags provide us with the topic about which the user has voiced his opinion. Analysis of hashtag is conducted in a way where a set of recent tweets containing the hash-tag term are collected, and the polarity of those tweets containing the term is analyzed whether people have a positive, negative or neutral opinion about it.
- The second part is analysis of remaining text and finding the polarity of the tweet whether it is positive, negative or neutral. Along with the text if tweet contains any emoticons, the polarity of the emoticon is calculated using non-bmp map.

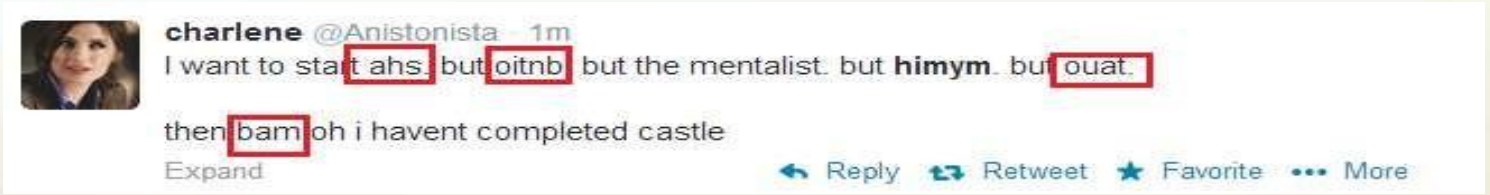
Tools used in Framework

1. Python
2. Django
3. Tweepy
4. Numpy
5. Pandas
6. Sklearn
7. NLTK

Challenges

- Tweets are highly unstructured and also non-grammatical

- Out of Vocabulary Words



- Lexical Variation

- Extensive usage of acronyms like *asap, lol, afaik*

PAUT Framework

Data Gathering:- To gather public opinion based on collected hashtags related to views about political parties including Twitter top trends

English Conversion:- Tweets gathered from user accounts were in the Urdu, English, and Roman Urdu Language. To add novelty in data analysis, we have included tweets of the Urdu language

EX:- ”انصاف ہو کس طرح کی دل صاف نہیں ہے
دل صاف ہو کس طرح کی انصاف نہیں ہے“

“The heart is not clear how to be justice The heart is not clear how justice is”.

PAUT Framework

Data Pre-processing

- Removing non-English Tweets
- Replacing Emoticons by their polarity
- Remove URL, Target Mentions, Hashtags, Numbers.
- Replace Negative Mentions
- Replace Sequence of Repeated Characters eg. ‘cooooooooool’ by ‘coool’
- Remove Nouns and Prepositions

Polarity Calculation and Analysis

Analysis can provide valuable insights from social media platforms by detecting emotions or opinions from a large volume of data present in unstructured format.

PREDICTING PERSONALITY

- To predict the score of a given personality feature, we performed a regression analysis in Weka .
- We used two regression algorithms: Gaussian Process and ZeroR, each with a 10-fold cross-validation with 10 iterations.

Python 3.6.6 Shell

File Edit Shell Debug Options Window Help

```
Python 3.6.6 (v3.6.6:4cf1f54eb7, Jun 27 2018, 03:37:03) [MSC v.1900 64 bit (AMD6
4)] on win32
Type "copyright", "credits" or "license()" for more information.
>>>
==== RESTART: G:\Sachin\Python Project\traitsPredictor-master\modelRun.py ====




processing trait:  ext
SGD Accuracy: 0.51 (+/- 0.09)
Random Forest Accuracy: 0.54 (+/- 0.03)
Multinomial NB Accuracy: 0.56 (+/- 0.01)
Bernoulli NB Accuracy: 0.56 (+/- 0.03)
Gradient Boosting Accuracy: 0.56 (+/- 0.02)
MLP Accuracy: 0.53 (+/- 0.02)


processing trait:  neu
SGD Accuracy: 0.51 (+/- 0.23)
Random Forest Accuracy: 0.60 (+/- 0.01)
Multinomial NB Accuracy: 0.63 (+/- 0.01)
Bernoulli NB Accuracy: 0.63 (+/- 0.01)
Gradient Boosting Accuracy: 0.62 (+/- 0.01)
```

Flow of script

1. Feature Extraction
2. Tweet Scraping
3. Data Process
4. Data Trainning
5. ModelRun

Application:

- The target audience varies from children to senior citizens who wish to know about people’s personality, response.
- The software can be used for people working in HR department of their offices in analyzing the personality of an applicant whether the applicant’s personality is suitable or not.
- People who are interested to know about their role models personality.

Conclusion

- Our subjects completed a personality test and through the Twitter API, we collected publicly accessible information from their profiles. After processing this data, we found many small correlations in the data. Using the profile data as a feature set, we were able to train two machine learning algorithms - ZeroR and Gaussian Processes – to predict scores on each of the five personality .With the ability to guess a user's personality traits, many opportunities are opened for personalizing interfaces and information.

- We considered two structural features - number of friends and network density- but we did not look at personality scores between friends. Understanding the connections between personality, tie strength, trust, and other related factors is an open space for research. By improving our knowledge of these relationships,we can begin to answer more sophisticated questions about how to present trusted, socially-relevant, and well-presented information to users.