Attitude Analysis through Social Media using Data Mining Techniques

Submitted in partial fulfilment of the requirements
of the degree of
Bachelor of Engineering

By

Apoorva Kuckian
Roll No: 2013140023

Kruti Mody
Roll No: 2013140030

Rishal Shah
Roll No: 2013140051

Supervisor:
Dr. Radha Shankarmani

Information Technology Department
Sardar Patel Institute of Technology
2016-17

CERTIFICATE

This is to certify that the project entitled Attitude Analysis through Social Media using Data Mining Techniques is a bonafide work of Apoorva Kuckian (Roll No: 2013140023), Kruti Mody (Roll No: 2013140030) and Rishal Shah (Roll No: 2013140051) submitted to the University of Mumbai in partial fulfilment of the requirement for the award of the degree of Undergraduate in Bachelor of Engineering –Information Technology

(Name and sign)
Supervisor/Guide

(Name and sign)
Project Co-ordinator

(Name and sign)
Head of Department

(Name and sign)
Principal

Project Report Approval for Bachelor of Engineering

Project report entitled Attitude Analysis through Social Media using Data Mining Techniques by Apoorva Kuckian, Kruti Mody and Rishal Shah is approved for the degree of Information Technology.

Examiners

1.--

2.--

Date:

Place:

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Apoorva Kuckian (2013140023)

Kruti Mody (2013140030)

Rishal Shah (2013140051)

Date:

Acknowledgements

We feel immense pleasure in presenting the synopsis report for our project entitled "Attitude Analysis through Social Media using Data Mining Techniques". We have channelized our best efforts towards a systematic approach to the project, keeping in mind the aim we need to achieve.

It is with great pleasure that we present the report on our project work at the end of 8th semester. We take this opportunity to share a few words of gratitude to all those who have supported us in making it possible. We extend our heartfelt gratitude to our project guide Dr. Radha Shankarmani for her able guidance and approachability. We would like to express our gratitude towards her constant encouragement, support and guidance throughout the development of the project. She was the one who never let our morale down and always supported us through our thick and thin. She is a constant source of inspiration for us and took utmost interest in our project. Our regular meetings proved to be a boon in the timely completion of this stage of the project.

We are very thankful to other teaching staff for their moral support and guidance throughout this final year.

Apoorva Kuckian

Kruti Mody

Rishal Shah

Abstract

Social Media is a platform which reflects a user's personality, revealing their personal information and giving insights into their lives. With this project, we aim to achieve a general perspective of one's attitude through social media domains and helping them in heading towards an idealized personality. We will be using their publicly available information to classify and correlate them with the Big Five Personality Traits - Openness to Experience, Conscientiousness, Extroversion, Agreeableness and Neuroticism. This involves predicting the polarity of sentiment words using collaborated opinion mining to successfully determine attitude. The analysis will comprise of the person's current attitude position on a scale along with their best possible model using regression.

Table of Contents

1. Chapter

Introduction

1.1. Problem Statement

This project aims to achieve a general perspective of one’s attitude through social media domains and helping them in heading towards an idealized attitude. The analysis will comprise of a person’s current attitude position on a scale along with their best possible model. The goal is to provide OCEAN value to an individual using his/her social media account with minimal error. On comparing the observed value with the ideal value (standards set by the organization) deciding whether to accept or reject the candidate is much simpler.

1.2. Scope

1.2.1. Data will be collected from Social Media Website datasets (Facebook, Twitter and Instagram).

1.2.2. Using the opinion mining algorithm, sentiment analysis will be conducted on the personal information, posts, etc.

1.2.3. The project will mine the information which is publically available.

1.2.4. The Data Visualization will be done using Tableau where the data will be provided in the form of text, graphs and charts.

1.2.5. The project will give the best possible attitude model of an individual.

1.2.6. The user will be provided with an in-detail analysis of his personality using the Big Five Personality traits.

1.3. Out of Scope

1.3.1. No analysis of images, media and emoticons. Only textual content (english only) will be analysed.No extensive personality tests taken.

1.3.2. No extensive personality tests taken.

1.3.3. Anything other than the Big 5 traits will not be analyzed.

1.4. Existing System

The relationship between the social media networks and personality has already been established. There are several online web portals, software, tools, etc which uses the Big Five Inventory to predict the personality of the subject under observation with minimal error rate. There have been previous studies that relates personality to social media by using classification and regression methodologies for processing the textual data. Also, sentiment analysis and opinion mining have been used so as to predict personality using twitter or any other social media. Also, Simple Percentage Analysis of Attitude Questionnaires are used to analyse attitude.

1.5. Proposed System

Attitude analysis will prove to be very helpful for recruiters for hiring deserving candidates by comparing the ideal and observed attitude. Social Media data pertaining to a specific social media website is obtained and is categorized into Big Five Personality

Traits - Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism (OCEAN). The attitude thus observed is termed as 'Observed Attitude'. The organization/ recruiter will set the trait value required in each case called 'Ideal Attitude'. Comparison between ideal and observed attitude will facilitate in decision making process.

1.6. Constraints

1.6.1. Only free data which is provided by various social media APIs will be used posing a limitation on data availability as paid data will not be considered during analysis.

1.6.2. System will work only in presence of APIs.

1.6.3. Larger search time due to huge dataset

1.7. System requirements

1.7.1. Functional Requirements

The following operations should be performed:

1.7.1.1. Map twitter handle to correct twitter ID:

When user enters his credentials in which the twitter handle should map to the appropriate twitter ID for the purpose of successful retrieving and storage of tweets

1.7.1.2. Providing summary of past reports:

If the user wishes to compare past reports, the summary (OCEAN values) should be easily provided on the front page.

1.7.1.3. Preprocessing tweets:

Remove stop words and images, audio, video and apply aggregation formula.

1.7.1.4. Provide a radar chart along with summary:

User can compare observed with ideal attitude using a graphical representation and short summary for comparison result on each trait.

1.7.2. Non-Functional Requirements

The non-functional requirements represent requirements that should work to assist the application to accomplish its goal.

1.7.2.1. Interface: The system must provide a web- based interface, which allows humans to interact with the system.

1.7.2.2. Performance: The system must be able to handle a user request in an acceptable amount of time.

1.7.2.3. Interoperability: Other applications can request the services of the system automatically (without direct human interaction with the GUI). The system must be able to interact with other applications using standard technologies.

1.7.2.4. Safety Requirements: For the safety requirements nothing but an operation of weekly backups for the database should take place.

1.7.2.5. Security and Privacy Requirements: Only authorized persons who are allowed to use and access the database, web pages and the product engine.

1.7.2.6. Software Quality Attributes:

- Reliability:The solution should provide reliability to the user that the product will run with all the features mentioned in this document are available and executing perfectly. It should be tested and debugged completely. All exceptions should be well handled.
- Accuracy: The solution should be able to reach the desired level of accuracy.

2. Chapter

Literature Survey

2.1. Survey

2.1.1. Our Twitter Profiles,Our Selves: Predicting Personality with Twitter - Daniele Quercia, Michal Kosinski, David Stillwell, Jon Crowcroft

Conference: Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on 9-11 Oct. 2011

Using this paper, we have used the concept of the Big Five Personality Traits that are - Openness to Experience, Conscientiousness, Extroversion, Agreeableness and Neuroticism. This helps in predicting the correct attitude of a person on the basis of the social media platform. We have also included the concept of analysis of Twitter data sets using Regression from this paper.

2.1.2. Predicting Personality with Social Media - Jennifer Golbeck,Cristina Robles,Karen Turner

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

This paper deals with the prediction of attitude and personality using Facebook data sets. This paper deals with the Data Collection using the available Facebook datasets followed by the Personality and Profile Correlations using the Pearson correlation between feature scores and personality scores. This paper also follows the concept of the Big Five Personality traits for the attitude analysis thus providing an insight into the various categories of attitudes of individuals along with their major and minor characteristic traits.

2.1.3. Sentiment Analysis Using Collaborated Opinion Mining - Deepali Virmani, Vikrant Malhotra, Ridhi Tyagi

This paper deals with the two important concepts of Real and Perceived Attitudes. In this paper, individuals answered a certain set of questions according to their point of view, as well as their beliefs about their friends' attitudes. This resulted in a Real vs Perceived Attitude agreement thus helping in understanding the attitude of a person with respect to their friends' attitude.

2.1.4. Real-World Behavior Analysis through a Social Media Lens - Mohammad-Ali Abbasi, Sun-Ki Chai, Huan Liu, Kiran Sagoo

This paper provides us predicting the real-world collective behaviour using the Social media platforms namely Twitter. This paper helps in finding the next probable event on the basis of the social activity of an individual about the event.

This helps in understanding the attitude of an individual according to their social behaviour. It also helps in tracking change of attitudes during a social movement by using a time series of tweets and blog posts.

2.1.5. Real and Perceived Attitude Agreement in Social Networks - Sharad Goel, Winter Mason, and Duncan J. Watts

This paper deals with the two important concepts of Real and Perceived Attitudes. In this paper, individuals answered a certain set of questions according to their point of view, as well as their beliefs about their friends' attitudes. This resulted in a Real vs Perceived Attitude agreement thus helping in understanding the attitude of a person with respect to their friends' attitude.

3. Chapter

Implementation

3.1. Big 5 Traits (OCEAN)

3.1.1. Openness:

It signifies imagination, creativity, curiosity,tolerance, political liberalism, and appreciation for culture. People scoring high on Openness like change, appreciate new and unusual ideas, and have a good sense of aesthetics.

3.1.2. Conscientiousness:

It signifies preference for an organized approach in contrast to a spontaneous one. People having high Conscientiousness are more likely to be well organized, reliable, and consistent. They enjoy planning, seeking achievements, and pursuing long-term goals.

3.1.3. Extraversion:

It signifies a tendency to seek stimulation in the external world, other's company, and express positive emotions. People having high Extroversion tend to be more outgoing, friendly, and socially active. They are usually energetic and talkative; they do not mind being at the center of attention, and make new friends more easily.

3.1.4. Agreeableness:

It signifies to a focus on maintaining positive social relations, being friendly, compassionate, and cooperative. People having high Agreeableness people tend to trust others and adapt to their needs.

3.1.5. Neuroticism:

It signifies the tendency to experience mood swings and emotions such as guilt, anger, anxiety, and depression. People having high Neuroticism are more likely to experience stress and nervousness, while people scoring low on Neuroticism tend to be calmer and self-confident.

3.2. Preprocessing

Data preprocessing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. Data preprocessing is a proven method of resolving such issues.

For eg:-
Preprocess the tweet before creating the FeatureVector for the tweet.

```
def processTweet(self, tweet):
        # process the tweets

        #Convert to lower case
        tweet = tweet.lower()
        #Convert www.* or https?://* to URL
        tweet = re.sub('((www\.[^\s]+)|(https?://[^\s]+))','URL',tweet)
        #Convert @username to AT_USER
        tweet = re.sub('@[^\s]+','AT_USER',tweet)
        #Remove additional white spaces
        tweet = re.sub('[\s]+', ' ', tweet)
        #Replace #word with word
        tweet = re.sub(r'#([^\s]+)', r'\1', tweet)
        #trim
        tweet = tweet.strip('\'"')
        return tweet
```

3.3. Tweepy

Tweepy is open-sourced, hosted on GitHub and enables Python to communicate with Twitter platform and use its API.

For eg:-
All the tweets of a user are retrieved using the twitter id of the user.

```
def get_all_tweets(screen_name):
          auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
```

```
    auth.set_access_token(access_key, access_secret)
    api = tweepy.API(auth)

    alltweets = []
    new_tweets = []
    outtweets = []

    new_tweets = api.user_timeline(screen_name=screen_name, count=200)

    alltweets.extend(new_tweets)

    # save the id of the oldest tweet less one
    oldest = alltweets[-1].id - 1

    # keep grabbing tweets until there are no tweets left to grab
    while len(new_tweets) > 0:
        print ("getting tweets before %s" % (oldest))

        # all subsiquent requests use the max_id param to prevent duplicates
        new_tweets = api.user_timeline(screen_name=screen_name,
count=200, max_id=oldest)

        # save most recent tweets
        alltweets.extend(new_tweets)

        # update the id of the oldest tweet less one
        oldest = alltweets[-1].id - 1

        print ("...%s tweets downloaded so far" % (len(alltweets)))

    # transform the tweepy tweets into a 2D array
    outtweets = [[tweet.id_str, tweet.created_at, tweet.coordinates, tweet.geo,
tweet.source, tweet.text] for tweet in
                alltweets]
    return outtweets
```

3.4. Flask

Flask is a web application framework written in Python. Flask is often referred to as a micro framework. It aims to keep the core of an application simple yet extensible. Flask does not have built-in abstraction layer for database handling, nor does it have form validation support. Instead, Flask supports the extensions to add such functionality to the application.

For eg:-
Flask connects the html front end with python back end.

```
app = Flask(__name__)

@app.route('/credentials', methods=['POST'])
def my_form_post():
        #credentials page code


if __name__ == '__main__':
  app.run()
```

3.5. D3.js

D3.js (or just D3 for Data-Driven Documents) is a JavaScript library for producing dynamic, interactive data visualizations in web browsers. It makes use of the widely implemented SVG, HTML5, and CSS standards. It is the successor to the earlier Protovis framework.

For eg:-
D3.js is used as a visualization tool to make the kiviat chart in report generation.

```
<script type="text/javascript">
                var w = 500,
                        h = 500;

                var colorscale = d3.scale.category10();

                //Legend titles
                var LegendOptions = ['Observed','Ideal'];

                //Data
                var d = [
                                  [
                                        {axis:"Openness to Experience",value:{{
ocean[0] }}},
                                        {axis:"Conscientiousness",value:{{ ocean[1]
}}},
                                        {axis:"Extraversion",value:{{ ocean[2] }}},
                                        {axis:"Agreeableness",value:{{ ocean[3]
}}},
                                        {axis:"Neuroticism",value:{{ ocean[4] }}},

                                  ],[
                                        {axis:"Openness to Experience",value:0.66},
                                        {axis:"Conscientiousness",value:0.71},
                                        {axis:"Extraversion",value:0.55},
                                        {axis:"Agreeableness",value:0.40},
                                        {axis:"Neuroticism",value:0.10},

                                ]
                               ];

                        //Options for the Radar chart, other than default
                        var mycfg = {
                          w: w,
                          h: h,
                          maxValue: 1.0,
                          levels: 5,
                          ExtraWidthX: 500
                        }

                        //Call function to draw the Radar chart
                        //Will expect that data is in %'s
                        RadarChart.draw("#chart", d, mycfg);

                        var svg = d3.select('#body')
                                .selectAll('svg')
                                .append('svg')
                                .attr("width", w+500)
                                .attr("height", h)
                </script>
```

4. Chapter

Methodology

4.1. Analysis

4.1.1. System flow diagram

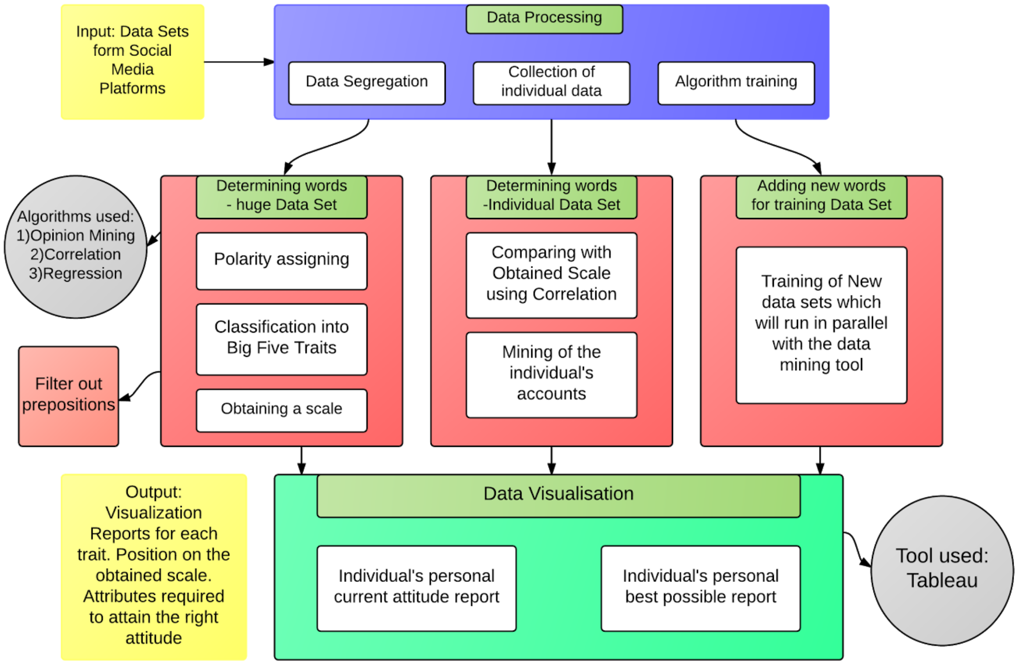

Input: Data Sets form Social Media Platforms
Data Processing
Data Segregation
Collection of individual data
Algorithm training
Algorithms used: 1)Opinion Mining 2)Correlation 3)Regression
Determining words - huge Data Set
Polarity assigning
Classification into Big Five Traits
Obtaining a scale
Filter out prepositions
Determining words -Individual Data Set
Comparing with Obtained Scale using Correlation
Mining of the individual's accounts
Adding new words for training Data Set
Training of New data sets which will run in parallel with the data mining tool
Output: Visualization Reports for each trait. Position on the obtained scale. Attributes required to attain the right attitude
Data Visualisation
Individual's personal current attitude report
Individual's personal best possible report
Tool used: Tableau

Fig. 4.1. System flow diagram

4.1.2. Architecture diagram

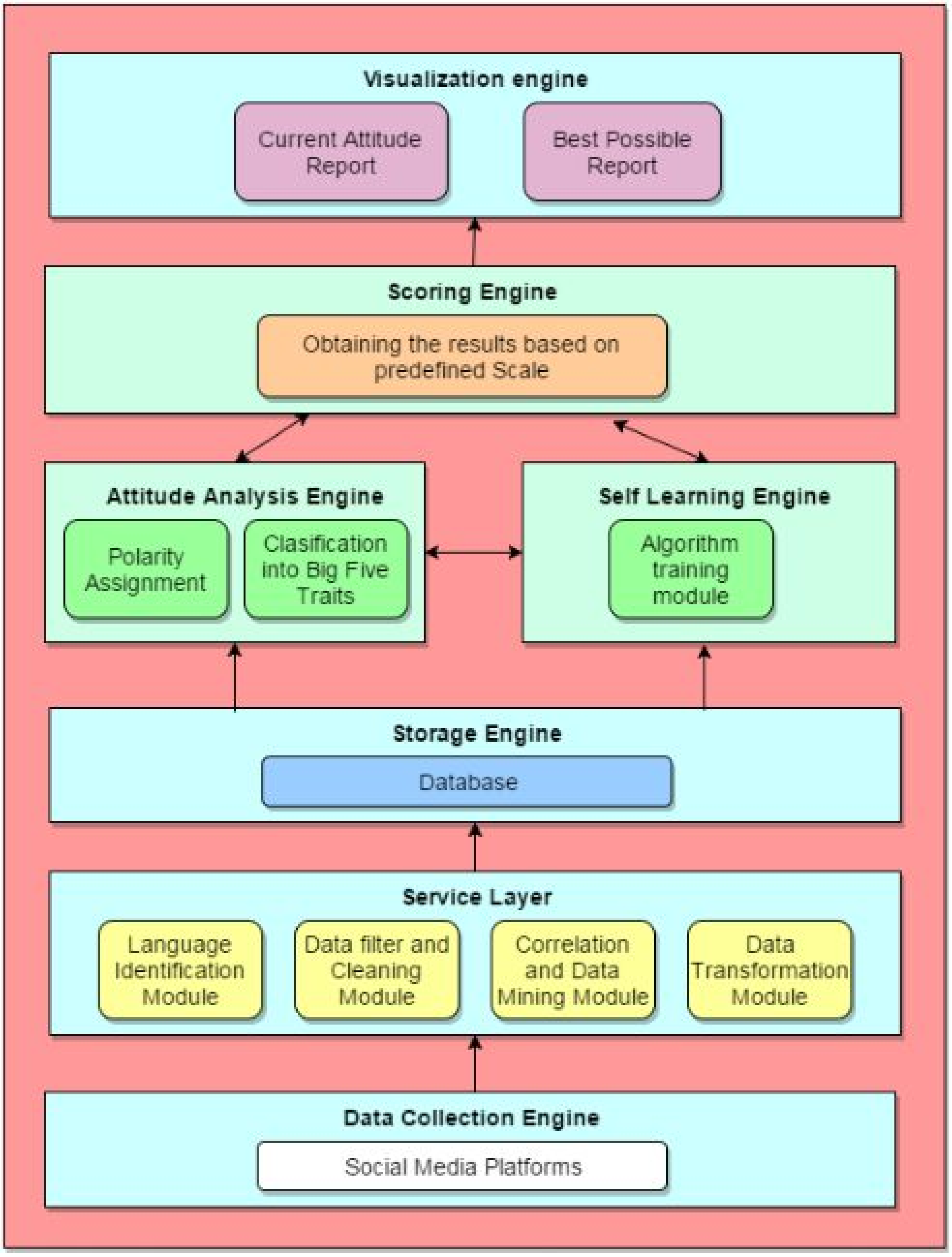

Visualization engine
Current Attitude Report
Best Possible Report
Scoring Engine
Obtaining the results based on predefined Scale
Attitude Analysis Engine
Polarity Assignment
Clasification into Big Five Traits
Self Learning Engine
Algorithm training module
Storage Engine
Database
Service Layer
Language Identification Module
Data filter and Cleaning Module
Correlation and Data Mining Module
Data Transformation Module
Data Collection Engine
Social Media Platforms

Fig. 4.2. Architecture Diagram

4.1.3. Use Case Diagram

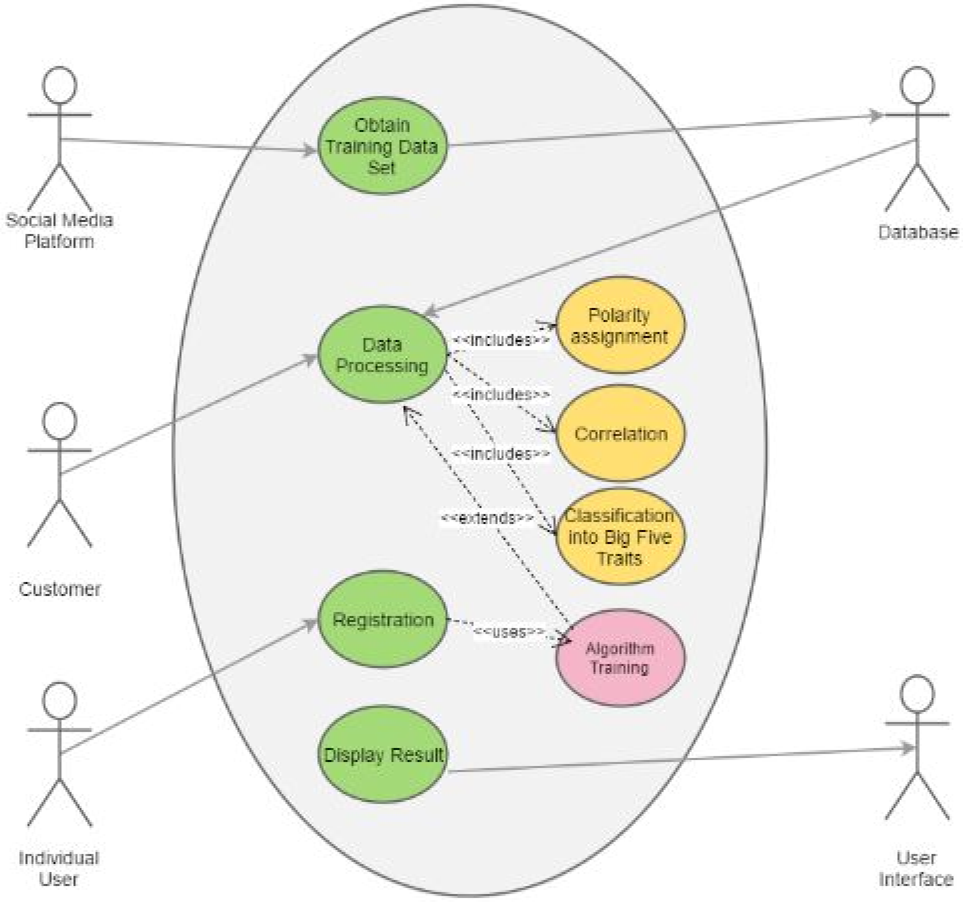


Fig. 4.3. Use Case Diagram

4.1.4. Sequence Diagram

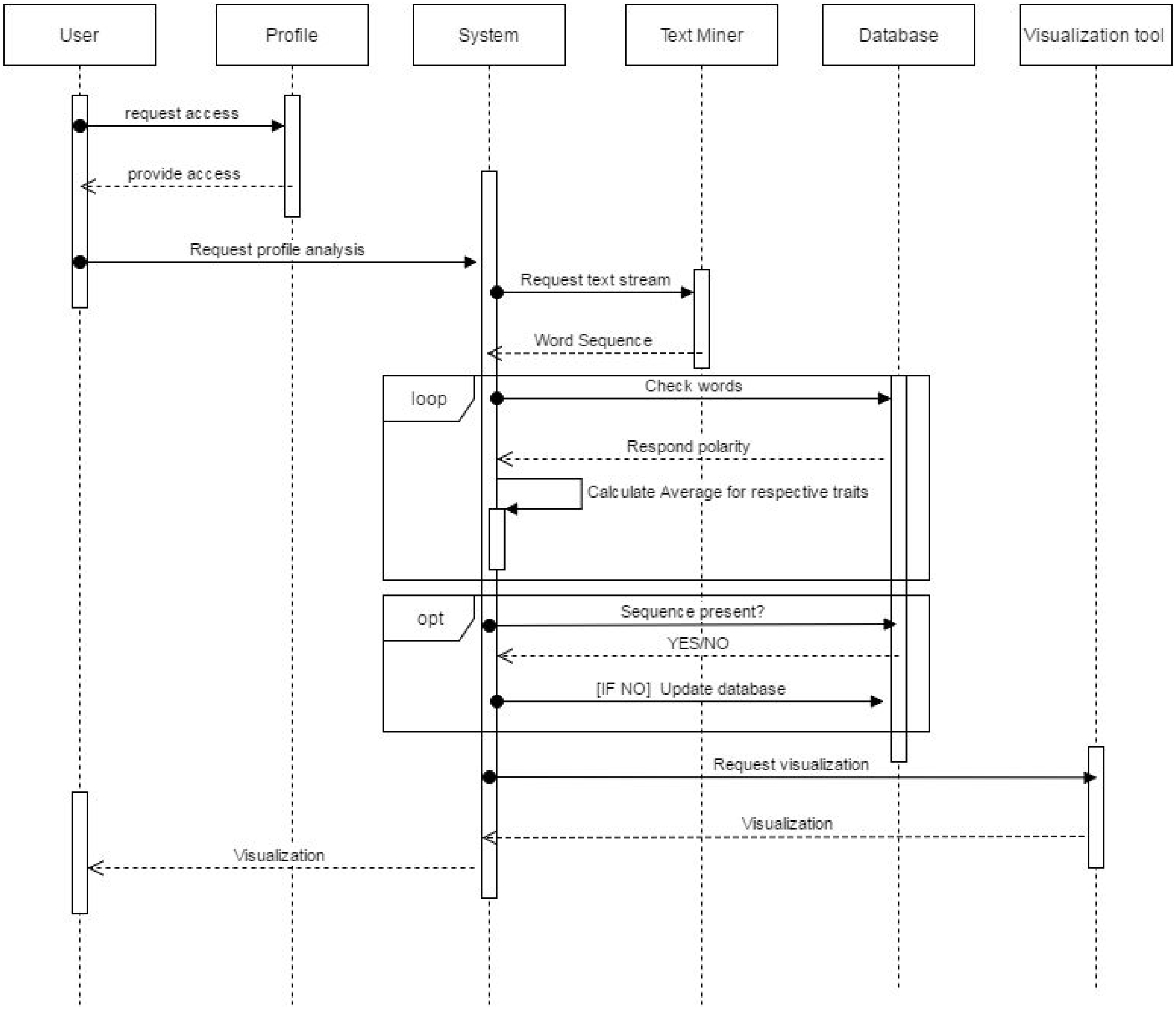


Fig. 4.4. Sequence Diagram

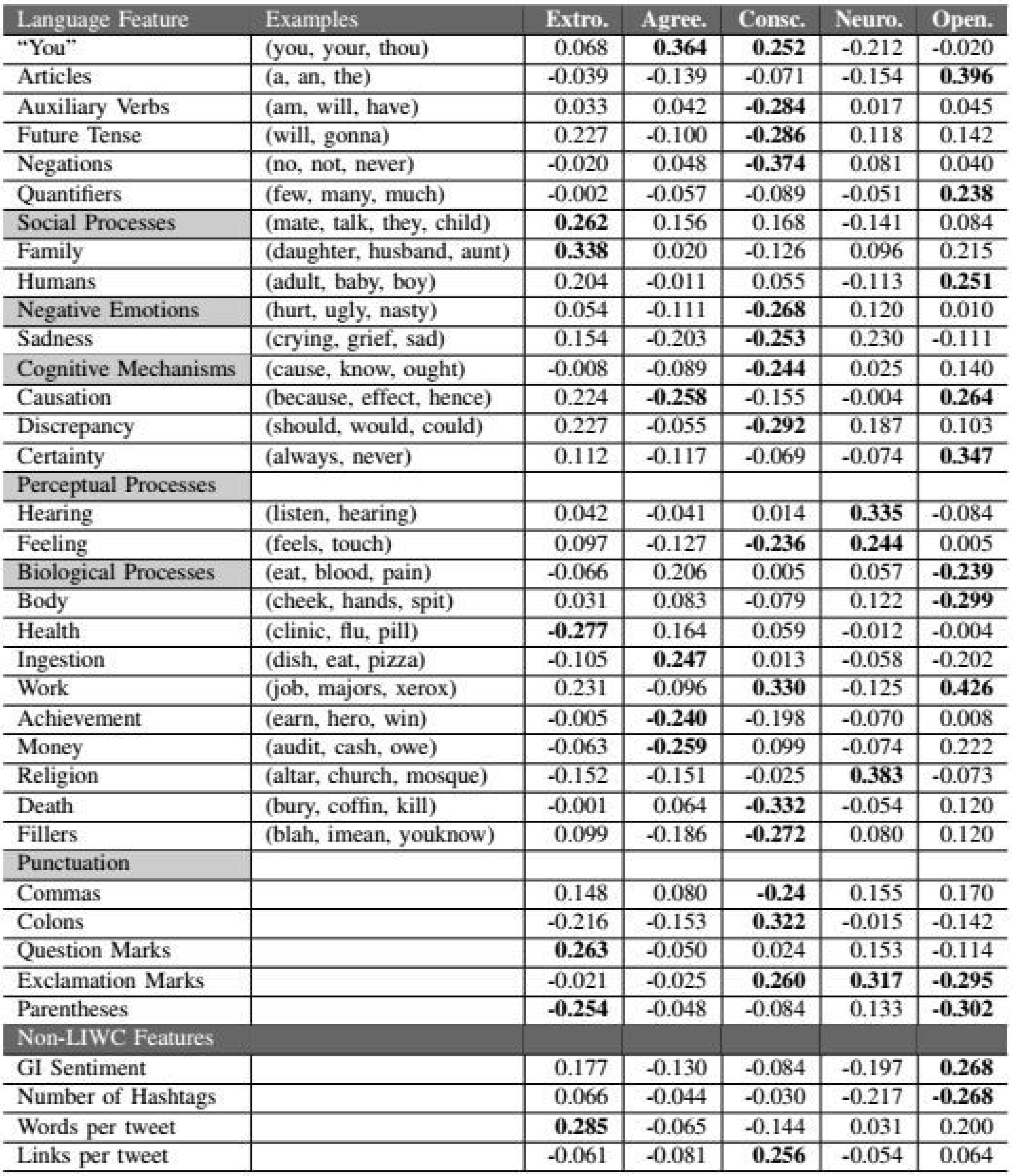
4.2. Data Collection

4.2.1. Extracting Tweets

Tweepy is open-sourced API, hosted on GitHub and enables Python to communicate with Twitter platform and use its API. Along with tweepy, Twython is also used to extract twitter data which is a python library. In order to use tweepy, it is necessary to log in and obtain four unique credentials namely, consumer key, consumer secret, access key and access secret.These credentials are then used in python code base wherever interaction with live tweets is required. However, there is limitation of 3200 tweets which can be retrieved per user, at once. The tweets of the user under study thus obtained are stored in an excel sheet which can be further processed.

4.2.2. Formation of Data Set

As publicly available data set proved to be inefficient and incomplete there was a need to create a new data set that could relate words obtained from twitter profile and OCEAN traits. Fig. 4.5. shows the categories in which all the words in a dictionary fall and it's respective OCEAN value [1]. Each word is categorized among these categories and assigned these OCEAN values. Initially the number of words taken in data set are limited but with continuously running the algorithm the data set is trained and new words are added to the data set.



Language Feature	Examples	Extro.	Agree.	Consc.	Neuro.	Open.
"You"	(you, your, thou)	0.068	**0.364**	**0.252**	-0.212	-0.020
Articles	(a, an, the)	-0.039	-0.139	-0.071	-0.154	**0.396**
Auxiliary Verbs	(am, will, have)	0.033	0.042	**-0.284**	0.017	0.045
Future Tense	(will, gonna)	0.227	-0.100	**-0.286**	0.118	0.142
Negations	(no, not, never)	-0.020	0.048	**-0.374**	0.081	0.040
Quantifiers	(few, many, much)	-0.002	-0.057	-0.089	-0.051	**0.238**
Social Processes	(mate, talk, they, child)	**0.262**	0.156	0.168	-0.141	0.084
Family	(daughter, husband, aunt)	**0.338**	0.020	-0.126	0.096	0.215
Humans	(adult, baby, boy)	0.204	-0.011	0.055	-0.113	**0.251**
Negative Emotions	(hurt, ugly, nasty)	0.054	-0.111	**-0.268**	0.120	0.010
Sadness	(crying, grief, sad)	0.154	-0.203	**-0.253**	0.230	-0.111
Cognitive Mechanisms	(cause, know, ought)	-0.008	-0.089	**-0.244**	0.025	0.140
Causation	(because, effect, hence)	0.224	**-0.258**	-0.155	-0.004	**0.264**
Discrepancy	(should, would, could)	0.227	-0.055	**-0.292**	0.187	0.103
Certainty	(always, never)	0.112	-0.117	-0.069	-0.074	**0.347**
Perceptual Processes						
Hearing	(listen, hearing)	0.042	-0.041	0.014	**0.335**	-0.084
Feeling	(feels, touch)	0.097	-0.127	**-0.236**	**0.244**	0.005
Biological Processes	(eat, blood, pain)	-0.066	0.206	0.005	0.057	**-0.239**
Body	(cheek, hands, spit)	0.031	0.083	-0.079	0.122	**-0.299**
Health	(clinic, flu, pill)	**-0.277**	0.164	0.059	-0.012	-0.004
Ingestion	(dish, eat, pizza)	-0.105	**0.247**	0.013	-0.058	-0.202
Work	(job, majors, xerox)	0.231	-0.096	**0.330**	-0.125	**0.426**
Achievement	(earn, hero, win)	-0.005	**-0.240**	-0.198	-0.070	0.008
Money	(audit, cash, owe)	-0.063	**-0.259**	0.099	-0.074	0.222
Religion	(altar, church, mosque)	-0.152	-0.151	-0.025	**0.383**	-0.073
Death	(bury, coffin, kill)	-0.001	0.064	**-0.332**	-0.054	0.120
Fillers	(blah, imean, youknow)	0.099	-0.186	**-0.272**	0.080	0.120
Punctuation						
Commas		0.148	0.080	**-0.24**	0.155	0.170
Colons		-0.216	-0.153	**0.322**	-0.015	-0.142
Question Marks		**0.263**	-0.050	0.024	0.153	-0.114
Exclamation Marks		-0.021	-0.025	**0.260**	**0.317**	**-0.295**
Parentheses		**-0.254**	-0.048	-0.084	0.133	**-0.302**
Non-LIWC Features						
GI Sentiment		0.177	-0.130	-0.084	-0.197	**0.268**
Number of Hashtags		0.066	-0.044	-0.030	-0.217	**-0.268**
Words per tweet		**0.285**	-0.065	-0.144	0.031	0.200
Links per tweet		-0.061	-0.081	**0.256**	-0.054	0.064

Fig. 4.5. Category mapping to OCEAN values

4.3. Preprocessing Tweets

Once the user enters his/her credentials on the web page his/her tweets are stored in an Excel sheet Twitter_Timeline.xlsx. These tweets won't be long but will contain pictures, URL, twitter user handles, emoticons, etc. These tweets are filtered such that stop words like a, the, in, of, etc are removed. Stop words are those words which cannot have any sentiment and are basically just connectives. Also, the URL is replaced with a word *"URL"* and twitter user references are replaced with *"AT_USER"*. Excessive punctuation marks at the start and at the end of the tweets are removed.

Now, after the tweets are filtered for stop words, URL and punctuation marks the resulting words obtained are stored to a vector array called feature vector. The words in feature vector account for some value in each OCEAN trait. These words are termed as *Feature Words (FW).*

4.4. OCEAN value for tweets

After formation of Feature Vector, each word is analysed. Each tweet has a separate feature vector. Let us consider a case where we have 5 tweets. So, we will have 5 feature vector each containing FW. Now, each tweet is processed as per the flow chart shown in Figure 3. FW is checked whether it is present in the data set or not. If it is present then the corresponding OCEAN values are inserted into the *sentiment_list*. If not, then FW is searched to have a synonym within the dictionary API. If synonym is found then the algorithm checks whether the synonym is present in the data set, if present in data set then the corresponding OCEAN value is added to sentiment_list. After adding the value of synonym the FW along with other remaining synonyms are inserted into the data set so that next time that word is encountered as FW it can be directly found from the data set and the need to invoke the dictionary API is eliminated. Now, if the synonym is also not found then the algorithm displays the error message of 'synonym not found' and the FW is added to *buffer.txt* file. The words in *buffer.txt* are checked whether they are capable of having sentiment or not and are then manually inserted to the data set.

For next tweet the process continues and the OCEAN values are appended in the sentiment_list such that length of the sentiment list will be five times the number of tweets evaluated. Here, since we considered 5 tweets the length of sentiment list will be 25 where 1st, 6th, 11th, 16th, 21st will correspond to the openness of tweet 1, 2, 3, 4 and 5; 2nd, 7th, 12th, 17th, 22nd will correspond to the conscientiousness of tweet 1, 2, 3, 4 and 5 and so on. In order to find the cumulative openness value for each tweet the weighted average is calculated. The openness value of each FW is multiplied by the weights assigned to the word.

Weights are assigned to the words based on degree of comparision and level of significance. The value thus obtained is added and is divided by the sum of weights of each FW (equation (1)). Hence the 1st, 6th, 11th, 16th and 21st values are the weighted average openness value of each tweet.

$$Openness_{tweet} = \frac{\sum_{i=1}^{no.ofFW} openness_i * weight_i}{\sum_{i=1}^{no.ofFW} weight_i}$$

----------Eq.4.1.

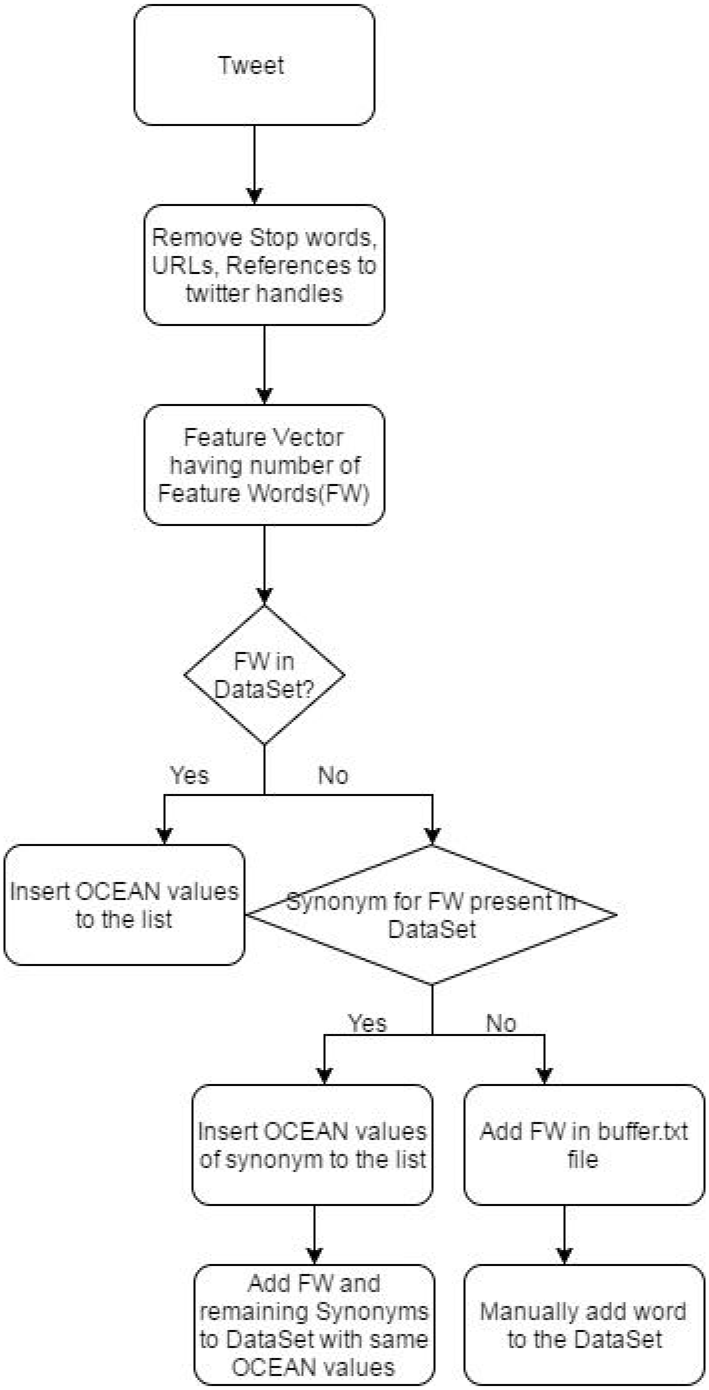

Tweet
Remove Stop words, URLs, References to twitter handles
Feature Vector having number of Feature Words(FW)
FW in DataSet?
Yes
No
Insert OCEAN values to the list
Synonym for FW present in DataSet
Yes
No
Insert OCEAN values of synonym to the list
Add FW in buffer.txt file
Add FW and remaining Synonyms to DataSet with same OCEAN values
Manually add word to the DataSet

Fig. 4.6. Conditional Flow

4.5. Output Visualization.

The kiviat or radar chart is obtained on the use interface just after he/she enters his credentials on the web page. The radar chart formed is made by using d3.js (Data Driven Documents, a JavaScript library).

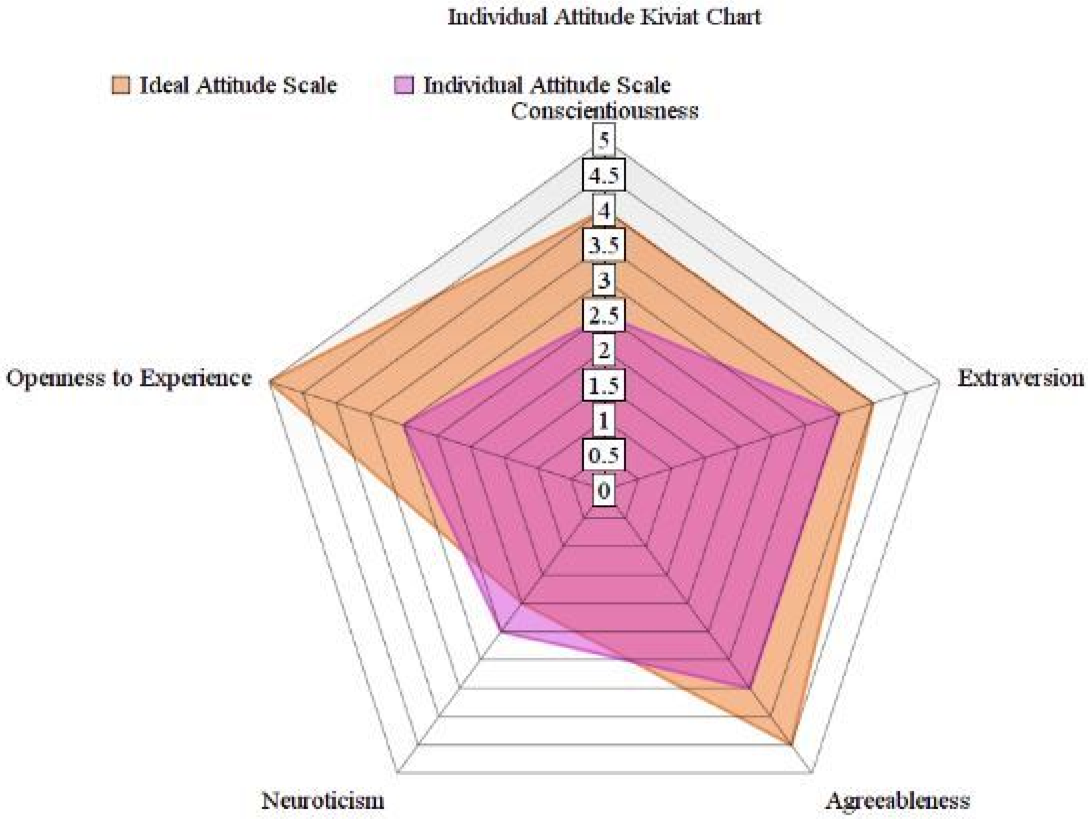


Fig. 4.7. Output Kiviat Chart

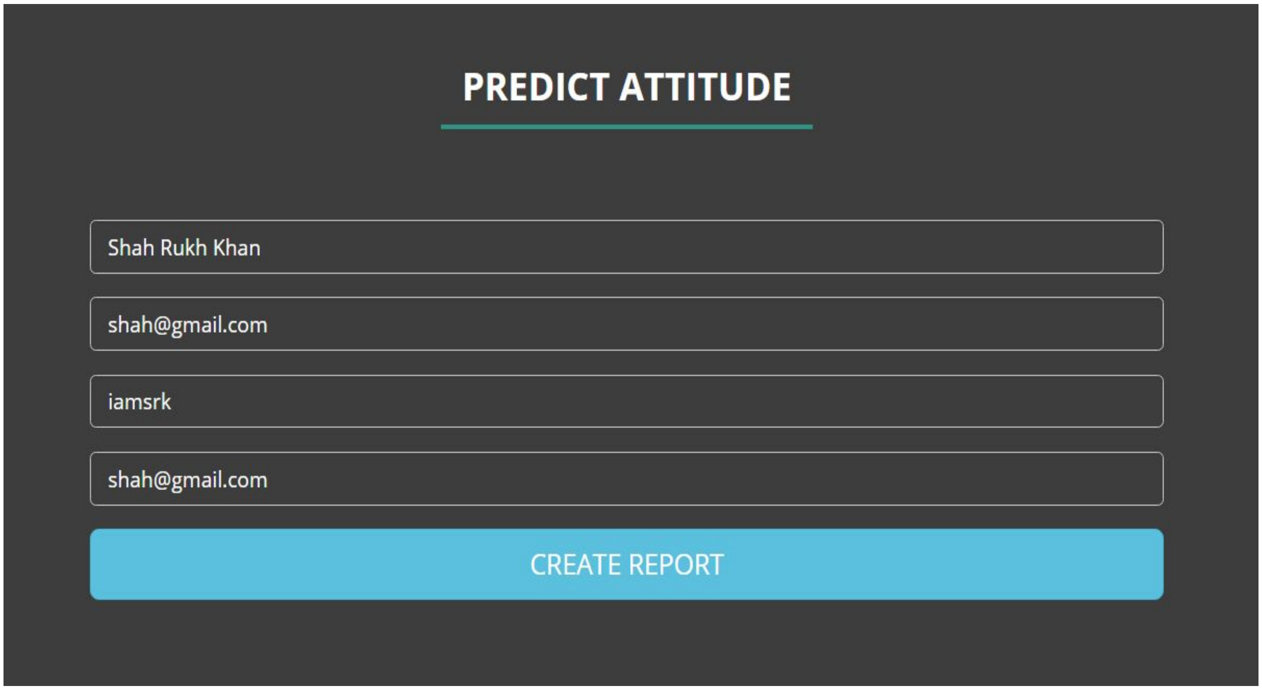
The OCEAN values calculated are represented on the radar chart as shown in Fig. 4.7. The individual's observed attitude value is shown along with the ideally required attitude value in each of the Big Five Trait.

5. Chapter

Results And Discussions

5.1. Result

The user enters his credentials via the credential page as shown in Fig. 5.1.



Shah Rukh Khan

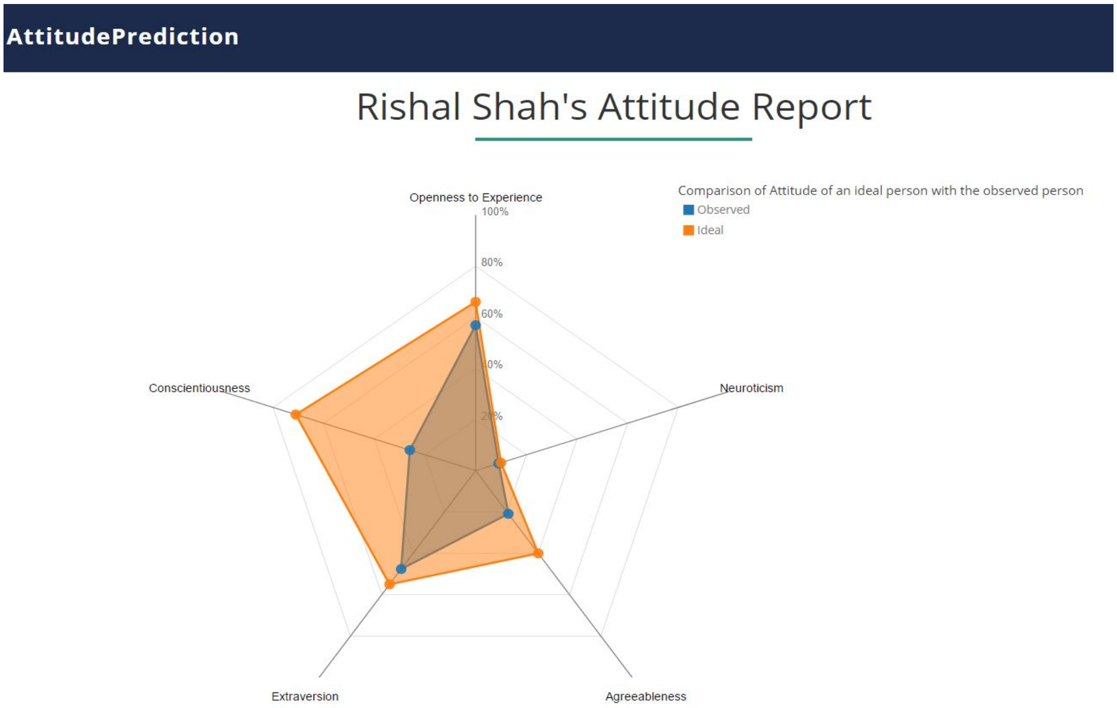
shah@gmail.com

iamsrk

shah@gmail.com

Fig. 5.1. Credential page

The following figure Fig. 5.2. Shows the kiviat chart of the user. The chart has values for ideal attitude as well as observed attitude which makes it easier to compare.



AttitudePrediction

Rishal Shah's Attitude Report

Fig. 5.2. Kiviat chart

5.2. Discussion

Users of social media reveal a lot about themselves, the question that arises from this research is how the results can be used. The kiviat chart can be studied to find out how close the \textit{observed value} of the subject is to the \textit{ideal value} defined by the organization which utilizes the tool for analysing the attitude. The difference between the observed and ideal value can be calculated and lesser the obtained value, greater is the closeness to the desired attitude. Considering the case previously listed, lesser the difference greater is the probability of acquiring the position offered by the organization. The person is more likely to be recruited if he/she fits perfectly to the expectations in terms of attitude. Organizations recruit employees based on skill set and by personal interview which lasts for maximum 1 hr. It is not possible to correctly determine the interviewee's attitude within few minutes or hours, that is when this research comes into picture. People easily acquire the skill set required for a position in an organization, what lacks is the attitude. What is the use of the skill set if you do not know where to apply it and how to apply it? Thus, with this research the person can work effectively so as to reach the ideal attitude standards along with the skill set thus helping him/her to excel. By knowing his/her attitude he/she will also know in which position does his attitude suits the best and discover his true potential.

6. Chapter

Conclusion And Future Work

6.1. Conclusion

In this paper we have shown how big five traits - openness, conscientiousness, extraversion, agreeableness and neuroticism and twitter data of an individual can be used to analyse attitude of that individual. Organizations set a standardized measure required for a position and based on the result obtained after analysing the candidate's twitter account the recruiter will know how close the candidate's attitude is to ideal attitude. The recruiter will thus be as ease while making a decision to accept or reject the candidate. The way it connects to the bigger picture is that it saves the overall time and effort spent in the existing recruitment system.

6.2. Future Work

The researched can be further extended by categorizing based on gender. Also, the algorithm implemented in this project can be used to analyse other social media domains. This project can be extended to analyze images, audios and videos. Large expanded organizations are always resourceful, so this research can be used for people who are less resourceful such as recruiting for startups where the recruitment of employee with correct attitude decides the progress of the startup. It can be used for finding a marriage partner, for hiring domestic help. Thus, this research has varied uses ranging from common man to mega organizations.

7. Chapter

References

[1] Jennifer Golbeck, Cristina Robles, Michon Edmondson, Karen Turner, “Predicting Personality from Twitter” IEEE International Conference on Social Computing, 2011

[2] Daniele Querica, Michal Kosinski, David Stillwell, Jon Crowcroft, Our Twitter Profiles, Our Selves: Predicting Personality with Twitter 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), 2011

[3] Ravikiran Janardhana, “Twitter Sentiment Analysis and Opinion Mining”, Department of Computer Science,university of North Carolina at Chapel Hill

[4] Margaret M. Bradely, Peter J. Lang, “Affective Norms for English Words: Instruction Manual and Affective Ratings”, Intelligent Sensors, Technical report C-1,The center for research in Psychophysiology, university of Florida, 1999

[5] Deepali Virmani, Vikrant Malhotra, Ridhi Tyagi, “Sentiment Analysis Using Collaborated opinion Mining”, NASA ADS , January 2014.

[6] Finn Arup Nielson, “A new ANEW: Evaluation of a word list sentiment analysis in microblogs”, Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages 718 in CEUR Workshop Proceedings : 93-98. 2011 May

[7] Francois Mairesse, Marilyn A Walker, Matthias R. Mehl, and roger K Moore, “Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text,” in Journal of Artificial Intelligence Research 30 (2007) 457-500,November 2007.

[8] Twitter Stats 2015, Twitter Inc.

[9] International Personality Item Pool, http://ipip.ori.org/

[10] Jennifer Golbeck, “Predicting Personality from Social Media Text”, Transactions on Replication Research, 2016.

[11] P. Rosen and D. Kluemper, “The Impact of the Big Five Personality Traits on the Acceptance of Social Networking Website”, AMCIS 2008 Proceedings, page 274, 2008.

[12] M. Selfhout, W. Burk, S. Branje, J. Denissen, M. van Aken, and W. Meeus, “Emerging Late Adolescent Friendship Networks and Big Five Personality Traits:A Social Network Approach”, Journal of personality,78(2):509–538, 2010.

ACKNOWLEDGMENTS

We have immense pleasure in presenting the synopsis report for our project entitled “Attitude Analysis through Social Media using Data Mining Techniques”

It is with great pleasure that we present the report on our project work at the end of 8th semester. We take this opportunity to share a few words of gratitude to all those who have supported us in making it possible. We extend our heartfelt gratitude to our project guide Dr. Radha Shankarmani for her able guidance and approachability. We would like to express our gratitude towards her constant encouragement, support and guidance throughout the development of the project. She was the one who never let our morale down and always supported us through our thick and thin. She is a constant source of inspiration for us and took utmost interest in our project. Our regular meetings proved to be a boon in the timely completion of this stage of the project.

We are very thankful to other teaching staff for their moral support and guidance throughout this final year.

We are also indebted to our college including the non-teaching staff for providing us with all the resources we required for completion of the project.