Personality Evaluation of Student Community using Sentiment Analysis

Ameen Khowaja[1], Mumtaz Hussain Mahar[2], Haque Nawaz1,[4], Shaukat Wasi[3] , Shafiq-ur-Rehman[4]

[1]Department of Computer Science, Sindh Madressatul Islam University, Karachi, Pakistan
[2]Department of Computer Science, Shah Abdul Latif University, Khairpur, Pakistan
[3]Department of Computer Science, Mohammad Ali Jinnah University, Karachi, Pakistan
[4]Department of Computer Science,SZABIST, Karachi, Pakistan

Summary
The social media is growing very rapidly, especially in the youth within the student community. The student community or youth communicates through their postings in form of textual data, images and videos to convey their messages within their circles of closely known and unknown friendships. The postings posted by the youth may be collected and used for the research purposes. In the large ocean of postings and the posters it is difficult and necessary to identify the personality traits. This identification may be useful to know about the students who is evaluating the teacher in an academic institution, a product reviewer who is reviewing and commenting on the product, the client feedback toward the product or the company, a teacher who is in process to build the foundation of students and other so many areas. Facebook is a very famous and common platform where youth feels comfortable to share their views in form of comments in textual form. We had tried to develop a tool on python language using three classifiers MNNB, RF and SVC to predicting or identifying the personality of students on social media using four labels including “sad”, “angry” (at negative side) and “happy”, “relax” (at positive side). A design-based approach has been used based on a dataset extracted from social media self-created page(s) where students of different mind-set were invited to comment on the questions relevant to their academics e.g. “The role of good grades is more important to build the career of student instead of having good practical knowledge?” or “In the evaluation system the home assignments should be discouraged by encouraging the classroom activities and participation”. In that case the dataset was collected fully a textual based having no emoticons used at all, the students expressed their views regarding the questions posted through social media page(s). Our approach succeeded to predict the personality with respect to posters Ids, time slots and shifts including morning, afternoon and evening using machine learning algorithms based on purely a textual dataset. In future the dataset and labels may be increased for more perfection in results to identify the personality, the personality may be predicted from the use of emoticons and roman English which is being used on social media very frequently.

Key Words:
Sentiment Analysis, personality Evaluation, personality identification, Text Classification.

1. Introduction

The reality cannot be denied that social media have become more popular within the student communities in comparison with all other communities. It became a very interesting platform within the communities of the world for communication, entertainment, social connections of various age groups, genders, researchers and the studies of various levels of students in most of the countries of the world. It is one of the favourites and famous live platform for researchers and other student communities to get the mass opinions about the researches, projects and specially to connect with the communities concerned to establish the social contacts within them for reviews and online guidelines, observing the behaviours etc.

The sentiment analysis is a technical way for identifying and categorizing point of views expressed in a textual form, especially toward to determine the attitude of the writer toward a topic in terms of Negative, Positive and Neutral states respectively [1]. The reviews may be toward a product or it could be a statement which is working independently describing any type of situation. The one of the best uses of sentimental analysis may be that, classify the normal and abnormal attitude of communities. The sentiment analysis is also a way to identify the user's opinions in form of views received in textual form as dataset. The decisions regarding the personality identification may be categorized like: happy, unhappy, sad, angry, positive, negative, refusal, dangerous, terrorist, funny, intelligent and other type of attitudes may be.

Sentimental analysis is also a process for extracting sentiments or views or opinions in textual form received as a message or textual opinion of the user, it also includes the deep analysis of the information which is hidden and covered as textual information, this hidden information is very important to insight the user’s behaviour whether positive or negative [3].

The sentimental analysis has a vast scope and became more popular by their evidential nature due to the speedy growth of electronic commerce, which is the dominant

Manuscript received March 5, 2019
Manuscript revised March 20, 2019

resource of expressing and analysing the point of views in form of text. Most of the clients depend on the reviews posted on the webs. Sentimental analysis is also known by the mining of views/opinions, behaviours concerned with various topics respectively. In developed and under developed countries the online social media replaces the off-line media rapidly, which encourages the people in country wide political, educational, socio-economic and other discussions, means the access of common people expanded globally to discuss various topics with their opinions in form of text [4].

It is also an art for processing the meaning less data and extract some meaningful information from it according to the settled parameters or categories. Data mining is widely used in different organizations / industries to make future oriented decisions for the rapid growth of the organizations. Data analysis is a way of verifying and transforming in order to retrieve meaningful information from the extracted data [5]. The one of excellent and famous methods of analysis as a technique is mining of data that concentrate on modelling and inventions of information for predicting way then the describing the purposes. Sentimental analysis extracts information automatically and analyses the subjects from the written text, this became a major area of research [5].The Sentimental Analysis belongs to the computational studies of people's views / opinions, both expressions may be interchanged, as some authors have reservations that they have some differences in meanings, The identification may be performed by the sentimental analysis while the opinion mining extracts people's opinion about a particular topic or subject [4] [7-9] [41] with respect to the step followed to identify whether the text is subjective or objective [6-7]. Most of users are keen interested to get the data from the e-solutions in form of online-advices and suggestions / opinion / reviews. The Sentimental Analysis is focused on behaviours, whereas text mining focused on the analytics of realities [7] [45].

The Sentimental analysis may also be known with the quality of mining the data which is focused to identify the textual representation of contents regarding their positivity and negativity might be expressed by different channels / platforms / peoples/authors toward their targets [8]. Most of the research for mining the opinion is focused on businesses and trade through websites and ecommerce applications. [9]. In Education the system for survey is we need the topic for extraction of statement and perform sentiment analysis on it / comments received through textual data [10]. It is more appropriate and possible through the algorithm of data mining, opinion-mining and sentiment analysis to find out the thinking of peoples of different regions [19].

While going through the literature reviews, we found that all the researcher of present and past have been worked a lot applied sentimental analysis on data mining for extracting the useful information from it using different datasets including the student dataset from the Facebook posts, profiles, but no or less have been succeeded on the evaluation and identification of the personalities of student communities on the social media on the basis of timeslot and shift, this research is well focused and have a distinguishing with the distinction to cover this issue and obtained the results. We tried our level best to identify the personality after evaluating the text through sentiment as well as the emotion analysis that will help the personnel or organizations to identify the student's personality on the basis of timeslot and shifts including morning, afternoon and evening that would be helpful taking proper decisions regarding the student community. Our research focuses on the following keywords with respect to timeslot and shift (Morning, Afternoon and Evening):

Sentiment Analysis, Emotion Analysis, Personality Evaluation, Personality Identification, Text Classification, Personality Evaluation of Student Community

1.1 Challenge of Emoticons identification in Emotion Analysis (SA)

The EA is a wide area for the process of natural language and opinion mining in form of text with the outcomes to determine the polarity of comments / posts [45], [51]. It is quite easy to determine the polarity when we are face to face because we feel, listen, see and observe, voice, tone, pitch, expressions on the basis of level of experience we may find the polarity and create an image of a person in our minds for future references, but in technological period of time people frequently use e-messages, emails, e-based network sites and emoticons. In this situation the text message can only provide the information or sense of information not the emotions at all, especially when the emotions are being expressed in a symbolic format like the emoticon codes being used on the social media [40] and display some images with respect to the states the human can have shown in table 1:

Table 1: Categories of Emotions

CATEGORY	EMOTIONS						
SAD							
ANGRY							
HAPPY							
RELAX							

And other in a different way to express the emotions on the e-culture based communication. In our research the dataset was different and comprised on purely textual data which was obtained from student community against the questions asked from them on the Facebook page, they have expressed their views in purely text format and no emoticons used at all.

1.2 Challenges with Personality identification

It would be a big challenge to identify and extract the sense from the symbols used for expression on Facebook [61]. We may see the symbols above are at least two or four which are used for a single expression like happy, kissing, cool, thumb is up and kidding as positive and sad or unhappy, confused or tensed, embarrassed, thumb is down and anger as negative.

The assigning polarity from a phrase as sentences may have ambiguous sense of information like: "I don't visit www.sad.com as I am quite happy "Here both "sad" and "happy" are used in a single phrase. So it would be again a challenge to extract the information and absorb the sensible meaning to assign the polarity.

1.3 Personality Types and Characteristics

As we know that, there is a big difference in personality type and their characteristics respectively. The personality type may be negative, positive, very negative, very positive and neutral in general, but the characteristics refer their properties. The personality is referred as the combination of characteristics which form a person's distinction. On the other side if the personality is being identified from unknown persons using a media like Facebook, twitter and other likelihood as platforms then it becomes critical to identify the personality. For that reason, we need the word and sentence data set with its description and context where these are used and when. There are some other called big-5 personality traits [71] might be Openness, conscientiousness, extraversion, introversion, extroversion, agreeableness, sensitiveness and neuroticism, which makes a person different from others.

In this research a personality can be identified through their lexica used on social media and lexica has many categories but here in this paper we considered and included the four main categories as: 0-Sad, 1-Angry, 2-Happy and at 3-Relax. The personality has many types and characteristics and traits, by which a personality can be obtained by applying the certain and appropriate process.

1.4 Personality Type v/s Characteristics and Traits

There may be many personality traits for identifying the personality out of them the most important and worldly recognized traits are:

Openness means the people who want to enjoy the adventures then they must have openness, openness means openness of experiences, they must be curious and appreciate the arts, imaginations and new things. People have less or no openness are just opposite and can't enjoy adventure even can't accept the adventures they stick or always happy with their habits, slow adaptation or avoidance toward new experiences and probably are not adventurous.

Conscientiousness means those who are well-organized and dutiful are conscientious, they are basically good planners and having good personalities.

Extraversion means more extravert and more socially active, chatty, sociable, they basically get energy from the crowds, assertive and cheerful in social interactions they are social butterflies.

Agreeableness measures the extent of a person's warmth and kindness. More agreeable the more trusting, helpful and compassionate. They are not much cold and suspicious.

Neuroticism these people basically worry about everything, they get tension on every small and big issues and become anxious and depressed very quickly, these peoples are basically very sensitive observe having even obsesses over germs and diseases. It becomes a major disease if reaches at its extreme.

Soft – Hatred Apathy Anger means annoyed, apathetic, bored, certain, cold, crabby, cranky, critical, cross, detached, displeased, frustrated, impatient, indifferent, peeved, imitated, and rankled come under the soft category of Hatred Apathy Anger.

Medium Hatred Apathy Anger means affronted, aggravated, angry, antagonized, arrogant, bristling, exasperated, incensed, indignant, inflamed, mad, offended, resentful, riled up and sarcastic come under the medium category of Hatred Apathy Anger.

Intense - Hatred Apathy Anger means aggressive, appalled, belligerent, bitter, contemptuous, disgusted, furious, hateful, hostile, irate, livid, menacing, outraged, ranting, raving, seething, spiteful, vengeful, vicious, vindictive, and violent come under Intensive category of Hatred Apathy Anger.

Soft – Shame and Guilt means abashed, awkward, discomfited, flushed, flustered, hesitant, humble, reticent, self-conscious, speechless and withdrawn come under Soft category of Shame and Guilt.

Medium – Shame and Guilt means Ashamed, Chagrined, Contrite, Culpable, Embarrassed, Guilty, Humbled, Intimidated, Penitent, Regretful, Remorseful, Reproachful, Rueful, and Sheepish come under Medium category of Shame and Guilt.

Intensive – Shame and Guilt Means Belittled, Degraded, Demeaned, Disgraced, Guild-Ridden, Guild-Stricken, Humiliated, Mortified, Ostracized, Self-Condemning, Self-Flagellating, Shamefaced, and Stigmatized come under Intensive category of Shame and Guilt.

Soft – Fear Anxiety and Panic Means Alert, Apprehensive, Cautious, Concerned, Confused, Curious, Disconcerted, Disoriented, Disquieted, Doubtful,Edgy,Fidgety,Hesitant,Indecisive,Insecure,Instinctive,Intuitive,Leery,Pensive,Shy,Timid,Uneasy, and

Watchful come under soft category of Fear and Anxiety and Panic.
Medium – Fear Anxiety and Panic means Afraid, Alarmed, Anxious, Aversive, Distrustful, Fearful,Jumpy,Nervous,Perturbed,Rattled,Shaky,Startled,Suspicious,Unnerved,Unsettled,Wary, and Worried come under medium category of Fear and Anxiety and Panic.
Intensive – Fear Anxiety and Panic Means Filled with Dread, Horrified, Panicked, Paralyzed, Petrified, Phobic, Shocked, and Terrorized come under Intensive category of Fear and Anxiety and Panic.
Soft - Jealousy and Envy Means Amused, Calm, Encouraged, Friendly, Hopeful, Inspired, Jovial, Open, Peaceful, and Smiling Upbeat come under the soft category of Jealousy and Envy.
Medium - Jealousy and Envy means Cheerful, Contented, Delighted, Excited, Fulfilled, Glad, Gleeful, Gratified, Happy, Healthy Self-esteem, Joyful, Lively, Merry, Optimistic, Playful, Pleased, Proud, Rejuvenated, Satisfied
Intensive - Jealousy and Envy Mean Awe-filled, Blissful, Ecstatic, Egocentric, Elated,Enthralled,Euphoric,Exhilarated,Giddy,Jubilant,Manic,Overconfident,Overjoyed,Radiant,Rapturous, Self-aggrandized,thrilled come under intensive category of jealousy and envy.
Happiness, Contentment and Joy Mean Contemplative, Disappointed, Disconnected, Distracted, Grounded, Listless, Low, Steady, Regretful and Wistful come under soft category of contentment and joy.
Happiness, Contentment and Joy Mean Dejected, Discouraged, Dispirited,Down,Downtrodden,Drained,Forlorn,Gloomy,Grieving, Heavy-hearted,Melancholy,Mournful,Sad,Sorrowful, Weepy and World-weary come under medium category of contentment and joy.
Happiness, Contentment and Joy mean Anguished, Bereaved, Bleak, Depressed, Despairing, Despondent, Grief-stricken, Heartbroken, Hopeless, Inconsolable and Morose come under medium category of contentment and joy.
Soft - Depression and Suicidal Urges mean Apathetic, Constantly Irritated, Angry, or Enraged (see the Anger list above),Depressed,Discouraged,Disinterested,Dispirited, Feeling Worthless,Flat,Helpless,Humourless,Impulsive,Indifferent, Isolated,Lethargic,Listless,Melancholy,Pessimistic,Purposeless, Withdrawn and World-weary come under soft category of Depression and Suicidal Urges
Medium - Depression and Suicidal Urges mean Bereft, Crushed, Desolate,Despairing,Desperate,Drained,Empty,Fatalistic,Hopeless,Joyless,Miserable,Morbid,Overwhelmed,Passionless, Pleasure less, Sullen come under medium category of Depression and Suicidal Urges.
Intensive - Depression and Suicidal Urges mean Agonized, Anguished, Bleak, Death-seeking,Devastated,Doomed,Gutted,Nihilistic,Numbed,Reckless, Self-destructive,Suicidal, Tormented and Tortured come under Intensive category of Depression and Suicidal Urges.
Beside all above traits, to identify the personality we could include some other traits which may be considered as the sub-categories of all above in general to finally disclose the personality on the basis of a dataset including positronic, poster comments and post time. The sub-categories are set and started from 0-Sad, 1-Angry, 2-Happy and 3-Relaxed. In the hypothesis it is clearly mentioned that the personality of a student community as a Facebook user may be declared with positive and negative. But by the sub-categories number 0 and 1 are the parameters of negative side (sad and angry) and the 2 and 3 are at positive side (happy and relaxed).
All the students have their own mindset for commenting on the social media, it depends and varies case to case and vicinity to vicinity. The students may behave differently in different timeslots, how the situation they are facing and what is the timeslot of comments they commented on. Based on above facts, the personality identification of student community became a question mark and difficult to identify. Various authors have given their input to identify the personality and opinion mining but not in a timeslot. The student is happy in morning and angry in afternoon however he is relaxed at evening and night so he may comment accordingly but the comment on same situation or question or post in different timeslots was not observed yet, by observing this the personality can be identified more accurately.
We decided to start our research on personality evaluation of student community in sentiment analysis and the same model may be applied on various other communities by extracting the data from the social media respectively.

2. Literature Review

The Sentiment Analysis is an approach to help the clients for making the decision more effective. It is quite difficult to analyse it manually that is why the sentiment analysis as well as the emotion analysis is applied for getting better and accurate results. In past so many researchers have been done the effort to analyse the data mining on the basis of sentiments, out of all some are referred as under: [12-18] [28-33] [38-40] [43] [50-52] [53] [56-57] and [62-63].
Keke Cai et al [12] proposed the technique for detection of contents, In the proposed approach, the contents are related to, two different types of opinions that are positive and negative opinions [55]. The major gap in this technique was analyzed regarding the classifier used with

it was not fulfilling the gap in the area "insight of what drives these sentiments". [13] introduced an opinion mining engine that used common-sense as a knowledge extracted from semantic Net and concept Net for performing the SA. He used a big dataset to extract the contents from news articles for testing the engine designed for data mining and applied the classifier which has obtained the results up to 71% accuracy with 91% for neutral opinion.

Federico Neri et al [14] worked on sentimental study. The applied dataset was comprised on 1000 live Facebook posts. the posts were relevant to the newscasts, comparison for sentiment of Rai (The Italian Public Broadcasting Service).

Ana CES. Lima et al [15] proposed the automated sentiment classifier for emotions as sentiment, the approach used for analysis was Naïve Bayes algorithm which was applied on the tweets containing the words. The gap in this approach was: classification for two aspects whether Positive or Negative while neglected the neutral state. Min Wang et al [16] focused on the polarity analysis of words have never been used, and the quantitative computation of sentiment words. The experiments shown was flexible and effective. ZHU Nanli et al [17]. Covered through their studies about the recent development in sentiment analysis. They have observed based on a survey conducted in three major fields: Sentiment Analysis, Feature Extraction and A framework.

Seyed-Ali Bahrainian et al [18] worked on the Sentiment Summarization and came up with the novel solution for SS and SA based on the tweets. The research was based on the comparison of various algorithms and methods for detecting the polarity of SA and SS. The major gap was in their research work was detection of sarcasm. [28] This work is focused on clustering-based sentiment analysis approach, claimed as new approach of SA. In the process TF-IDF weighting method was applied. They claimed that this technique has competitive advantages over the two existing approaches: symbolic techniques and supervised learning methods. As there were some drawbacks found in current techniques are: symbolic technique completely relies on the scores to generate the class of document, in that case the experiment shows relatively low accuracy rates. On the other side SLA have good results but this technique is very costly, however the larger amount of data for training needs to be pre-defined by classes manually [49]. While having a look on Table-II, there is no a big different in current and existing techniques (ST 65.83%, SLA 77% - 88%, while CBA 77.17% - 78.33%), but the great advantage of the current approach is it is faster than both with no human participation. Tan Li Im et al proposed lexicon approach is used to analyse the financial news, sentiments particularly for positive and negative polarity. In this paper lexical based approach is proposed to identify and extract the sentiments from the financial news. Two experiments conducted, one is non-stemmed token and obtained 74% accuracy and stemmed token with 79.1% accuracy. The gap and a big drawback in this approach is the word before and after the targeted word may affect the sentiment of the word, means this approach couldn’t find the solution to detect the sentiment by considering a single word. In future they committed to fill this gap for solving the problem of phrase or more than one word.

S. Setty et al [30] classification of user’s posts on Facebook was identified as a problem and applied classifier (SVM) for classifying the Facebook news feeds. First of all, news feed fetched from the Facebook and tagging classifier built which classified the Facebook news feeds based on the tag’s category comprised on liking page posts and friend’s posts. Various classifiers built to classify the news feeds / posts and the algorithm such as binary logistic regression, naïve Bayes, SVM, Bias Net and J48 used to achieve the better results, it further classifies the friend’s posts, event posts and entertaining posts. Comparatively Weka and our learning model on the basis of 200 Facebook posts as dataset, the SVM with our learning model got the better and more accurate results [30]

Kaili Mao et al [31] novel sentiment analysis method which combines Lexicon-based and Learn-based techniques (CLL) for analysing the cross-domain sentiment of reviews for chines product. The corpus used in this demonstration were books, hotels and electronics, and four categories of features comprised of 16 features in total for building six classifiers. The research claims that CLL performs better for books and hotels while low in performance for electronics.

Chien-Liang Liu [33] designed and developed a review-summarization and movie-rating system in the mobile environment. Sentiment classification is applied on rating information as well as in movies reviews based on classification results. They claim that this approach played a vital role, in this research author proposed novel based approach LSA to analysis the related features about product, and moreover statistical method identifies the sentiment of word(s). Sentiment words and product related feature will be used on the basis of feature-based summarization. In this experiment for performance analysis, the SVM model is applied with the number of features that classify sentiment reviews and takes less than 6 seconds to load. This filtering approach used to reduce the size of summary based on the user’s preferences by various aspects. The LSA-Based design proposed fully utilizes the internet contents.

Dejan Markovikj et al. [53] worked on data mining of facebook statuses for personality predictive model using SVM, SMO, MBAB, ABM1of 250 FB users. This model was limited to analyse the relevance and sensitivity of

investigation for indicators prediction and personality traits.
Randall Wald and Taghi Khoshgoftaar [49] worked on personality prediction through machine of Facebook profiles using DM and ML techniques on big-5 model of personality traits on the basis of 537 Facebook users and 45 questions. The model presented by them was limited due to small dataset and personality evaluation through the automated data mining.
Adyan Marendra Ramadhani [54] used MLP (Multilayer Perceptron) with regular MLP settings. In this they used 1000 datasets for positive and negative for testing and training of the system. The database in total is about 4000. The experiment is highly trained with 100 epochs. The accuracy for testing and training datasets for twitter SA is based on deep learning method with the results: Trains: 77.45% and Test: 75.03%. However, the accuracy for testing and training datasets for twitter SA using MLP with the results: Train: 67.45% and Test: 52.60%. The problem demonstrated by the author in this paper was a huge amount of unstructured data was a difficult task to extract and identify regarding the sentiment analysis, that is why they used the deep learning techniques based on feed-forward neural network with many hidden layers in terms of neural network with the results about 70% accuracy for achieving better results.
Fazel Keshtkar et al. [57] worked using data mining technique for detecting personality of players in an educational game using machine learning techniques (NB, RF and DT). The data set was Leary's Rose framework comprised on 200 student's textual data excerpts from student chat. This model was also limited to apply experiment by extending the model on big-5 personality traits.
Kuei-Hsiang Peng et al. [59] worked on predicting personality traits of Chinese users based on Facebook wall posts of 222 FB users using the SVM technique but the system was limited to find accuracy in in classification of the remaining four factors except extroversion as a trait.
Jeevanandam and S.Koteeswaran [62] worked on Feature Selection using Random Forest Method for Sentiment Analysis using Naïve Bayes, LVQ and Random Forest with the combination of MLNN for classification however the model was limited to perform with LVQ
Elif Uysal et al. [63] performed sentiment analysis on General Elections in Turkey "A Case Study" and used Stemming and number of NLP operations using tweeter dataset of 93653 tweeters but the software used for results was limited to handle negative verbs and having less dataset for sentiment words.
Danny Azucar et al. [71] worked on prediction of big 5 traits for personality from digitized footprints of social media using the analysis of metadata, the system was limited in performing the comparison of accuracy for personality identification and prediction on specific social media platforms.

3. Research Methodology

This research is followed by experimental methods by using supervised techniques of machine learning and found the personality of student community based on categorization, filtering and classifiers. The filters and classifiers applied on the comments and evaluated the personality of student community. Dataset was extracted from Facebook page(s) about 2706 comments collected for two month's timespan and extracted for the dataset. Out of 100% dataset, about 70% data assigned to the training set and rest approximately 30% data assigned to the testing set. The personality identified based on subcategories such as; sad and angry, happy and relax.

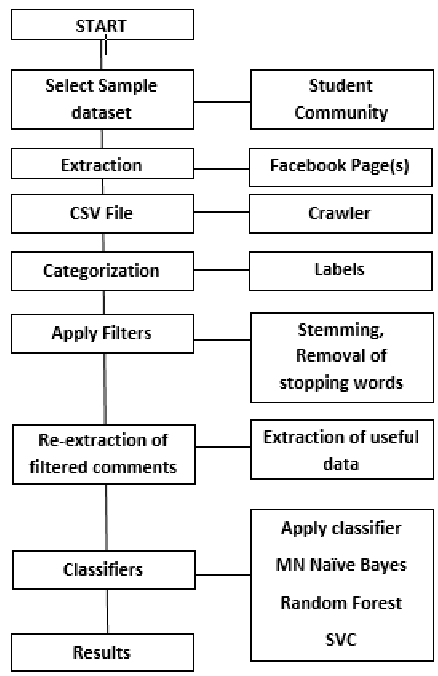

Fig. 1 Designed Model

3.1 Data Collection

The dataset was required for this problem was not only the words which are positive and negative but the sentences / phrases which was frequently being used on the Facebook by the student communities, for that reason we have focused on extraction of the data from Facebook pages of different universities. In this regard some issues were critical to face during the extraction which were extraction was restricted by the page owner, an extracting tool was needed to extract the data from Facebook pages, preparing of extracted data into the csv comma delimited format which could be required for removal of stopping words, filtering of positive and negative words, filtering of sub-personality traits like 0-Sad, 1-Angry, 2-Happy and 3-Relaxed which was focus of this paper to predict the personality of students based on the comments posted on the Facebook with respect to the shift and timeslot. This whole process was needed, and a huge dataset comprised on long and short phrases were also required for applying the algorithms. So a database comprised on phrases / comments gathered, crawled/extracted. The dataset was collected on the basis of questionnaire relevant to the academia and the student community where students have posted the comments in simple English and expressed their emotions without using the emoticons, we have simply removed stopping words, categorized the dataset in four labels manually the classified them into sub-personality traits and obtained the results using three famous text classifiers that are: SVC, Random Forest and Naïve Bayes. The dataset comprised on approximate 2706 comment of almost 25 to 30 student's / Facebook users collected during the two months' timespan from October 2018 to December 2018, the data was collected from Facebook pages of various universities, which provided the post, their comments and other relevant data including the poster ID. As the scenario was to collect the data and classify but we all know that there is a difference in personality types and their characteristics at different levels like soft, medium and intensive levels of traits could be considered, out of so many traits some well-known traits considered as the sub-categories of positive and negative traits in general are:

3.2 Sub-Categories as labels in dataset

The deciding following sub-categories were a challenge to identify. For that a huge dataset was needed with different categories of comments posted by the student community within different timespan. The sub-categories included in this research are very richer and relevant to the behaviour or nature of a person. The nature is kind and created personalities of student/person that can be changed through the situations passed through them with the different timespan. For example: a person is negative in the morning because he could not enjoy his resting time at night and commented negatively but same person is positive in the noon or even in after noon because he heard a good news in the university regarding good results or from anywhere else, and he becomes negative again in the night because he is restless or afraid of anything at home, in that case the personality traits may be categorized further with emotions and sub-categories for observing more accuracy to predict the personality [68-69]. Following categories or we may say the labels in the dataset would be helping us to decide the personality of student are:

3.2.1 Sad(0)

This category includes the sadness of a student in any circumstances like sometime he/she/they is/are afraid, apologetic, awkward, bewildered, concerned, confused, dark, depressed, despondent, discontented, discouraged, dismayed, disparaging, jealous, lonely, lazy, spiritually weak, physically weak, mentally weak, pitiful, poor, depressed, suppressed, insecure, gloomy, fussy, fearful, dull, faithless, distressed, dissatisfied, nervous, ill, sick, quarrelsome, dismayed, doubtful, illegal at first time, unforgiving, vacant, jobless, worried, withdrawn, tired, sorry for something done wrong, sensitive, serious, restless quiet, solitary, reserved, unfriendly, unhappy, shocked and many other states which a person can face and become sad.

3.2.2 Angry(1)

This category may include the bad-tempered, irritable, annoyed, impatient, angry, unhappy, disagreed, unrecognized though experienced, upset, touchy, moody, short-tempered, sensitive, crazy, ego, hot-tempered, fractious, temperamental, irascible, ignored, negligence, inflexible, anal, bilious, cantankerous, censorious, choleric, choosy, selective, crabby, cranky, crusty, crotchety, demanding, difficult, excitable, fiery, finicky, foul, bad-mood, financially weak, hot-headed, hypercritical, ill-tempered, jaundiced, insatiable, miserable, nagging, not suffer fools gladly, obstreperous, ornery, particular, pedantic, peevish, peppery, pernickety, pernickety, petty, pettifogging, picky, prickly, querulous, quick-tempered, ratty, snappy, stroppy, testy, tetchy, truculent, unforgiving, revengeful, uptight and many other states which a person can have and become angry.

3.2.3 Happy(2)

This category may include the happy, bright, growing, excited, exuberant, forgiving, fortunate, lucky, friendly, laughs, smiles, funny, joyous, healthy, restful, strong, quick, hopeful, humorous, innovative, imaginative, affectionate, adventurous, ambitious, animated, animated, charming, charismatic, dreamy, eager, easy-going, energetic, giddy, cheerful, zany, wild, unconcerned, warm, helpful, social, thrilled, tireless, trusting, thoughtful, thankful, talkative, sweet, supportive, self-confident, secure, rich, rewarded, awarded, skilled, recognized, positive, playful, peaceful, optimistic and many other states which a person can have and become happy.

3.2.4 Relax(3)

This category may include the so what, chill, cool, no worries, relax, relaxed, satisfied, satisfaction, no work, less work, easy, enjoying, joy, picnic, pleasure trip, trip, confident, not stressed, less stressed, go ahead, can do, take-it-easy, ease, smile, smiles, smiled, fun, joyous, healthy, energetic, synergetic, no-ambitions, cheer, drink, drunk, tireless and many others states which a person can have and become relax.

3.3 Testing and Training of dataset

All the data passed through two famous strategies and assigned almost the 70% of data for training through which the system could be trained and ready for testing, and rest of 30% of dataset could be assigned for testing.

3.4 Tools and Techniques

As this paper is based on the experimental approach to find the results, for this we needed different tools such the Facebook pages with the permission to extract the comments from pages, the extractor which could be used to extract the data from the pages, transformation of extracted data, text classification, pre-processing, categorization and finalizing the results. The provided data was in csv format and could be converted in any of csv or .xlsx or .xls format but needed to be arranged as required in any spread sheet, however; using NLTK there are various algorithms are available for pre-processing and classification, out of all we used the Naïve Bayer (a classifier), Random Forest (a classifier) and the SVC (Support Vector Machine a classifier) as classifier which have distinction in classifying the textual data above all. These algorithms helped to classify the collected data in categories labelled in the dataset. In the pre-processing tokenization, stemming is also applied for splitting the data to separate the inputted words and remove the suffix from these words. The WordNet dictionary is also used for increasing the range of words and comparing synonyms of inputted words for getting better results.

3.5 Tokenization

The process of tokenization has great importance for separating the actual words from the dataset and splitting the sentences which are including punctuations and unnecessary verbs should be discarded. This is very important for lexical based sentiment and emotion analysis.

3.6 Stemming

In pre-processing the stemming plays a key role for reducing the derived words in suffix to their stem of words. For that stemming algorithm is also applied to remove suffix from given dataset.

In some cases, stemming has ambiguous results like when stem applied on the "university" and "universe" both stemmed as "universe" but generally stemming is very supportive to stem the suffix with the accuracy for removal of suffix.

3.7 Removal of Stop words

The removal of stop words was an important task before the classification of the extracted data which was including some irrelevant or stopping characters which had no meaning or sense like @ , . ! ? etcetera, and other stopping words like the , there, then, what, when, where, why, who, bhai (a roman urdu word(like brother in English language), and other so many words which could be needed to be removed from the actual dataset to make the extracted comments more perfect for sentiment analysis and categorization. Most of stop worlds removed through predefined algorithm of NLTK library and are shown below:

	PST_ID	PSTR_ID	LABEL	COMMENT	TIME	Comments_without_stopwords
0	1.0	A101	2	MY BEST WISHES : venue?? I knew It that you ar...	8:56	MY BEST WISHES : venue?? I knew It good man he...
1	1.0	A102	3	venue plz : I will try to come !	9:56	venue plz : I try come !
2	1.0	A103	1	Why You have to come bhai	9:22	Why You come bhai
3	1.0	A104	1	no its not suitable time : always you set the ...	8:56	suitable time : always set programs extempore
4	1.0	A105	2	Try to understand and be part of someone's hap...	9:56	Try understand part someone's happiness : read...

Fig. 2 Removal of Stop Words

3.8 Categorization

In this process the extracted comments are categorized in different sub-categories like (0) Sad (Approx. 1222 comments), (1) Angry (Approx. 637 comments), (2) Happy (Approx. 425 comments) and (3) Relax (Approx. 422 Comments) using the three most famous algorithms such as; Naïve Bayes, Random Forest and SVC.

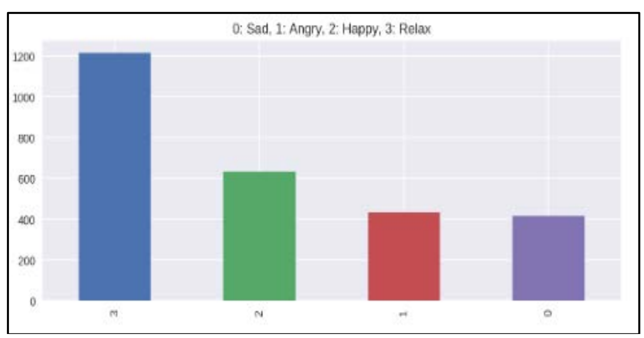

Fig. 3 Size of dataset with labels

The above image is the pictorial representation of the dataset which is used for training and testing, comprised on four categories. The category 0 and 1 belong to the negative however; 2 and 3 relevant to the positive side of the sentiment analysis.

3.9 Machine Learning Techniques

3.9.1 Naïve Bayes Classifier

These classifiers are the probabilistic classifiers which applies the Bayes theorem with the strong naïve used for assumption among the features, the conditional probability can calculate the probability of the event sing its predefined knowledge. It contains the probability of hypothesis, evidence, evidence given that hypothesis is true and the probability of the hypothesis that the evidence is there. The main reasons to apply NB in this research are:

- Simple and very efficient yet.
- The objective of NB comes from assumption that the database features are independent mutually.
- It is used for normalization of results, removing that does not effect.
- If dependence is condition and NB actually holds the same then the Naïve Bayes will quickly converge than other models such as regression.
- We need less dataset to obtain the best results.
- Best for text classification.
- It is faster, a model which is highly scalable, it scales linearly on the basis of number of rows and predictors and process is parallelized.
- A supervised technique.

3.9.2 Random Forest

Random Forest is a supervised technique or learning algorithm, having a big advantage is that it can be used for both classification and the regression problems. In the classification it predicts the class of the given data in 0s and 1s it is also known as the binary classifier. In the regression algorithm it predicts the discrete values that can be used to identify the linear relationship between attributes. The reason for selecting the RF in this research for obtaining results are:

- Faster algorithm in comparison to other machine learning techniques
- Able to deal with imbalances and missing data in dataset.
- A supervised technique.
- Perform on the number of trees

3.9.3 Support Vector Machine

A supervised ML technique which works for text classification and the regression, it is mostly preferred that it should be used for classification as compared to regression for better results.
In this paper SVC (support vector classifier) is used which is preferred for data mining. The reasons for selecting this technique for text classification in this research are:

- Best fits the data you provide
- Supervised ML technique used for text classification
- Can be differently implemented using same technique.
- It finds the boundary which is optimal between the most possible outputs
- Transforms the inputted data into the desired form.
- Used for prediction and can be applied in linear and nonlinear problems.

4. Results and Discussion

4.1 Precision and Recall

Precision and Recall is a fractional technique, the "precision" is used for the prediction of positive values, it finds the relevance of instances among the obtained instances from the dataset, while the relevant instances which have been retrieved from the total instances is called the "recall". The exactness of classifier can be given by precision that can be calculated through [43]:

$$Precision\ = \frac{TP\!: True\ Positive}{TP\!: True\ Positive + FP\!: False\ Positive}$$

$$Recall = \frac{TP\!: True\ Positive}{TP\!: True\ Positive + FN\!: False\ Nagtive}$$

	precision	recall	f1-score
Sad	0.47	0.62	0.54
Angry	0.49	0.62	0.55
Happy	0.39	0.34	0.36
Relax	0.62	0.53	0.57
micro avg	0.52	0.52	0.52
macro avg	0.49	0.53	0.51
weighted avg	0.52	0.52	0.52

Fig. 4 Precision and Recall MultinomialNB

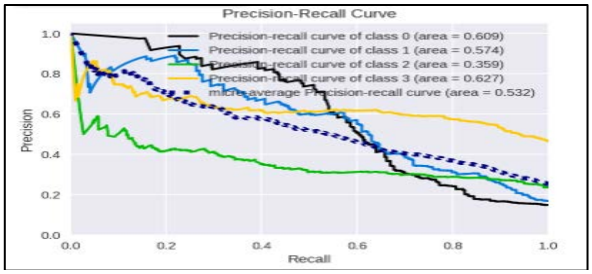

Fig. 5 Precision and Recall Curve MultinomialNB

	precision	recall	f1-score
Sad	0.90	0.69	0.78
Angry	0.80	0.72	0.76
Happy	0.57	0.57	0.57
Relax	0.71	0.78	0.74
micro avg	0.71	0.71	0.71
macro avg	0.74	0.69	0.71
weighted avg	0.72	0.71	0.71

Fig. 6 Precision and Recall Random Forest

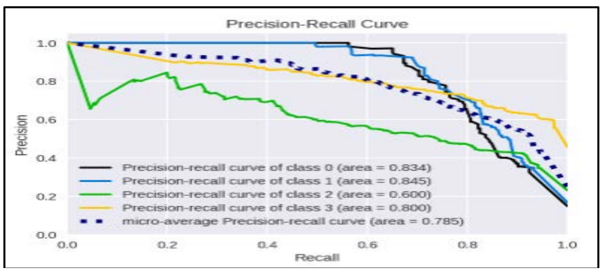

Fig. 7 Precision and Recall Curve Random Forest

	precision	recall	f1-score
Sad	0.86	0.70	0.78
Angry	0.83	0.73	0.78
Happy	0.57	0.55	0.56
Relax	0.70	0.79	0.75
micro avg	0.71	0.71	0.71
macro avg	0.74	0.69	0.71
weighted avg	0.72	0.71	0.71

Fig. 8 Precision and Recall Linear SVC

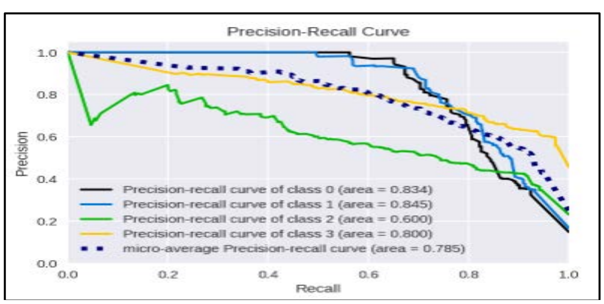

Fig. 9 Precision and Recall Curve Linear SVC

4.2 Confusion Matrix

The accuracy can be obtained through a performance evaluation process known as accuracy, the accuracy can be measured through true positive plus true negative divided by true positive + true negative + false positive +false negative [43].

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{TP + TN}}{\boldsymbol{TP + TN + FP + FN}}$$

In fig 10 matrix shown in the results below predicting the true positive and true negatives in the matrix, however; other both sides are depicting the false positive and false negative. Model is tested on 892 which is 033% of total comments in the dataset, in which the diagonal values contain 0.62% is obtained for true negative, 0.62% is obtained for true positive however 0.11% is obtained for false positive and the remaining 0.9% is obtained for false negative for label 0 and 1. Similarly for labels 2 and 3, 0.34% is obtained for true negative, 0.53% is obtained for true positive, 0.4% is obtained for false positive and 0.21% is obtained for false negative.

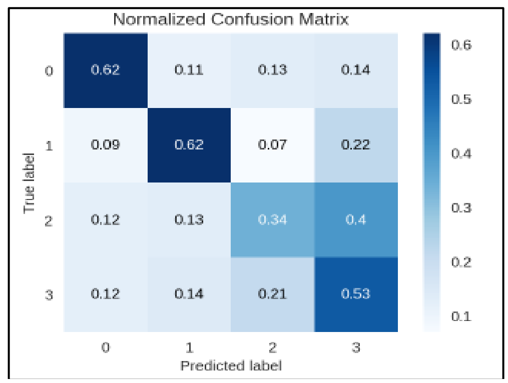

Fig. 10 Confusion Matrix MultinomialNBClassifier

4.3 Random ForestClassifier Confusion Matrix Results

In fig 11, the diagonal sides of the matrix show the correct prediction of model as true positive and true negative. However; two adjacent sides are depicting the false positive and false negative. Model is tested on 892 (0.33% of total dataset) comments, in which the diagonal values contain 0.7% is obtained for true negative, 0.73% is obtained for true positive however 0.02% is obtained for false positive and the remaining 0.02% is obtained for false negative for label 0 and 1. Similarly for labels 2 and 3, 0.55% is obtained for true negative, 0.78% is obtained for true positive, 0.4% is obtained for false positive and 0.15% is obtained for false negative.

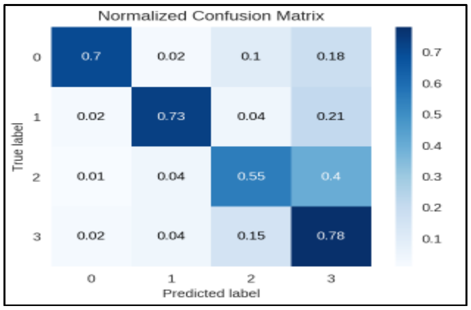

Fig. 11 Confusion Matrix Random Forest Classifier

4.4 LinearSVC Confusion Matrix Results

In fig 12, the matrix predicted the model as true positive and true negative. However; like the above two models, two adjacent sides are demonstrating the false positive and false negative. Similarly, the model is tested on 892 (0.33% of total dataset) comments, in which the diagonal values contain 0.7% is obtained for true negative, 0.73% is obtained for true positive however 0.03% is obtained for false positive and the remaining 0.02% is obtained for false negative for label 0 and 1 and Similarly for labels 2 and 3, 0.55% is obtained for true negative, 0.8% is obtained for true positive, 0.41% is obtained for false positive and 0.15% is obtained for false negative.

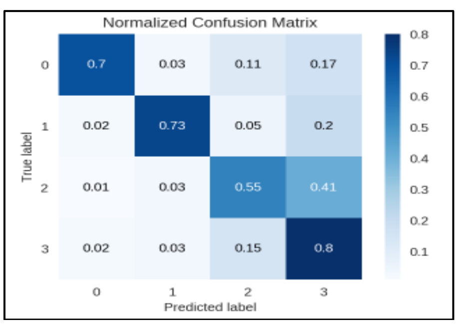

Fig. 12 Confusion Matrix Linear SVC

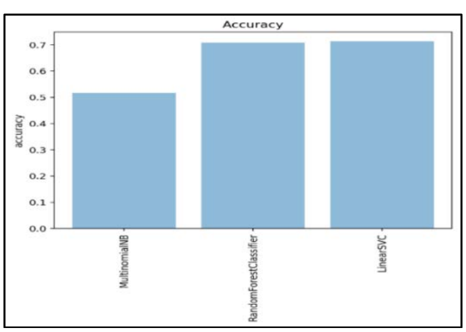

Fig. 13 Model's Accuracy Results

In fig. 14 demonstrates the retrieval of a particular id# 102 of a studentthe said student is relaxed in all three shifts with respect to the timeslot.

	PSTR_ID	COMMENT	EMOTION	TIME
0	A102	venue plz : I try come !	Relax	9:56
1	A102	May blessed always, respect others	Relax	21:00
2	A102	even human may ghost	Relax	8:56
3	A102	let inform higher management soon	Relax	0:00
4	A102	generous offer sir	Relax	20:00
5	A102	tension may release trips	Relax	9:56
6	A102	Practical work important explore it.	Relax	8:56
7	A102	But provide practical exposure.	Relax	22:00
8	A102	theoretical learning easy practical practical ...	Sad	5:34
9	A102	In opinion best grades experience inportant ma...	Relax	5:11

Fig. 14 Comments of a particular ID#102 with Emotion and Timeslot

In fig 15 the results show that the id# 102 is relaxed in all shifts with respect to the timeslots for morning, afternoon and evening.

Fig. 15 Comments of a particular ID#102 with Emotion and Timeslot

5. Conclusion

The evaluation of a personality especially in case of student community was an issue since the social media was emerged and got popularity very speedily. The popularity of social media cannot be ignored, but the social media fully ignored the hidden identity of persons and their personalities, hidden genders and the differences, inaccuracy in ages and the differences, and restriction over the comments posted by the persons / students or any other community, the vulgarity, unnecessary trainings of individuals especially the students who are under age. In that unclear or ambiguous scenarios, the identification of personality of students was a big issue. This paper covers the identification of personalities of student community on the basis of personality traits using Naïve Bayes, Random forests and SVC the most popular text classifiers for classifying the data posted on the Facebook and disclose the personality as Sad, Angry, Happy and Relaxed. The future work for this domain may be to increase the traits and sub-categories for personality identification and make it more specific and accurate.

References

[1] Sinha, Harsha, and Arashdeep Kaur. "A detailed survey and comparative study of sentiment analysis algorithms." Communication Control and Intelligent Systems (CCIS), 2016 2nd International Conference on. IEEE, 2016

[2] Rani, Sujata, and Parteek Kumar. "A Sentiment Analysis System to Improve Teaching and Learning." Computer 50.5 (2017): 36-43.
[3] Kaur, Harpreet, and Veenu Mangat. "A survey of sentiment analysis techniques." I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), 2017 International Conference on. IEEE, 2017.
[4] Ravi, Kumar, and Vadlamani Ravi. "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-Based Systems 89 (2015): 14-46.
[5] Abirami, A. M., and V. Gayathri. "A survey on sentiment analysis methods and approach." Advanced Computing (ICoAC), 2016 Eighth International Conference on. IEEE, 2017.
[6] Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. "Sentiment analysis algorithms and applications: A survey." Ain Shams Engineering Journal 5.4 (2014): 1093-1113.
[7] [Vinodhini, G., and R. M. Chandrasekaran. "Sentiment analysis and opinion mining: a survey." International Journal 2.6 (2012): 282-292.
[8] Alhojely, Suad. "Sentiment Analysis and Opinion Mining: A Survey."
[9] Nasim, Zarmeen, Quratulain Rajput, and Sajjad Haider. "Sentiment analysis of student feedback using machine learning and lexicon based approaches." Research and Innovation in Information Systems (ICRIIS), 2017 International Conference on. IEEE, 2017.
[10] Aung, Khin Zezawar, and Nyein Nyein Myo. "Sentiment analysis of students' comment using lexicon based approach." Computer and Information Science (ICIS), 2017 IEEE/ACIS 16th International Conference on. IEEE, 2017.
[11] Vo, Hung T., et al. "Topic Classification and Sentiment Analysis for Vietnamese Education Survey System" (2016).
[12] Keke Cai, Scott Spangler, Ying Chen, Li Zhang., "Leveraging SentimentAnalysis for Topic Detection", IEEE/WIC/ACM InternationalConference on Web Intelligence and Intelligent Agent Technology,2008, pp. 265-271.
[13] Prashant Raina., "Sentiment analysis in News articles using Senticomputing",1EEE 131h International conference on Data MiningWorkshops, 2013, pp.959-962.
[14] Federico Neri, Carlo Aliprandi, Federico Capeci, Monstserrat Cuadros,Tomas, "Sentiment Analysis on social media", IEEE/ACM Internationalconference on Advances in Social Networks analysIs andmining,20 12,pp. 919-926.
[15] Ana C. E.S.Lima, Leandro N.de Castro., "Automatic Sentiment Analysis oftwitter Messages", Publisher IEEE, 2012, pp. 52-57.
[16] Min Wang, Hanxio Shi., "Research on Sentiment Analysis Technology and Polarity Computation of Sentiment words", Publisher IEEE, 2010,pp.331-334.
[17] ZHU Nanli, ZOU Ping, Ll Weign, CHENG Meng., "Sentiment Analysis: A Literature Review", in proceedings of the 2012 IEEE ISMOT,pp. 572-576.
[18] Seyed-Ali Bahrainian, Andreas Dengel., "Sentiment Analysis and Summarization of Twitter Data", IEEE 161h International Conference onComputational Science and Engineering, 20 13,pp. 227-234.
[19] Pang, Bo, and Lillian Lee. "Opinion mining and sentiment analysis." Foundations and Trends® in Information Retrieval 2.1–2 (2008): 1-135.
[20] Asur, Sitaram, and Bernardo A. Huberman. "Predicting the future with social media." Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on. Vol. 1. IEEE, 2010.
[21] Kumar, Akshi, and Mary Sebastian Teeja. "Sentiment analysis: A perspective on its past, present and future." International Journal of Intelligent Systems and Applications 4.10 (2012): 1.
[22] B. O'Connor, M. Krieger, D. Ahn. TweetMotif: Exploratory Search and Topic Summarization for Twitter. ICWSM-2010.
[23] K. Gimpel, N. Schneider, B. O'Connor, D. Das, D. Mills, J. Eisenstein, N.A. Smith, Part-of-speech tagging for twitter: annotation, features, and experiments, in: Carnegie-Mellon Univ Pittsburgh Pa School of Computer Science, 2010.
[24] L. Derczynski, A. Ritter, S. Clarke, and K. Bontcheva. 2013, Twitter Part-of-Speech Tagging for All: Overcoming Sparse and Noisy Data, In Proceedings of the International Conference on Recent Advances in Natural Language Processing, ACL.
[25] O. Owoputi, B. O'Connor, C. Dyer, K. Gimpel, N. Schneider, and N.A. Smith, Improved Part-of-Speech Tagging for Online Conversational Text with Word Clusters, In HLT-NAACL, pp. 380-390. 2013.
[26] K. Atkinson, Gnu Aspell 0.60.4 (2006).
[27] M.F. Porter, Snowball: A language for stemming algorithms, (2001).
[28] Li, Gang, and Fei Liu. "A clustering-based approach on sentiment analysis." Intelligent Systems and Knowledge Engineering (ISKE), 2010 International Conference on. IEEE, 2010.
[29] Im, Tan Li, et al. "Analysing market sentiment in financial news using lexical approach." Open Systems (ICOS), 2013 IEEE Conference on. IEEE, 2013.
[30] Setty, Shankar, et al. "Classification of facebook news feeds and sentiment analysis." Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on. IEEE, 2014.
[31] Mao, Kaili, et al. "Cross-domain sentiment analysis of product reviews by combining lexicon-based and learn-based techniques." High Performance Computing and Communications (HPCC), 2015 IEEE 7th International Symposium on Cyberspace Safety and Security (CSS), 2015 IEEE 12th International Conferen on Embedded Software and Systems (ICESS), 2015 IEEE 17th International Conference on. IEEE, 2015.
[32] Liu, Chien-Liang, et al. "Movie rating and review summarization in mobile environment." IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 42.3 (2012): 397-407.
[33] Ramadhani, Adyan Marendra, and Hong Soon Goo. "Twitter sentiment analysis using deep learning methods." Engineering Seminar (InAES), 2017 7th International Annual. IEEE, 2017.
[34] Aung, Khin Zezawar, and Nyein Nyein Myo. "Sentiment analysis of students' comment using lexicon based approach." Computer and Information Science (ICIS), 2017 IEEE/ACIS 16th International Conference on. IEEE, 2017.
[35] Nitin, Gokarn Ila, Gottipati Swapna, and Venky Shankararaman. "Analyzing educational comments for topics and sentiments: A text analytics approach." Frontiers in Education Conference (FIE), 2015 IEEE. IEEE, 2015.

[36] Ortigosa, Alvaro, José Ignacio Quiroga, and Rosa M. Carro. "Inferring user personality in social networks: A case study in Facebook." Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on. IEEE, 2011.
[37] Chen, Baiyun, and Justin Marcus. "Students' self-presentation on Facebook: An examination of personality and self-construal factors." Computers in Human Behavior 28.6 (2012): 2091-2099.
[38] Ortigosa, Alvaro, Rosa M. Carro, and José Ignacio Quiroga. "Predicting user personality by mining social interactions in Facebook." Journal of computer and System Sciences 80.1 (2014): 57-71.
[39] Settanni, Michele, and Davide Marengo. "Sharing feelings online: studying emotional well-being via automated text analysis of Facebook posts." Frontiers in psychology 6 (2015): 1045.
[40] Vashisht, Geetika, and Sangharsh Thakur. "Facebook as a corpus for emoticons-based sentiment analysis." Int. J. Emerg. Technol. Advan. Eng 4 (2014): 904-908.
[41] Esuli, Andrea, and Fabrizio Sebastiani. "SentiWordNet: a high-coverage lexical resource for opinion mining." Evaluation (2007): 1-26.
[42] Nielsen, Finn Årup. "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs." arXiv preprint arXiv:1103.2903 (2011).
[43] Tripathy, Abinash, Ankit Agrawal, and Santanu Kumar Rath. "Classification of Sentimental Reviews Using Machine Learning Techniques." Procedia Computer Science 57 (2015): 821-829.
[44] Louis, F.and Wojciech, C., Book "Advances in The Human Side of Service Engineering", 5th International Conference on Applied Human Factors and Ergonomics, Volume Set, Proceedings of the 5th AHEE Conference 19-23 July 2014.
[45] Liu, B., Sentiment Analysis and Opinion Mining. Morgan and Claypool Publishers, 2012.
[46] Yin, Z., Rong, J., and Zhi-Hua, Z., "Understanding Bag-of-Words Model: A Statistical Framework", International Journal of Machine Learning and Cybernetics, 2010.
[47] Nisha, J., and Dr.E. Kirubakaran, "M-Learning sentiment analysis with Data Mining Techniques", International Journal of Computer Science and Telecommunications, Volume 3, Issue 8, 2012.
[48] Richard, S., Alex, P., Jean, Y.W., Jason, C., Christopher, D.M., Andrew, Y.N. and Christopher, P., "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", Conference on Empirical Methods in Natural Language Processing, 2013.
[49] Wald, Randall, Taghi Khoshgoftaar, and Chris Sumner. "Machine prediction of personality from Facebook profiles." Information Reuse and Integration (IRI), 2012 IEEE 13th International Conference on. IEEE, 2012.
[50] Pandey, Umesh Kumar, and Saurabh Pal. "Mining data to find adept teachers in dealing with students." International Journal of Intelligent Systems and Applications 4.3 (2012): 27.
[51] Sobkowicz, Pawel, Michael Kaschesky, and Guillaume Bouchard. "Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web." Government Information Quarterly 29.4 (2012): 470-479.
[52] Poria, Soujanya, et al. "Common sense knowledge based personality recognition from text." Mexican International Conference on Artificial Intelligence. Springer, Berlin, Heidelberg, 2013.
[53] Markovikj, Dejan, et al. "Mining facebook data for predictive personality modeling." Proceedings of the 7th international AAAI conference on Weblogs and Social Media (ICWSM 2013), Boston, MA, USA. 2013.
[54] Lima, Ana Carolina ES, and Leandro Nunes De Castro. "A multi-label, semi-supervised classification approach applied to personality prediction in social media." Neural Networks 58 (2014): 122-130.
[55] Zuber, Mohammad. "A survey of data mining techniques for social network analysis." International Journal of Research in Computer Engineering and Electronics 3.6 (2014).
[56] Ortigosa, Alvaro, José M. Martín, and Rosa M. Carro. "Sentiment analysis in Facebook and its application to e-learning." Computers in human behavior 31 (2014): 527-541.
[57] Keshtkar, Fazel, et al. "Using data mining techniques to detect the personality of players in an educational game." Educational Data Mining. Springer, Cham, 2014. 125-150.
[58] Mahajan, Chetan, and Preeti Mulay. "E3: effective emoticon extractor for behavior analysis from social media." Procedia Computer Science 50 (2015): 610-616.
[59] Peng, Kuei-Hsiang, et al. "Predicting personality traits of Chinese users based on Facebook wall posts." Wireless and Optical Communication Conference (WOCC), 2015 24th. IEEE, 2015.
[60] Sagadevan, Saravanan, Nurul Hashimah Ahamed Hassain Malim, and Mohd Heikal Husin. "Sentiment valences for automatic personality detection of online social networks users using three factor model." Procedia Computer Science 72 (2015): 201-208.
[61] Novielli, Nicole, Fabio Calefato, and Filippo Lanubile. "The challenges of sentiment detection in the social programmer ecosystem." Proceedings of the 7th International Workshop on Social Software Engineering. ACM, 2015.
[62] Jotheeswaran, Jeevanandam, and S. Koteeswaran. "Feature selection using random forest method for sentiment analysis." Indian Journal of Science and Technology 9.3 (2016).
[63] Uysal, Elif, et al. "Sentiment Analysis for the Social Media: A Case Study for Turkish General Elections." Proceedings of the SouthEast Conference. ACM, 2017.
[64] Buettner, Ricardo. "Predicting user behavior in electronic markets based on personality-mining in large online social networks." Electronic Markets 27.3 (2017): 247-265.
[65] Yarkoni, Tal, and Jacob Westfall. "Choosing prediction over explanation in psychology: Lessons from machine learning." Perspectives on Psychological Science 12.6 (2017): 1100-1122.
[66] Yadollahi, Ali, Ameneh Gholipour Shahraki, and Osmar R. Zaiane. "Current state of text sentiment analysis from opinion to emotion mining." ACM Computing Surveys (CSUR) 50.2 (2017): 25.
[67] Xue, Di, et al. "Personality recognition on social media with label distribution learning." IEEE Access 5 (2017): 13478-13488.

[68] Buettner, Ricardo. "Predicting user behavior in electronic markets based on personality-mining in large online social networks." Electronic Markets 27.3 (2017): 247-265.
[69] Uysal, Elif, et al. "Sentiment Analysis for the Social Media: A Case Study for Turkish General Elections." Proceedings of the SouthEast Conference. ACM, 2017.
[70] Imran, Azhar, Muhammad Faiyaz, and Faheem Akhtar. "An Enhanced Approach for Quantitative Prediction of Personality in Facebook Posts." International Journal of Education and Management Engineering (IJEME) 8.2 (2018): 8-19.
[71] Azucar, Danny, Davide Marengo, and Michele Settanni. "Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis." Personality and Individual Differences 124 (2018): 150-159.
[72] Bleidorn, Wiebke, and Christopher James Hopwood. "Using machine learning to advance personality assessment and theory." Personality and Social Psychology Review (2018): 1088868318772990.
[73] W. Shaikh, A. Yusuf, and S.-R. Massan, “Data Analytics & Classification Through Advanced Machine Learning Methods,” in SZABIST Research Conference, 2015, p. 1.

Mr. Ameen received his MS degree from Mohammad Ali Jinnah University Karachi in Computer Science in 2019. He is working as Lecturer at Sindh Madressatul Islam University, Karachi.He has more than 15 years of teaching and administrative experience. His research interest areas are: Sentiment Analysis, Internet of Things, Databases and programming languages.

Dr. Mumtaz Hussain Mahar has completed his Ph.D. from England. Currently he is working as Dean, Faculty of Physical Sciences, Shah Abdul Latif University, Khairpur.

Dr. Shaukat Wasi received his PhD (CS) and MS (CS) degrees from FAST-National University of Computer and Emerging Sciences (NUCES) in 2015 and 2006 respectively. He did his Bachelors in Computer Science from University of Karachi in 2003. He started his teaching career at FAST-NUCES in 2007. He, then joined DHA Suffa University (DSU) as the founding teaching faculty. Currently he is engaged at Mohammad Ali Jinnah University (MAJU) as Associate Professor and the leading the Department of Computer Science. Dr. Shaukat wasi has been working and interested to do research in HCI, IR, IE, Text Classification and Mining, and Email Research.

Mr. Haque Nawaz Lashari is persuing his PhD in Computer Science from Shaheed Zulfikar Ali Bhutto Institue of Science and Technology, Karachi, He received his MS degree from Mohammad Ali Jinnah University Karachi from Network and Telecommunication in in 2010. He is working as Lecturer at Sindh Madressatul Islam University, Karachi. He has more than 23 research publications in national and international journals. His areas of research interests are wireless communication, network security, routing protocols, optimization algorithms and mobility management in mobile ad hoc networks

Mr. Shafiq-ur-Rehman Massan is persuing his PhD in Computer Science from Shaheed Zulfikar Ali Bhutto Institue of Science and Technology, Karachi. His areas of research interests are Artificial Intelligence, Energy and Control systems. Metaheuristic Algorithms and optimization techniques, Image Processing, Data Mining, Machine Learning, Neural Networks and Deep Learning.