IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 18, Issue 3, Ver. II (May-Jun. 2016), PP 44-52
www.iosrjournals.org

Human Being Character Analysis from Their Social Networking Profiles

Biswaranjan Samal[1],Mrutyunjaya Panda[2]
Department Of Computer Science And Application, Utkal University,Vanivihar/ Bhubaneswar, India
Department Of Computer Science And Applications, Utkal University,Vani Vihar, / Bhubaneswar, India

Abstract : *In this paper, characteristics of human beings obtained from profile statement present in their social networking profile status are analyzed in terms of introvert, extrovert or ambivert. Recently, Machine learning plays a vital role in classifying the human characteristics. The user profile status is collected from LinkedIn, a popular professional social networking application. Oauth2.0 protocol is used for login into the LinkedIn and web scrapping using JavaScript is used for information extraction of the registered users. Then, Word Net: a lexical database is used for forming the word clusters such as: extrovert and introvert using semi-supervised learning techniques. K-nearest neighbor classification algorithm is finally considered for classifying the profiles into various available categories. The results obtained in the proposed method are encouraging with good accuracy.*

Keywords*: accuracy ,Clustur, Corpus, K-NN, Oauth, Social media, Unigrams, WebScrapping ,Word Net*

I. Introduction

Human beings are known as social animals who love to share their feelings through various ways of communications. The communications might be in the form of face to face or in the form of written or by using the electronic medias or over internet. In current scenario social networking applications are playing vital roles for communication of the peoples. People use these applications for building their networks wider by adding new peoples to their profiles or by joining various available groups and also make updates about themselves, by sharing their moods, current status, photos. User tries to make their profile to look smarter by adding their personal information as well as writing summary about their characteristics, which results them to stay more updated and connected with their nearest and dearest one. These applications help people to meet new peoples using their various facilities, so it is much more important for a user to make his/her profile much more attractive and strong.

These social networking applications can be categorized according to their uses, such as: Facebook for connecting peoples and share information or interact with the peoples and LinkedIn for professionals to connect with various professional groups or forums. As a result "social recruiting" is also becoming a popular word along with these social applications [10] that allows us to understand how the web technologies and social networking applications are used by the peoples and organizations for recruiting the talented resources according to their need, saving lots of time, money and resources. Because of this, lots of people are registering themselves with this type of professional websites and tries to show case their personality , talent, knowledge, by providing various types of information's which may or may not be accessible publicly depending upon their own preferences. Likewise various organizations and human resource (HR) groups also used to get registered in the same professional social networking websites for social recruiting purpose and they visit over the user profiles and use searching facilities of these websites to reach the targeted users or profiles [10]. That's why it is necessary for the users of these professional social networking applications to build their profile in smart way which will impress the recruiters.

In our approach character of a human being can be thought of as: (1) Extrovert – A person whose character is interest turns outward, exposed towards external world, strong desire for activity, interested in athletics, dominant, not easily embarrassed, aggressive, unscrupulous, popular with people , egoistic , speaks fluently, and behaves friendly (2) Introverts includes characters such as lack of activity , book and magazines lover , interest in inner life, self-centered, submissive attitude , scrupulous , inclined to worry , not popular , good at writing than speaking ,and Ambivert are the mixture of both characters , it may behave as introvert in one situation and as extrovert in another situation.

Machine Learning (ML) is a field of Computer Science which can be defined as "It includes the area of study in which machines or computers are able to learn without being programmed explicitly "[11]. Based on the desired outcome of algorithms, common and popular algorithm types are

- Supervised Learning – This ML is used for classification purpose. Where learner is required to learn a function which will map a vector into a class from available classes by considering the functions input and output values.

- Unsupervised Learning – Here set of inputs are not available and algorithms set a model.
- Semi-supervised Model – This ML is used with the algorithm where a set of labeled and unlabeled inputs are used for generating appropriate classifier or model. And this model best suits for our research work to train the classifier model for classifying.

Natural Language Processing (NLP) is a field of Computer Science, which combines the artificial intelligence and computational linguistics which is responsible for interaction between the computers and human languages. Most of the algorithms that are used by modern NLP are based on ML. So NLP has vast uses such as in Optical Character Recognition (Given an image to the computer, and it has to recognize the characters present in it), Natural Language Generation (Here database is given to the computer, and it has to convert it into the human readable language) and many more.

Now-a-days, sentiment analysis [12] and opinion mining [13] have attracted many a researchers to carry out research using NLP and ML, which seems to be a challenging field. This motivated us to pursue further investigation in this direction to validate the proposed approach in human being character analysis.

The objective of this paper is centered at the following:

- To identify the users characteristics who are registered with LinkedIn, by considering their profile statement available in LinkedIn and categorizing them as Extrovert, Introvert or Ambivert. This is mainly because of the users generally intends to keep a positive profile with a large number of positive words and hesitant to include their shortcomings in the personal profile. Some time it happens that user thinks that, as the profiles may be seen in public, so most of the users tries to add so many positive things about them and hides the negative things about them in their profile statements, for making their profile to look positive. Which may result in a negative way and their profile statement may looks more extrovert type. Where as it is observed that people having ambivert characteristics which includes characteristics from both introvert and extrovert are more liked by other peoples [9].
- This research will tries to analyze user profiles, and tries to categorize the profile type. So that it can suggest the users about their profile type and guide them to write a good profile statement which may include characteristics from both introvert and extrovert and makes the profile as ambivert.

Rest of the paper is organized as follows. Section II discusses some of the related work available in the literature. Section III describes about the classification technique used in our research. Section IV addresses the detailed methodologies adopted in a step by step manner. Experimental results and discussions are provided in Section V followed by conclusions and future scope of research in Section VI.

II. Related Works

In [6], it studied the relationship between language used on Twitter and personality traits. This paper shows how various linguistic features correlate with each personality trait and to what extent personality traits can be predicted from language. It gathered personality data from Myers-Briggs Type Indicator (MBTI) personality test which contains thinking, feeling, sensation, intuition, introversion, extroversion, judging and perceiving of the users. It has used n-grams, Twitter POS tags, and word vectors to explore the most related linguistic features for different personality traits. It able to predict the personality traits with an average accuracy of 66.1%.

In [7], authors have predicted the personality by collecting data from standard mobile phone logs, Simple Vector Machine (SVM) classifier has been used as the model, resulting in a prediction whether phone users were low, average, or high in neuroticism, extraversion, conscientiousness, agreeableness, and openness with an accuracy of 54%, 61%, 51%, 51%, and 49%, respectively.

In [9], author’s purpose was to determine whether there are significant gender-based differences in academic achievement, test anxiety and personality type (introversion, extroversion) among high school students in Papumpare district of Arunachal Pradesh, India. It has used mean, standard deviation and t-test for analysis of data.

In [21] authors have taken Facebook messages of 75,000 volunteers and found variants in language with personality, age and gender. They have considered general language only ignoring the part of the online behaviours.

It is also reported by several researchers that used patterns from the usage of social media such as Facebook and Twitter for the user’s personality prediction [22, 23].

The authors have gathered information about the usage of mobile phone applications from YouTube, Internet, etc ., and provided the meaningful insights of the mobile phone user and their personality trait [24, 25].

III. K-Nearest Neighbor Classification Algorithm

Data can be analysis either by using Classification or by using the Prediction models. Detailed descriptions about these models are given below.

Prediction models are used to predict continuous-valued functions.

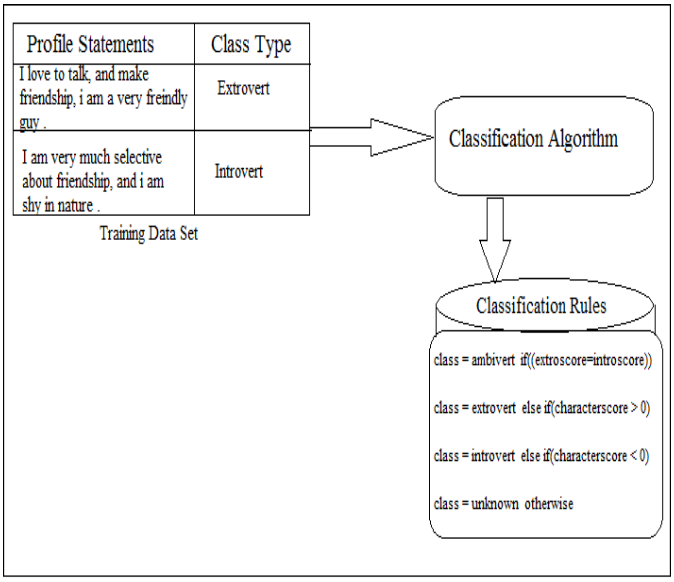
For example let's assume that a teacher wants to calculate how much time is taken by a student for answering particular questions answer. This is a numeric prediction.

Classification models are used for predicting the category class labels.

For example classifying a text message spam or not or classifying characteristic of a human being extrovert or introvert.

From above example we can observe that the model which we want to use for our data analysis purpose is best suit to Classification model, so we are going to adopt the classification model for doing data analysis on human being character analysis. The data classification includes below two steps

Step 1- Building the model or Classifier, this is known as the learning phase in which we have to use the classification algorithm to build the classifier. The classifier is built using the training set, in our research the training set contains the profile statements along with their type classification types. Which will work as sample data to train the model? It can be observed from Fig 1.

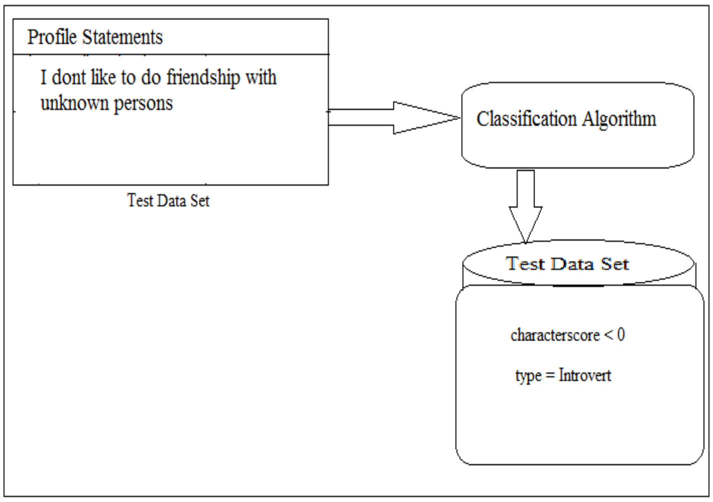


Profile Statements	Class Type
I love to talk, and make friendship, i am a very freindly guy .	Extrovert
I am very much selective about friendship, and i am shy in nature .	Introvert

Training Data Set
Classification Algorithm
Classification Rules
class = ambivert if((extroscore=introscore))
class = extrovert else if(characterscore > 0)
class = introvert else if(characterscore < 0)
class = unknown otherwise

Figure 1 . Demonstrating Model/Classifier Building

Step 2 – Use Classifier for classification , here the classifier is used for classifying the test set or the new data set by applying the classification rules that we have trained the classifier in our Step 1. Fig 2 will demonstrate this step.



Profile Statements
I dont like to do friendship with unknown persons

Figure 2. Demonstrating the use of classifier for classification of new data

The available algorithms for classification are K-NN, Naive Bayes, Support Vector Machine(SVM) and Decision Tree [19].

Decision Tree, these are the trees which classifies instances by sorting them based on their feature values, it follows divide and conquer paradigm [19].

Naive Bayes, this classification algorithm is used to train the classifier with a set of objects which are belongs to set of classes and aim of this algorithm is to construct a rule which will further classify the new objects. This is supervised learning model [19].

Support Vector Machine (SVM), it is known because of it s high generalization performance without the need of priori knowledge [19].

K-Nearest Neighbor, This algorithm is otherwise known as Memory-Based Reasoning, Example-Based Reasoning, and Lazy learning. The basic idea behind using the K-NN algorithm is that, this algorithm can be used with a dataset in which the data are separated into several classification groups or classes for predicting the class of a new sample or given data point. It can be used as a semi supervised learning model.

In K-NN algorithm the initial dataset is known as training dataset which is used by the algorithm for training itself. And the data which we will give for testing is known as the testing dataset, here the way in which the algorithm decides the new point to be classified into an existing class by observing the training set are to pick the k closest data points to the new observation and take the most common class among those. In other words we can simply say that this algorithm will observe or retrieve the neighbors and accordingly do's the classification.

So, here we can say that larger the size of the available training dataset more is the accuracy of classification. We can observe that k =1, which says to select the nearest neighbor is used for efficiency but again it depends upon the "noise" present in the dataset. If locality is preserved then larger "k" can give a smoother boundary which is best for the generalization. The pseudo code for the K-NN algorithm is provided in Table 1.

Table 1: K-NN Algorithm

```
K – Represents the number of nearest neighbor to consider,

P – Represents the test_data.

Q – Represents the training_data

For each object obj in P do

•   Calculate the distance D(obj,y), between "obj" and every object y in Q
•   neighborhood = the k neighbors in the Q closest to obj
•   obj.class = setClass (neighborhood)
end For
```

Fig. 3 demonstrates the training_set with red, yellow, and green color and the test_set with black color which needs to be classified. Euclidean distance is used to measure the similarity between the words those falls in different classes.

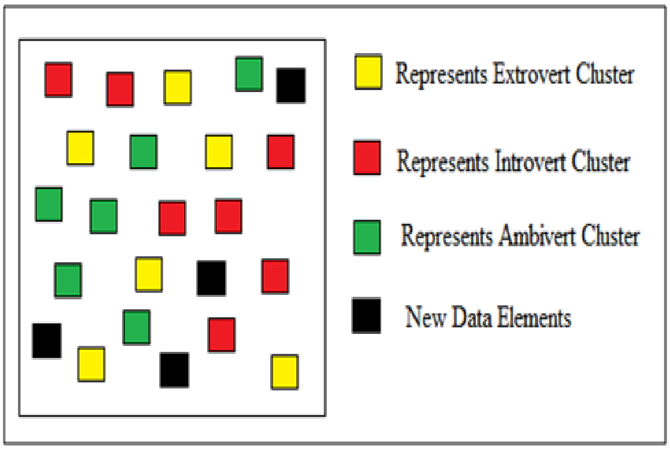

Represents Extrovert Cluster
Represents Introvert Cluster
Represents Ambivert Cluster
New Data Elements

Figure 3 . Demonstrating K-NN algorithm

IV. Proposed Methodology

The step by step procedure for our experimental set up is as follows:

A. Collecting Profile Data from LinkedIn

Extracting useful information's from the web or websites is known as "Web Mining". It is the most tedious work, while one wants to collect a large number of data. This can be achieved by some of useful web technologies known as Web Usage Mining, Web Scrapping and Semantic Annotations [14].

- Web Usage Mining – This web mining is used to get the information's about the visitors or users of a website [14].
- Web Scrapping – Popularly known as "Web Data Extraction" is a technique for retrieving useful data from the websites by using a scripting language supported by the web technologies [14].
- Semantic Annotations: This is used to structure the unstructured data present over the web, for this we need external tools [14].

Oauth 2.0 is an open protocol which is used for authentication of the applications, so that applications can interact with each other as well as access each other. According to this protocol a third party application is allowed to access some of the basic information's of an user from a web application on behalf of the user by providing an interface to the user for log in to the system from which the third party needs information [15].

JavaScript is a client side as well as server side scripting language which is a widely used and most popular scripting language used by almost in every web applications. Either in client side processing or in server side processing.

In our first step we have used Oauth2.0 authentication techniques to sign in into the LinkedIn and then adopted the web scrapping technique, for which we have used JavaScript as our scripting language, for web mining and collected approximately 3000 users' summary data, and stored them into a text file.

B. Forming Word Clusters

Cluster - A group of similar things occurring closely together or have common characteristics. The process in which things are grouped is known as clustering. It is a technique for statistical data analysis. Clustering can be done by using the ML techniques such as Supervised, Semi Supervised, and Unsupervised. As the name suggest in supervised clustering user interaction is required throughout the clustering process. But in semi supervised learning initially the user is required to train the algorithm on how to do clustering with some examples and then the algorithm do its work and build the cluster according to the given training. Where as in unsupervised learning the algorithm is designed to learn from a set of input and output and accordingly it works.

Lexical Database is a collection of databases. Lexical database contains information about lexical category, synonyms of words as well as it also contains the semantic and phonological relations between words.

WordNet [3][4] : It is a freely available lexical database consisting of English nouns, verbs, adjectives, adverbs which are grouped together into set of synonyms (synsets), each expressing a distinct concept. This database can be used in online and also can be used in offline [18]. Its design is inspired by psycholinguistic theories of human lexical memory [18].We have used offline version of WordNet as our word dictionary for collecting various synsets and similar words building the extrovert and introvert word cluster.

Now it's turn for building up the word cluster for the introvert and extrovert characteristics [4]. Here instead of following traditional way of collecting all the similar kind of words for both the classes we have adopted a semi supervised machine learning approach [1]. In this approach we don't need to collect manually all the related words for the classes which is a lengthy and tedious task, instead we initially collect some of the related words for each class and form the initial word cluster [2] .

Next we trained our system to do the machine learning and retrieve all the synonyms and similar words for the words present in different cluster, from WordNet [3][4] library and merge them into the corresponding word clusters.

We repeat this process recursively and finally formed the word cluster's having words more than 1500 in each cluster. And stored the cluster data into text files.

C. Corpus Building and Corpus Cleaning

Linguistic is the scientific study of language. Linguistic can be used for relating human sounds with their meaning, or it can be used to encode the relationship between the related words present inside a text block and so on. Here we have used this for building the corpuses from the clusters.

Corpus: In linguistic corpus is the plural form of the corpora [16], it can be of two types Text Corpus and Speech Corpus, and again these types can be sub divided as monolingual corpus or multilingual corpus [16].

- Speech Corpus - It represents a large set of audio files.
- Text Corpus – It represents a large set of structured texts.
- Monolingual Corpus – This is the type of the corpus whose contents are consisting of only one language type.

- Multilingual Corpus – This is the type of corpus where contents of the corpus can be of more than one language type.

In our research we are going to use text and monolingual corpus as all the cluster we have formed previously are the text clusters, consisting of only a single language which is English. The languages present other than English should be removed.

Word cloud: An image composed of the words used in a particular text or subject, where the size of the words present in the image indicates the frequency of that word.

The cluster that we have developed contains numbers, special characters, unrecognized symbols, emoticons, white spaces, unnecessary spaces, icons. So we have to clean the clusters and reform the clusters according to our need so that ambiguity will be less and the processing of the texts will be smoother. During this step we have also taken one more step and have removed all the English stop words as well as the words having length less than 3 as we found that the words having length less than 3 are unused for our work.

While developing the profile corpus which contains all the summary of LinkedIn users we have put a special string "xxxtxxx" as a separator between the profiles, for easy recognition and build the profile corpus. We have also build the Word Cloud for the profile corpus by taking few of words from them. "Fig. 3" shows the word cloud for the profile summary corpus.

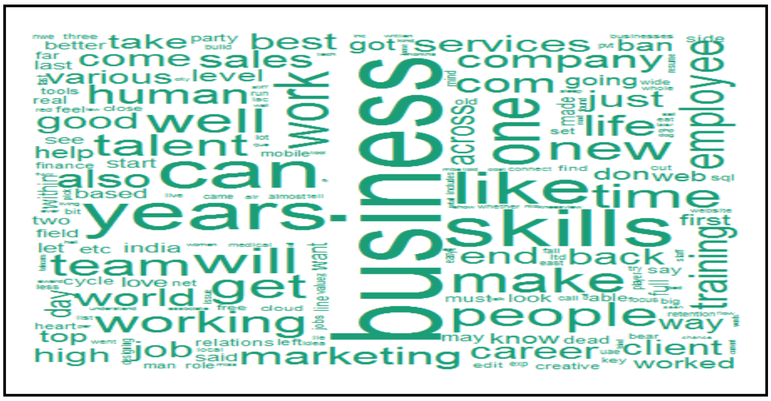


Figure 4. Profile Word Cloud

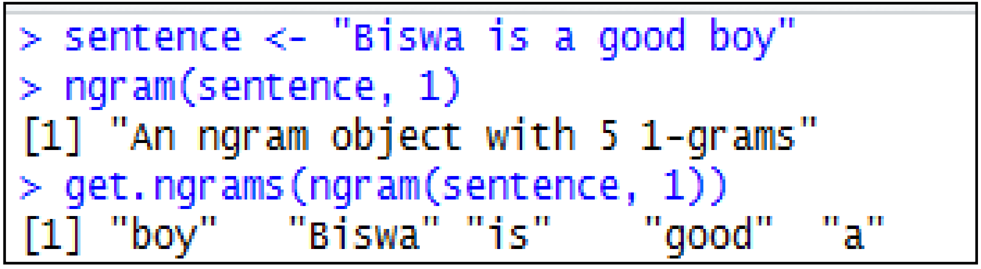
In Fig .4, the words that are look more bolder and larger indicates that these words are being frequently used by the peoples in their profile statements and the words which are small in size indicates that they have used in less frequently. Here in this word cloud I have placed the words whose frequency of occurrence is at least "10".

D. Identifying Unigrams

Known as N-gram of size 1. In NLP and linguistic world N-gram refers to the continuous sequence of n items in a given sequence of text or speech. N-grams are generally collected from text or speech corpus [17]. Where N= 1 the N-grams are known as Unigrams, likewise "bigrams" for N = 2, "trigrams" for N = 3 and so on. In our research it is necessary to retrieve all the unigrams present in a profile statement for measuring which words belongs to which cluster type, introvert or extrovert. And accordingly we will assign the classification to the profile statements.

Example:

R is a programming language and software environment used for various statistical computing and data mining works . We have used this language for our research work, because our work best suits the purpose for which the "R" language has been developed.



```
> sentence <- "Biswa is a good boy"
> ngram(sentence, 1)
[1] "An ngram object with 5 1-grams"
> get.ngrams(ngram(sentence, 1))
[1] "boy"    "Biswa" "is"    "good"   "a"
```

Figure 5. R Code sample of retrieving unigrams from a sentence

Fig .5 , Demonstrates a "R" code sample for generating available unigrams from a given sentence.

Here in this figure we are considering a sentence “ Biswa is a good boy “ in first line of the image and the last line in the image shows the generated unigrams from the entered sentence.

We will separate each statement by recognizing the special character we have added and then retrieve all the unigrams present in that sentence and compare each unigram for identifying, weather the unigrams belongs to extrovert type or introvert type. Here we will maintain 3 counter variables, such as extroscore, introscore, characterscore. If we will find the unigram belongs to the extrovert cluster then we will increase the extroscore by “+1” and characterscore by “+1”, where as if the unigram belongs to the introvert cluster then we will increase the introscore by “+1” and characterscore by “-1”.

And finally when we finished retrieving all the unigrams from a sentence then we will calculate the values of introscore, extroscore, characterscore and accordingly decides the class in which the sentence or the profile of a user falls (Extrovert, Ambivert, Introvert).

```
class = ambivert  if((extroscore == introscore))
class = extrovert  else if(characterscore > 0)
class = introvert  else if(characterscore < 0)
class = unknown  otherwise
```

And we will combine the sentence with the type and form a new sentence which will contains the sentence along with its type, which can be used as input in our further steps.

E. Building and Training the Classification Model

As discussed earlier instead of using traditional supervised learning algorithm for training the model and manually assigning levels to the training set, which is a time consuming and tedious task [2]. We follow the semi supervised learning approach in which we will train the model initially by providing a training data set, which contains both the sentence and its classifications type. And expects that the model will train itself using the training set given to it and for developing the model we are going to use K-Nearest Neighbor’s classification algorithm.

Here we have use K-Nearest Neighbor classification algorithm for training our model.

The algorithm overflow used in this paper is presented in Table 2.

Table 2: Algorithm overflow of the proposed approach

Step 1: Build extrovert, introvert, profile cluster.

Step 2: Build corpus from the clusters.

Step 3: Clean corpus (remove numbers, stop words, punctuations, special characters, white spaces, etc), and add a string “xxxtxxx” at the end of each profile statement.

Step 4: Retrieve the unigrams and identify the cluster they belongs to and classify them accordingly.

Step 5: Bind the class (extrovert, introvert, ambivert) with each profile sentence.

Step 6: Build data frame from the profile corpus.

Step 7: Build TermDocumentMatrix (profiletdm) from the above data frame.

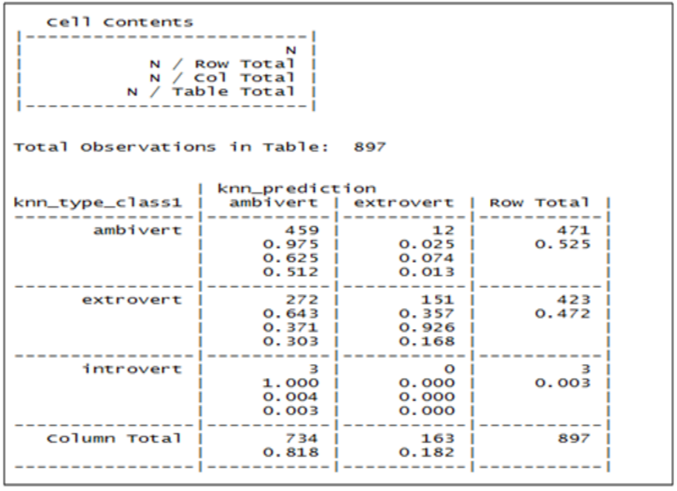
Step 8: Build the model with K-NN classification algorithm.

Step 9: Train the model by providing 70% data of the profiletdm.

V. Experimental Results And Discussion

All the experiments are conducted in an Intel I3 processor with 2.00 GHz CPU with 1TB HDD with 4GB RAM in Windows 10 Operating System. We have used Oauth 2.0 for login to LinkedIn, Web Scrapping using JavaScript for extracting information from LinkedIn, WorldNet for word extraction and R as programming language for carrying out our proposed research.

The observations and confusion matrix obtained after simulation is provided in Fig. 6 and Fig. 7 respectively.

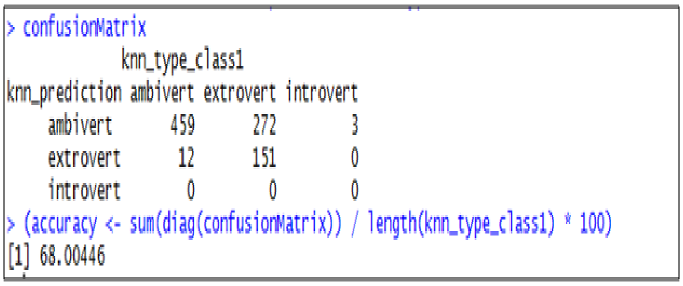


```
   Cell Contents
|-------------------------|
|                       N |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|


Total Observations in Table:  897


               | knn_prediction
knn_type_class1 |  ambivert | extrovert | Row Total |
---------------|-----------|-----------|-----------|
      ambivert |       459 |        12 |       471 |
               |     0.975 |     0.025 |     0.525 |
               |     0.625 |     0.074 |           |
               |     0.512 |     0.013 |           |
---------------|-----------|-----------|-----------|
     extrovert |       272 |       151 |       423 |
               |     0.643 |     0.357 |     0.472 |
               |     0.371 |     0.926 |           |
               |     0.303 |     0.168 |           |
---------------|-----------|-----------|-----------|
     introvert |         3 |         0 |         3 |
               |     1.000 |     0.000 |     0.003 |
               |     0.004 |     0.000 |           |
               |     0.003 |     0.000 |           |
---------------|-----------|-----------|-----------|
  Column Total |       734 |       163 |       897 |
               |     0.818 |     0.182 |           |
---------------|-----------|-----------|-----------|
```

Figure 6 Experimental Observations



```
> confusionMatrix
              knn_type_class1
knn_prediction ambivert extrovert introvert
     ambivert       459       272         3
     extrovert       12       151         0
     introvert        0         0         0
> (accuracy <- sum(diag(confusionMatrix)) / length(knn_type_class1) * 100)
[1] 68.00446
```

Figure 7 Confusion Matrix for Accuracy

It can be observed from Fig 6 and Fig 7 that the overall accuracy of our proposed approach is 68% which is quite encouraging.

VI. Conclusions And Future Works

We present a novel approach on how to classify characteristics of the human beings considering their LinkedIn profile status or profile summary using K-NN classification technique. This may provide an impetus to the social network users for improving their profiles in a smarter way.

The results obtained are satisfactory that encourage us to extend our work to make more accurate prediction by varying the size of the N-Gram and other efficient machine learning techniques in future.

References

[1]. Bing Liu, "Synthetic structure of industrial plastics (Handbook of Natural Language Processing, Second Edition, editors: N. Indurkhya and F. J. Damerau)," 2010.

[2]. Bing Liu, Xiaoli Li, Wee Sun Lee and Philip S. Yu, "Text Classification by Labeling Words", American Association for Artificial Intelligence. 2004.

[3]. George A. Miller, "WordNet: (A Lexical Database for English. Communications of the ACM Vol. 38, No. 11: 39-41)", 1995 .

[4]. Christiane Fellbaum, "(WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press)", 1998.

[5]. Dekang Lin, "Automatic Retrieval and Clustering of Similar Words" ., COLING '98 Proceedings of the 17th international conference on Computational linguistics - Volume 2 , 768-774, 1998

[6]. YilunWang,"UnderstandingPersonalitythroughSocialMedia".YilunWang,"UnderstandingPersonalitythroughSocialMedia". Accesed on 30/04/2016.

[7]. Yves-Alexandre de Montjoye,, Jordi Quoidbach, Florent Robic, and Alex (Sandy) Pentland, "Predicting Personality Using Novel Mobile Phone-Based Metrics" in: A.M. Greenberg, W.G. Kennedy, and N.D. Bos (Eds.): SBP 2013, LNCS 7812, pp. 48–55, 2013.

[8]. Hassan Seif, "Naïve Bayes and J48 Classification Algorithms on Swahili Tweets: Perfomance Evaluation((IJCSIS) International Journal of Computer Science and Information Security, Vol. 14, No. 1,)", January 2016 .

[9]. Dr.B.Reena Tok, SubhanginiBoruwa, A Study On Gender-Based Differences In Relation To Test Anxiety, Academic Achievement And Personality Type Among High School Students In North East India- With Reference To Papumpare District Of Arunachal Pradesh .Excellence International Journal Of Education And Research Volume 2, Issue 5 ,Issn 2322-0147, May 2014

[10]. Vicknair, Jamie; Elkersh, Dalia; Yancey, Katie; Budden, Michael C. (The Use Of Social Networking Websites As A Recruiting Tool For Employers) , American Journal of Business Education 3.11 (Nov 2010): 7-12.

[11]. Taiwo Oladipupo Ayodele . Types of Machine Learning Algorithms, New Advances in Machine Learning, Yagang Zhang (Ed.), InTech,2010, DOI: 10.5772/9385. Available from: http://www.intechopen.com/books/new-advances-in-machine-learning/types-of-machine-learning-algorithms

[12]. Bing Liu, "Sentiment Analysis and Subjectivity" Handbook of Natural Language Processing, Second Edition, (editors: N. Indurkhya and F. J. Damerau), 2010 .

[13]. Bing Liu. Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, May 2012.

[14]. Sanjay Kumar Malik ; , SAM Rizvi, Information Extraction Using Web Usage Mining, Web Scrapping and Semantic Annotation " International Conference on Computational Intelligence and Communication Networks (CICN), 465 – 469,2011 .

[15]. D. Hardt, Ed, The OAuth 2.0 Authorization Framework, Category: Standards Track,ISSN: 2070-1721, October 2012.

[16]. Marianne Hundt, Nandja Nesselhauf,Carolin Biewer, " Corpus linguistics and the web", May 2006.

[17]. William B. Cavnar and John M. Trenkle , "N-Gram-Based Text Categorization", 1994 .

[18]. Alberto J. Caæas, Alejandro Valerio, Juan Lalinde-Pulido, Marco Carvalho, Marco Arguedas, "Using WordNet for Word Sense Disambiguation to Support Concept Map Construction" 10th International Symposium, SPIRE 2003, Manaus, Brazil, October 8-10, 2003. Proceedings.

[19]. Raj Kumar, Dr. Rajesh Verma, Classification Algorithms for Data Mining: A Survey, International Journal of Innovations in Engineering and Technology (IJIET) Vol. 1 Issue , August 2012 7 ISSN: 2319 – 1058 .

[20]. Pang-Ning Tan, Michael Steinbach and Vipin Kumar. Introduction to Data Mining, Addison Wesley, 2006 .

[21]. HAndrewSchwartz,JohannesCEichstaedt,MargaretLKern,LukaszDziurzynski,StephanieMRamones,MeghaAgrawal,AchalShah,MichalKosinski, David Stillwell, Martin EP Seligman, et al. 2013. Personality, gender, and age in the language of social media: The Back, M.D., et al.: Facebook profiles reflect actual personality, not self-idealization. Psychological Science 21(3), 372–374 (2010) .

[22]. Counts, S., Stecher, K.: Self-presentation of personality during online profile creation. In: Proc. AAAI Conf. on Weblogs and Social Media (ICWSM) (2009).

[23]. Chittaranjan, G., Blom, J., Gatica-Perez, D.: Mining large-scale smartphone data for personality studies. In: Personal and Ubiquitous Computing (2012).

[24]. Staiano, J., et al.: Friends dont Lie–Inferring Personality Traits from Social Network Structure (2012).