A Machine Learning-based Online Social Network Analysis for 360-degree User Profiling

R.Kanniga Devi

bstract: *This paper aims to analyse the online social network for reconnaissance of people for finding their potentiality. The work considers one of the famous social networking sites, Twitter, where people express their thoughts and ideas, through which the people in the site knowingly or unknowingly reveal the information about themselves such as personal interests, likes and dislikes. The Machine Learning technique facilitates the work to mine the tweet data of a person to get his/her 360-degree profiling. This profiling is helpful to identify the personality type of a person, which is essential for the Government to identify the people involved in spreading the fake news in Twitter.*

Keywords : Machine Learning, Natural Language Processing, Online Social Network, Personality Test, Profiling, Sentimental Analysis, Twitter.

I. INTRODUCTION

Socially relevant set of nodes are connected with some relation to represent friendships or communication in a social network [1]. The social networking sites facilitate the real-time communication and interaction among its users [2]. The zero cost for the social networking site usage attracts many people to share their day-to-day experiences and opinions on boundless topics such as product usage, movie review and current affairs [3]. By holding an enormous amount of shared information, the social networking sites become a rich source of information [4]. The rich source of information catches the attention of many business organizations, who tend to transform the offline marketing strategies into online marketing in order to brand their products. To survive and to be successful in the competitive business market, business organizations strive to understand the mindset of the people. Twitter is a famous social networking micro blogging site and potential market for the businesses [5]. The networking model in Twitter allows people to follow other people without any permission [6]. The community in Twitter varies from the common people to the

celebrities. From the high-level interaction in Twitter community, the business organizations need to identify the influential users to brand their content. The Social Network Analysis (SNA) is the branch of sociology, which studies the structure of the social network and relationship among the social system using mathematics to extract some interesting patterns out of the analysis [7]. Many researchers performed SNA on various social networking sites to identify the influential persons for target marketing, but failed to make 360-degree profiling of the users [8-12]. Hence, this work

aims to mine the Twitter data to get 360-degree profiling of users, which may help the business organizations to identify the influential persons and increase the efficiency of marketing campaigns. For analyzing the Twitter data, this work adopts Machine Learning technique.

Machine learning (ML) is a computer science field that facilitates computers to learn and write its own code to perform a task that the user wants to achieve [13]. For instance, online recommendation engines use ML to personalize the ad delivery. ML can also be used to filter spam and detection of network security threats [14]. There exists two categories of Machine learning algorithms: supervised and unsupervised [15], [16]. The supervised machine learning algorithm takes input data with data labels and determines a rule to map inputs to outputs. The unsupervised machine learning algorithm takes input data without data labels and determines its own rule to map inputs to outputs.

Regression analysis is a kind of Machine Learning method, which is used for predictive modelling and has variants of analytical methods [17]. In particular, this work focuses on the Regression analysis method, Logistic Regression. The Logistic Regression multiplies the input data with a weight value to determine the most important input features to discriminate among the different classes [18]. In this work, Supervised Machine Learning Classification algorithm named Logistic Regression is user to analyze the Twitter. Logistic regression provide greater accuracy as it is a binary classifier over other multiclass classification algorithms/classifiers namely Random Forest, Decision tree, *K*-nearest neighbour algorithms. Since this work is based on multiclass classification, but the Logistic Regression classifier is a binary classifier, this work uses a principle named oneVsRest (OVR) to classify the multiclass classification problem.

The natural language is the language used by people for every day communication. NLP applies computational techniques to analyze and synthesize the natural language [19]. In this research, the text is the tweet posted by the person. NLP involves analysis of text to gather information from it. This work processes the raw-tweet into information rich processed tweet using NLP.

The opinion of the mass is important in many cases. For example, the business organizations may want to know whether people support their product or not. A company might want to find out the reviews or sentiments of its products. Sentiment Analysis (SA) is a kind of computational technique in Natural Language Processing. It helps in finding the sentiment or opinion hidden within a text.

R.Kanniga Devi*, Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Krishnankoil, India. Email: rkannigadevi@gmail.com

A Machine Learning-based Online Social Network Analysis for 360-degree User Profiling

Analyzing the Twitter data is more challenging due to its unstructured nature, limited size, use of slangs, misspells and abbreviations. This research focuses on finding sentiments from Twitter data. Most of the researchers dealt with various ML approaches of sentiment analysis [20-25], and just classified sentiments as positive and negative sentiments. Hence, this research work addresses the likes, dislikes and character of the persons in addition to regular sentiment analysis, which will be helpful to know the complete personality of a person.

Behaviour analysis is the science of controlling and predicting human behaviour. A behaviour is everything a person does. Analyzing human behavior has impact on social influence. Investigation of user behaviour analysis is an important research area in computational social network for various reasons. Identification of criminal activities, anomaly detection, information diffusion in short period, sentiment analysis of social groups, popularity estimation, product rating belong to behaviour analysis in social networks. Most of the social networks are dynamic in nature, this poses tremendous challenges in characterization of exact behaviour. The studies on social network user behaviour needs more investigations by research communities to address various challenges in accurate analysis of behaviour. There exists a large number of social network user behaviour analysis methods with special emphasis on behaviour prediction [26-30]. By analyzing various social network data sets and by observing past activities, future behaviour of user can be predicted.

Although many ML techniques were used for profiling, if we adopt a hybrid approach which involves a sentiment analyzer and a personality test, the profiling can be strengthened further. In addition, this paper presents the result obtained from various machine-learning classifiers. In particular, sentiment change of the public opinions can be predicted very well in order to lessen the negative sentiments and its spread.

The paper is structured as follows: The section 2 presents the detailed literature survey. The section 3 elaborates the proposed work. The section 4 presents the comparative study. Finally, section 5 provides concluding remarks.

II. LITERATURE SURVEY

This section reviews the sentiment analysis, behaviour analysis and user behaviour prediction on social networks with special emphasis on Twitter.

A. Sentiment analysis

Maite Taboada et al. [20] presented a lexicon-based approach for extracting sentiment from text. This work developed a Semantic orientation calculator, which used dictionary of words with their semantics. This calculator was applied to polarity classification task in order to assign pessimistic or optimistic opinion label to a data. The work outperformed the compared approaches namely, manual or automatic dictionaries. In addition, this solution performed well on blog postings and video game reviews. Le T. Nguyen et al. [21] built a machine learning model for the social media dynamics to predict collective sentiment dynamics. This

Support Vector Machine (SVM) model predicted the sentiment changes of a topic over a period. This model achieved above 85% accuracy on directional sentiment prediction on large-scale Twitter data. However, this work did not delve into micro analysis of individual users or tweets and the low level network structures. Michele Di Capua et al. [22] proposed an unsupervised approach for automatic detection of torment traces over a social network, using NLP (Natural Language Processing) and machine learning. This work designed a model based on Growing Hierarchical Self Organizing Maps (SOM), which clustered documents containing torment trace. This model worked well with the social networks Twitter, YouTube and FormSpring to detect bully traces. Geetika Gautam et al. [23] performed sentiment analysis on Twitter data by constructing feature vector and applied SVM, Naive Bayes, and Maximum entropy along with the Semantic Orientation based WordNet in order to extract similarity from the content feature. This work measured the performance of classifier in terms of recall, precision and accuracy. Yan Dang et al. [24] proposed sentiment classification using lexicon-enhanced method, which combined machine learning and semantic-orientation to improve sentiment classification performance. This work conducted experiment on online product reviews and measured the performance in terms of overall accuracy, average precision, average recall and average F-Measure. Shubhodip Saha et al. [25] detected sarcasm by analysing the Twitter posts using Naïve Bayes and SVM classifiers and its accuracy was calculated using WEKA tool. For pre processing of data, TextBlob was used, which is one of the Natural Language Toolkit (NLT) packages. This work found the polarity of tweets using RapidMiner. Ali Hasan et al. [13] adopted a hybrid approach: sentiment analyzer with machine learning. This work applied Naïve Bayes and SVM to classify the election sentiments. The work compared different sentiment lexicon libraries namely, Word Sequence Disambiguation (W-WSD), SentiWordNet and TextBlob and claimed that W-WSD outperformed the other lexicons.

B. Behavioural analysis

Qiang Yan et al.[31] built a microblog social network to analyze the user behavior. This work characterized the behavior of following and being followed as out-degree and in-degree respectively. This model had short diameter of connected graph, short average path length and high average clustering coefficient. Based on empirical analysis, this paper reported that inducing drive and spontaneous drive lead to the behavior of posting microblogs. This work performed simulation of the proposed model, which matched well with Twitter. Reza Zafarani et al. [32] analyzed the behaviours observed on social media and proposed that it is possible to classify these behaviours into individual and collective behaviour. Individual behaviour is exhibited by a single user, whereas collective behaviour is observed when a group of users behave together. This work used data mining and statistical approach to measure group level user behavior. This work considered case studies of individual behaviour analysis (community membership behaviour and connecting users across sites) and movie revenue prediction using Twitter and user migration in social media as instances of collective behaviour analysis. Qiwei He et al.

993

[33] investigated the relationship between the posts on Face Book (FB) and Self-Monitoring(SM) skills and reported that textual posts on the FB could partially predict the users' SM skills. This work presented a model to evaluate the relationship between the posts and SM skills. It evaluated the quality of responses to the Snyder's Self-Monitoring Questionnaire collected via the Internet using Item Response Theory (IRT) model and explored the textual features of the posts in different SM-level groups. The prediction of posts led to 60% accuracy compared with the Snyder's SM scale classification. Huiqi Zhang et al. [34] developed a socioscope model for social-network and human behavior analysis based on mobile-phone call-detail records. This work used multiple probability and statistical methods for quantifying social groups, relationships, and communication patterns and for detecting human-behavior changes. Additionally, proposed a new index to measure the level of reciprocity between users and their communication partners. This work used they used real-life call logs for the validation of the results. Jose Simoes et al. [35] extended regular user profile data to accommodate social, context, device and policy related information to make it user behavior aware. By doing this, social network providers can predict user behaviours and will enable them to provide user-oriented services. Hajra Waheed et al. [36] aimed to identify, analyze, and classify the characteristics associated with user behavior using Thematic analysis for the classification of user behavior. This study reported that activities performed on Social Networking Sites (SNS) are either associated with user behavior or reflect personality characteristics. The findings of this study could be used by SNS providers to develop more user-centered applications.

C. Behaviour Prediction

Nhat Hai Phan et al. [26] introduced an ontology-based Restricted Boltzmann Machine (ORBM) model for human behaviour prediction in health social networks. This work proposed a bottom-up algorithm to learn the user representation from ontologies. Then the user representation was used to incorporate self-motivation, social influences, and environmental events together in a human behaviour prediction model. A well-known deep learning method, Restricted Boltzmann Machines (RBMs) was used to simulate the interactions among the behaviour determinants. Pasquale de Meo et al. [27] presented an in-depth analysis of the user behaviours on different Social Sharing platforms namely, Flickr, Delicious and Stumble Upon. This work combined social network analysis with semantic analysis to characterize the tagging behaviour and the tendency to create friendship relationships of the users of these platforms. Sarka Zehnalova et al. [28] studied user roles in social network. These roles describe the node and its activity in the network over time. Self-organizing maps (SOMs) are used for nodes role identification and for discovery of all the roles present in the network. This work experimented with a large scale co-authorship network constructed from a DBLP dataset. Monika Mital et al. [29] investigated the multihoming behaviour of users on social networking web sites. This work explored how social networking websites might achieve critical network size due to multihoming by users. When users use products/services from two competing product/service providers to avail benefits of network size, it is called

multihoming. The multihoming resulted in increased utility for the users of social networking websites, when the two websites are differentiated in terms of features. Radoslaw Brendel et al. [30] proposed a behavioural model based on user roles in social network considering attributes of actors and attributes of relations. For dynamic social network, pattern subgraphs and deterministic finite-state automata were proposed and each role was modeled by an automaton. The analysis is conducted in regard to attributes of actors and relations.

The existing works on sentiment and behaviour analysis were merely a classification of sentiments into positive, negative and neutral sentiments. But failed to analyze the posted text to understand it's inner meaning like what the tweet is about, whether it is a general statement or a hate speech or an offensive statement and what type of personality a particular user is. In addition, most of the existing works used the (single class) supervised classifiers namely Naive Bayes, Support Vector Machine and Random Forest which did not show accurate classification and prediction results. Hence, there is a need to develop a machine learning-based model in order to perform more deep analysis and accurate prediction of the personality of a user from his/her posted data, which is essential for identification and prevention of immoral activities happening in a social network.

III. PROPOSED WORK

The social networking site considered in this work is Twitter. Twitter is a micro blogging service, with 320 millions of users tweet messages about a maximum of 140 characters [37]. A follower on Twitter can view all of the tweets posted by the followed persons. The user can follow any other user, but there is no compulsion for following back. Some influence accounts have millions of followers without following them back, hence there is no requirement on reciprocity of relationship. Twitter provides an Application Programming Interface (API) that enables real time and easy access to all tweets and information of users. The proposed work comprised of many phases namely, data pre-processing, classification, sentiment analysis, topic analysis and personality analysis.

A. Data Pre-processing

The first and foremost step in SNA is data pre-processing, which is done to structure the unstructured tweet data in order to extract the vital information from the raw data and to keep the data ready for analysis. For example, Twitter data is highly unstructured as it follows an informal communication.

Having a more structured and meaningful can improve the knowledge of the whole data mining process. Tweets are highly unstructured and non-grammatical and extensive usage of acronyms. We collected text data from real time tweet data as sample data and performed pre processing techniques like stop words removal, removing URLs, punctuations removal and converting all to lower case.

Pre-processing Phases:

The multi-phase data pre-processing technique works on the highly unstructured data to produce a meaningful
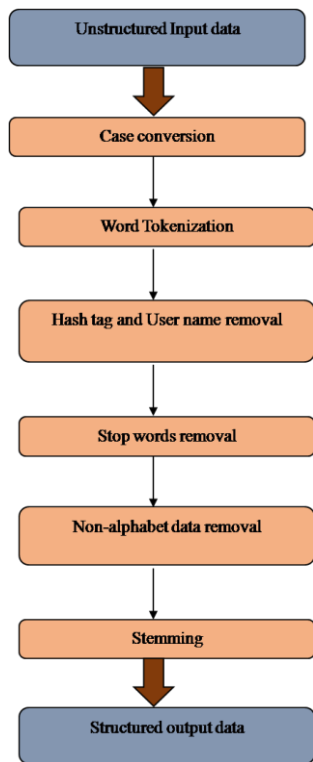
structured data and is illustrated in Figure 1.

Fig. 1. Data Pre-Processing Phases

The pre-processing phases illustrated in Figure 1 are explained clearly with a real-time twitter data. For example, the tweet by our Honourable Indian Prime Minister is an example for raw data and there is a need to extract vital information from this raw data by means of pre-processing the data.

Fig. 2. Sample Tweet Data

Here is the text from his tweet,

"The fraternity of doctors is special. During #MannKiBaat, I conveyed greetings for Doctor's Day. Doctors from India are admired world over for their skills and excellence."

Step 1: Case conversion: Converting the data to lower case

the fraternity of doctors is special. during #mannkibaat, i conveyed greetings for doctor's day. doctors from india are admired world over for their skills and excellence.

Step 2: Word tokenization: Splitting the sentence into individual words

['the', 'fraternity', 'of', 'doctors', 'is', 'special', '.', 'during', '#mannkibaat', ',', 'i', 'conveyed', 'greetings', 'for', "doctor's", 'day', '.', 'doctors', 'from', 'india', 'are', 'admired', 'world', 'over', 'for', 'their', 'skills', 'and', 'excellence', '.']

Step 3: Removal of hash tags and usernames

A tweet may have a hash tag, which starts with a '#' symbol and username, which starts with '@' symbol. Generally, these data are not contributing to analysis and are removed.

['the', 'fraternity', 'of', 'doctors', 'is', 'special', '.', 'during', ',', 'i', 'conveyed', 'greetings', 'for', "doctor's", 'day', '.', 'doctors', 'from', 'india', 'are', 'admired', 'world', 'over', 'for', 'their', 'skills', 'and', 'excellence', '.']

Step 4: Removal of stop-words

Stop words are generally meaningless pieces of data, which needs to be removed. For example, *a, was, the, etc* are stop words.

['fraternity', 'doctors', 'special', '.', ',', 'conveyed', 'greetings', "doctor's", 'day', '.', 'doctors', 'india', 'admired', 'world', 'skills', 'excellence', '.']

Step 5: Removal of non-alphabetical data

['fraternity', 'doctors', 'special', 'conveyed', 'greetings', 'doctors', 'day', 'doctors', 'india', 'admired', 'world', 'skills', 'excellence']

Step 6: Stemming

Stemming converts the words into its root form. For example, the word *loving* is converted into *love*. The resulting root word need not to be in a proper English language. This process is also called as lemmatization.

[fratern doctor special convey greet doctor day doctor indiaadmir world skill excel]

Finally, the important information from the data is mined. The Algorithm 1 depicts the multi-phase data pre-processing on a tweet.

Algorithm 1: Data Pre-processing

```
function_preprocess (T)
    for every tweet from T do
        tweet_text = lower_case (tweet);
        words= tokenizer(tweet_text);
        for word in words do
            if word.startswith (@ or #) then
                pass;
            else
                preprocessed_word (word);
            end if
        end for
        for word in preprocessed_word do
          if not in english.stop_words then
            stop_words_removed= word;
        end for
        for word in stop_words_removed do
    if not in list.punctuation then
            punctuation_removed= word;
        end for
        for word in punctuation_removed do
            stemmed_word= stem(word);
        end for
        for word in stemmed_word do
            processed tweet = join(word);
        end for
    end for
    return processed tweet
```

B. Logistic Regression-based Multi-Class Classification

This work applies one of the regression methods namely, Logistic Regression. Logistic Regression analyses a data set which contains one or more independent variables that determine an outcome. For instance, determining whether an email is spam or not have only two possible outcomes. This is called binary classification. In machine learning, multiclass classification is the problem of classifying instances into one of three or more classes. The existing multi-class classification techniques can be categorized into (i) Transformation to binary (ii) Extension from binary and (iii) Hierarchical classification and this work applies transformation to binary approach.

C. Transformation to binary

One vs Rest (OvR)

The OVR reduces the problem of multiclass classification to multiple binary classification problems. It can be categorized into One vs Rest and One vs One. The techniques developed based on reducing the multi-class problem into multiple binary problems can also be called problem transformation techniques. Since, Logistic regression is a binary classifier, this work uses One vs Rest for multi class Labeling.

To analyse and classify the Twitter data, a Supervised Machine Learning Classification algorithm named Logistic Regression is used. The logistic regression works by using a sigmoid function (Logistic function) $\hat{p}$, which always returns a value between 0 and 1. Sigmoid is an activation function, no matter how large the input size is, the sigmoid function will return the values between 0 to 1. The values greater than 0.5 denotes positive or 1, else 0. The Equation 1 denotes a model for the logistic prediction $\hat{y}$.

$$\hat{y} = \begin{cases} 0 \ if \ \hat{p} < 0.5 \\ 1 \ if \ \hat{p} \geq 0.5 \end{cases}$$

(1)

D. Dataset

The adequate data are needed in order to make any Machine Learning algorithm work. Hence, the Twitter data set is downloaded using the developed application and Figure 3 shows the downloading process. The Twitter Application Programming Interface (API) is used for downloading the data from the twitter. The data are stored in a local SQLite database as it comes along with Python package. The SQLite database columns are tweet id, username, date, tweet text, likes, retweet, sentiment (which ranges from -1 to 1), which expresses the thoughts of people from negative, neutral and positive sentiment. This sentiment is not present in the twitter and hence this work uses an NLP library named ,Text-Blob to process the sentiment data.
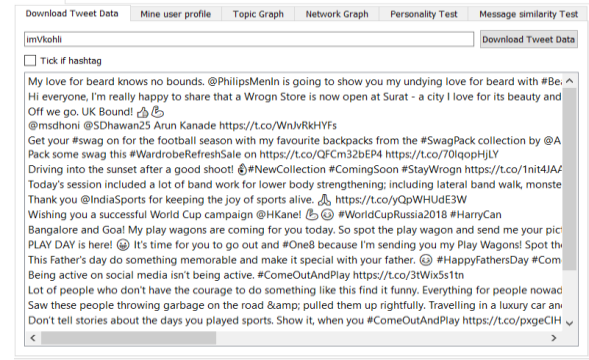

Fig.3. Downloading Twitter data

E. Sentiment Analysis

Sentiment analysis is used to determine the attitude of a person. Generally, sentiment values range between -1 and 1. This work uses a publicly available data-set from data world.[38] This data-set consists of about 15 thousand entries and only two columns namely, "tweet_text" and "orig_does_this_tweet_contain_hate_speech" are extracted for analysis. The Logistic regression-based machine learning model is trained to learn from this extracted data in order to predict the future tweets posted by the persons. The requirement for any mathematical algorithm to work is that the data must be in numeric. Hence, the text-based data set need to be converted to numerical data. For this, a bag-of-words model is used to convert the data set. The bag-of-words model is a simplifying representation used in NLP. In this model, a text is represented as a bag of its words without regard to its grammar and word order. For example, the text "Roses are red. Jasmines are White." is converted to numerical data, where the numeric value under each word in the Table 1 represents the frequency of the word in the text.

Table- I: Bag-of-Words Model

Roses	Are	Red	Jasmines	White
1	2	1	1	1

(1)

All the data in the *tweet_text* column in SQLite database is converted into numerical data. The *tweet_text* column is the independent variable, where the dependent variable is the result (eg. offensive, hate speech, general statement). Since Logistic regression is a binary classifier, this work uses a One vs. Rest methodology for multi-class classification. Once the model has been trained, its data is pickled or serialized, so that we need not train the model again and again while running the application. This pickled model is un-pickled while predicting the results. The sentiment analysis is done on the tweets posted by a person and the result is presented in Table II.

Table- II: Sentiment Analysis

Sentiment Type	Number of sentiments
Positive Sentiment	227
Negative Sentiment	149
Neutral Sentiment	573
Average Sentiment score	**0.037**

F. Analysing the personality of the person by his tweets

This work performs MBTI (Myers–Briggs Type Indicator) personality test on a person using the tweets posted by him.[39] For this, a publicly available data-set from Kaggle is used This data-set consists about 8,600 data for personality type and their speeches. Each row consists of the personality type and fifty speeches given by the person and the high level description of the sixteen personality types is depicted in Figure 4. The principles followed in sentiment analysis is also used for analysing the personality of the person. The machine learning model is trained in this data-set and predicts the tweets posted by the person. Finally, the mean of the result is calculated and result is updated in the Graphical User Interface (GUI) depicted in Figure 5. It is quite evident from the Figure 5 that the person belongs to ISFJ personality.

Fig. 4. High-Level Description of the Sixteen Personality Types by MBTI
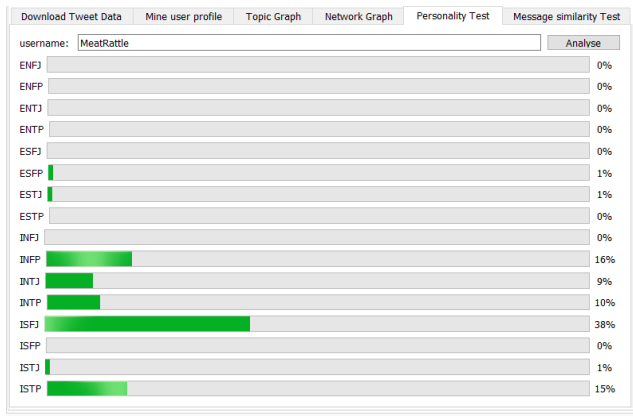

Fig.5. Personality Analysis

IV. COMPARATIVE STUDY

For comparative purpose, this study considers other supervised classifiers namely, Decision Tree and Random Forest. All data-sets are trained to scikit-learn's compatible ML model using a Core i7 processor with extreme pipelining along with 16 Giga Bytes of RAM. The confusion matrix for offensive /hate speech is plotted based on the general confusion matrix depicted in Fig 6.
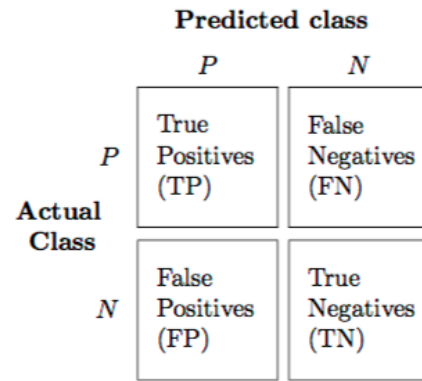

Fig. 6. Confusion Matrix

Using the Confusion Matrix in Figure 7, the equation for accuracy measure can be formed and is represented in Equation 2. The algorithms are compared on the accuracy measure and the result is presented in Table 3.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \qquad (2)$$

Table- III: Accuracy score of Logistic Regression over other famous algorithms

Compared Algorithms	Accuracy score for Personality Classification
Logistic Regression	0.564
Decision Tree classifier	0.488
Random Forest classifier	0.361

Hence, it is evident from the result in Table 3 that the Logistic Regression outperforms other classifiers on the accuracy.

V. CONCLUSION

This paper analyses the online social network, Twitter for reconnaissance of people for finding their potentiality. The Machine Learning technique, Logistic Regression mines the tweet data of a person to get his/her 360-degree profiling. This profiling is helpful to identify the personality type of a person using MBTI personality test and Government can use this analysis to identify the persons involved in spreading the fake news. In future, we can apply on other social media sites like Instagram, Snapchat, but the challenge is downloading the data especially images and processing them.

REFERENCES

[1] Alexandra Marin, Barry Wellman: The SAGE Handbook of Social Network Analysis, 2014.

[2] Fazli Can, Tansel Özyer, Faruk Polat: State of the Art Applications of Social Network Analysis, Lecture Notes in Social Networks, Springer International Publishing, 2014.

[3] Pooja Wadhwa, M.P.S Bhatia: Social Networks Analysis: Trends, Techniques and Future Prospects, IET Fourth International Conference on Advances in Recent Technologies in Communication and Computing (ARTCom2012), 2012, pp.1-6.

[4] Bogdan Batrinca, Philip C. Treleaven: Social media analytics: A survey of techniques, tools and platforms, AI & Soc, Springer, 30, 2015, pp.89–116.

[5] Martin Grandjean: A social network analysis of Twitter:

Mapping the digital humanities community, Cogent Arts & Humanities, 3: 1171458, 2016, pp. 1-14.

[6] Manav Masrani, Poornalatha G: Twitter Sentiment Analysis Using a Modified Naïve Bayes Algorithm, Information Systems Architecture and Technology: Proceedings of 38th International Conference on Information Systems Architecture and Technology – ISAT 2017, Advances in Intelligent Systems and Computing 655, 2017, pp.171-181.

[7] Jimeng Sun, Jie Tang: A Survey of Models and Algorithms for Social Influence Analysis, Social Network Data Analytics, Springer Science+Business Media, LLC 2011.

[8] Christoph Trattner, Frank Kappe: Social Stream Marketing on Facebook: A Case Study, International Journal of Social and Humanistic Computing, 2(1), 2013, pp.86-103.

[9] [9] Richard M Medina: Social Network Analysis: A case study of the Islamist terrorist network, Security Journal, vol. 27, 1, 2014, pp.97–121

[10] Jose Antonio Iglesias, Aaron Garcıa-Cuerva, Agapito Ledezma and Araceli Sanchis: Social Network Analysis: Evolving Twitter Mining, IEEE International Conference on Systems, Man, and Cybernetics (SMC 2016), 2016, pp.1809-1814.

[11] Anna Squicciarini, Sarah Rajtmajer, Christopher Griffin: Positive and negative behavioral analysis in social networks, Advanced Review, 2017, pp.1-12.

[12] David Décary-Hétu, Carlo Morselli: Gang Presence in Social Network Sites, International Journal of Cyber Criminology, 5 (2), 2011, pp.876–890.

[13] Ali Hasan, Sana Moin, Ahmad Karim and Shahaboddin Shamshirband: Machine Learning-Based Sentiment Analysis for Twitter Accounts, Mathematical and Computational Applications, 2018, pp.1-15.

[14] Sheena Angra, Sachin Ahuja: Machine learning and its applications: A review, 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), IEEE, pp.57-60.

[15] Amanpreet Singh, Narina Thakur, Aakanksha Sharma: A Review of Supervised Machine Learning Algorithms, 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), IEEE, pp.1310-1315.

[16] Parth Mehta , Hetasha Shah, Vineet Kori, Soumya Shukla , Mihir Shenoy: Survey of unsupervised machine learning algorithms on precision agricultural data, 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), IEEE, 2015, pp.1-8.

[17] Itır Önal, Ali Mert Ertuğrul: Effect of using regression in sentiment analysis, 2014 22nd Signal Processing and Communications Applications Conference (SIU), 2014, pp.1822-1825.

[18] Indra, Liza Wikarsa, Rinaldo Turang: Using Logistic Regression Method to Classify Tweets into the Selected Topics, Proceedings of the 2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS) 2016, pp.385-390.

[19] Muhammad Taimoor Khan, Mehr Durrani, Armughan Ali , Irum Inayat, Shehzad Khalid and Kamran Habib Khan: Sentiment analysis and the complex natural language, Complex Adaptive Systems Modeling, 4 (2), 2016, pp.1-19.

[20] Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, Manfred Stede: Lexicon-based methods for sentiment analysis, Computational Linguistics, 2011, 37, pp.267–307.

[21] Le T. Nguyen, Pang Wu, William Chan, Wei Peng, Ying Zhang: Predicting Collective Sentiment Dynamics from Time-series Social Media, Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM '12), ACM, 2012.

[22] Michele Di Capua, Emanuel Di Nardo, Alfredo Petrosino: Unsupervised Cyber Bullying Detection in Social Networks, Proceedings of the 23rd International Conference on Pattern Recognition (ICPR), IEEE, 2016, pp.432-437.

[23] Geetika Gautam, Divakar yadav, Sentiment Analysis of Twitter Data Using Machine Learning Approaches and Semantic Analysis: Proceedings of the Seventh International Conference on Contemporary Computing (IC3), IEEE, 2014, pp.437-442.

[24] Yan Dang, Yulei Zhang, HsinChun Chen: A Lexicon-enhanced method for sentiment classification: An experiment on online product reviews, IEEE Intelligent Systems, 25, 2010, pp.46–53.

[25] Shubhodip Saha, Jainath Yadav, Prabhat Ranjan: Proposed approach for sarcasm detection in Twitter, Indian Journal of Science and Technology, 10(25), 2017, pp.1-8.

[26] Nhat Hai Phan, Dejing Dou, Hao Wang, David Kil, Brigitte Piniewski: Ontology-based Deep Learning for Human Behavior Prediction in Health Social Networks, Information Sciences, 384, 2017, pp.298-313.

[27] Pasquale de Meo, Emilio Ferrara, Fabian Abel, Lora Aroyo, Geert-Jan Houben: Analyzing User Behavior across Social Sharing Environments, ACM Transactions on Intelligent Systems and Technology (TIST), 5(1), 2013, pp.14-42.

[28] Sarka Zehnalova, Zdenek Horak, Milos Kudelka, Vaclav Snasel: Using Self-organizing Maps for Identification of Roles in Social Networks, Proceedings of the IEEE Fifth International Conference on Computational Aspects of Social Networks, 2013, pp.44-49.

[29] Monika Mital, Sumit Sarkar: Multihoming behavior of users in social networking web sites: A theoretical model, Information Technology & People, 24(4), 2012, pp.378-392.

[30] Radoslaw Brendel, Henryk Krawczyk: Primary role identification in dynamic social networks, Proceedings of the IEEE International Conference on Computational Aspects of Social Networks (CASoN), 2011, pp.54 - 59.

[31] Qiang Yan, Lianren Wu, Lan Zheng: Social network based microblog user behavior analysis, Physica A 392, 2013, pp.1712–1723.

[32] Reza Zafarani, Huan Liu: Behavior Analysis in Social Media, IEEE Intelligent Systems, 29(4), 2014, pp.9–11.

[33] Qiwei He, Cees A.W.Glas: Michal Kosinski, David J.Stillwell, Bernard P.Veldkamp, Predicting self-monitoring skills using textual posts on Facebook, Computers in Human Behavior, 33, 2014, pp.69–78.

[34] Huiqi Zhang, Ram Dantu, João W. Cangussu: Socioscope: Human relationship and behavior analysis in social networks, IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 41(6), 2011, pp.1122–1143.

[35] Jose Simoes, Thomas Magedanz: Understanding and Predicting Human Behavior for Social Communities, Handbook of Social Network Technologies and Applications, Springer, 2010, pp.427-445.

[36] Hajra Waheed, Maria Anjum, Mariam Rehman, Amina Khawaja: Investigation of user behavior on social networking sites, Plos One, 2017, pp.1-19.

[37] Daniel E. O'leary: Twitter Mining for Discovery, Prediction and Causality: Applications and Methodologies, Intelligent Systems in Accounting, Finance and Management, Wiley, 22, 2015, pp.227–247.

[38] Thomas Davidson, Dana Warmsley, Michael Macy,Ingmar Weber: Automated Hate Speech Detection and the Problem of Offensive Language, Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017), pp.512-515.

[39] https://www.myersbriggs.org/my-mbti-personality-type/mbti-basics/home.htm?bhcp=1 [online- last accessed- October 2018]

AUTHORS PROFILE

R.Kanniga Devi received her Bachelor of Engineering in Computer Science and Engineering from Madurai Kamaraj University, Madurai in the year 2001 and Master of Engineering from Anna University, Chennai in the year 2008. She pursued her doctoral program at Kalasalingam Academy of Research and Education, Tamilnadu, India in the year 2018. She is a member of CiE, CSI, IAEng, ISTE and IEEE. She has 16 years of teaching experience. She has published 12 research papers in International Journals and Conferences. She got Science Academy Summer Research Fellowship and Tamilnadu Young Scientist Fellowship. Her research interest include Cloud computing, Graph theory applications in Cloud Computing, and Machine Learning. She is currently working as an Associate Professor in the Department of Computer Science and Engineering at Kalasalingam Academy of Research and Education, Tamil Nadu, India.