# Big Data Analysis of Facebook Users Personality Reconition using Map Reduce Back Propagation Neural Networks

3 authors:

Solomon A Akinboro
University of Lagos
36 PUBLICATIONS   52 CITATIONS

SEE PROFILE

Ibrahim Kazeem
Osun State University
20 PUBLICATIONS   35 CITATIONS

SEE PROFILE

Ayobami Taiwo Olusesi
17 PUBLICATIONS   39 CITATIONS

SEE PROFILE

# Big Data Analysis of Facebook Users Personality Recognition using Map Reduce Back Propagation Neural Networks

[1]Solomon Akinboro, [2]Ibrahim K. Ogundoyin and [3]Ayobami.  T. Olusesi
[1]Department of Computer Science, University of Lagos, Akoka, Nigeria
[2]Department of Information and Communications Technology, Osun State University, Osogbo, Nigeria
[3]Department of Electrical/Electronic and Computer Engineering, Bells University of Technology, Otta, Nigeria
{akinboro2002|ayobamiolusesi}@yahoo.com|olaac@ymail.com

**Abstract**- Machine learning has been an effective tool to connect networks of enormous information for predicting personality. Identification of personality-related indicators encrypted in Facebook profiles and activities are of special concern in most research efforts. This research modeled user personality based on set of features extracted from the Facebook data using Map-Reduce Back Propagation Neural Network (MRBPNN). The performance of the MRBPNN classification model was evaluated in terms of five basic personality dimensions: Extraversion (EXT), Agreeableness (AGR), Conscientiousness (CON), Neuroticism (NEU), and Openness to Experience (OPN) using True positive, False Positive, accuracy, precision and F-measure as metrics at the threshold value of 0.32. The experimental results reveal that MRBPNN model has accuracy of 91.40%, 93.89%, 91.33%, 90.43% and 89.13% CON, OPN, EXT, NEU and AGR respectively for personality recognition which is more computationally efficient than Back Propagation Neural Network (BPNN) and Support Vector Machine (SVM). Therefore, personality recognition based on MRBPNN would produce a reliable prediction system for various personality traits with data having a very large instance.

**Keywords**- Machine learning, Facebook, MRBPNN, Personality Recognition, Neuroticism, Agreeableness.

———————————— ◆ ————————————

## 1 INTRODUCTION

In recent years, huge amount of unstructured data has been generated from the social media websites like twitter, facebook and google. The data is in the form of text, images, videos and documents. Such huge volume of data is called a big data (*Sulochana*, 2016). Big Data analysis is a process of gathering data from different resources and organizes such data to discover meaningful facts and figures. It is the process of finding knowledge from bulk variety of data.  Analysis of data does not only help to determine the hidden facts and figures, but also rank the data with respect to the important information it provides (Shoro  and Soomro, 2015).

Social networking sites are applications that enable users to connect to the Internet by creating personal information profiles. These personal profiles can include any type of information, such as photos, video, audio files, and blogs. Facebook is the most used social network in almost every part of the world, recording about 2.7 billion individuals using at least one of the Facebook-owned platforms (Newsfeed.org, 2019). Facebook users dominate the social network market with 2.4 billion users, followed by WhatsApp (1.6 billion), and Instagram (one billion), both also owned by Facebook (Statista.com, 2020). Personality is a way person respond to a particular situation. It is the combination of characteristics that make an individual unique. Assessment of personality over the past two decades in various researches has revealed that personality can be defined by five dimensions known as Big Five personality traits (Tripathi, 2010).

The Big Five personality traits are extraversion, neuroticism, agreeableness, openness and conscientiousness. The Five Factor Model (FFM) is the most widespread and generally acceptable model for personality. It represents the basic structure that reflects human personality. FFM provides the nomenclature and conceptual framework that brings together research findings in psychology of individual differences and personality (Bachrach *et. al*, 2012). Personality recognition consists of the automatic classification of authors' personality traits, which can be compared against gold standard labels obtained by means of personality tests (Mairesse *et.al*. 2007). A variety of Facebook variables were expected to play a prominent role in establishing appropriate context for these investigations. Facebook profiles and activities provide valuable indicators of user's personality, revealing the actual rather than idealized or projected personality (Back *et.al*, 2010).

Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention. Machine learning is popular because computation is cheap and has abundant data to learn from. There are four types of machine learning algorithms, which are supervised, semi-supervised, unsupervised and reinforcement. Artificial Neural Networks (ANN) is an example of supervised machine learning algorithm.  ANN are models that are inspired by the structure or function of biological neural networks. They are a class of pattern matching that are commonly used for regression and classification problems. The most popular artificial neural network algorithms are: Perceptron, Multilayer Perceptron (MLP), Back-Propagation, Stochastic Gradient Descent, Hopfield Network and Radial Basis Function Network (RBFN).

Conventional classification algorithms such as back-propagation neural network and support vector machine

---

might not be suitable for analysing Facebook data. Moreover, when the data scale is large (usually in the size of gigabytes), such an algorithm might be very slow and sometimes cannot even run out a result at all (Liu, Hongyan & Gaoshan, 2010). It is a big challenge to analyze the vast volume of Facebook data to get relevance and different patterns of information on timely manner. Hence, this research work to explores the intrinsic attributes of Map Reduce Back-Propagation Neural Network (MRBPNN) for modelling complex relationships in Facebook data with the aim of finding reasonable patterns in the data for personality recognition. The proposed MRBPNN will improve the efficiency and scalability over large scale Facebook data

## 2 RELATED WORK

Personality is one of the fundamental and stable individual characteristics that can be detected from human behavioural data (Tripathi, 2010). Rise in social media usage has increased the attention to recognize personality traits by analysing the contents of user-generated text. The review of existing users' personality prediction techniques was done to identify the approaches adopted. Kalghatgi, Manjula, & Nandini (2015) reviews the techniques used for analysing social media data to identify important personality traits, that is, characteristics or qualities particular to a person. This can be used in a variety of areas such as marketing, business intelligence, psychology and sociology. A parallelism among individual's personality traits and his/her linguistic information is explored for analytics.

Liu, Jiajia, & Yuanchun (2016) proposed a new probabilistic topic model (PT-LDA model) to predict the personality traits within the framework of Five Factor Model. The proposed model extends the Latent Dirichlet Allocation (LDA) model to integrate the n-gram features into few latent topics and each topic is characterized by not only the multinomial distribution over words but also the Gaussian distributions over personality traits. They developed a Gibbs-EM algorithm to solve the proposed model iteratively based on Gibb's sampling and expectation maximization. Quantitative evaluation shows that PT-LDA is more accurate, efficient and robust than several baselines. Their experiment also shows that the proposed model can be used to extract the interpretable topics associated with each personality trait, which provides a new way to uncover user behaviours in online social network.

Kalani & Fernando (2016) developed persons' personality recognition system using machine learning techniques. They identified the personality traits from a face image using three separate methods; ANN, SVM and deep learning. Face area of an image is identified by a colour segmentation algorithm. Then the extracted image is input to personality recognition process. They identified features of the face manually and input them to ANN and SVM. Each personality trait is valued from 1 to 9. In their second attempt, m-SVM was used because outputs are multi-valued. ANN gave better results than m-SVM. In their third attempt they propose a methodology to identify personality traits using deep learning to iimprove the accuracy of the classification.

Siddique & Fung (2017) proposed a multilingual personality classifier that uses text data from social media and YouTube Vlog transcriptions. They mapped them into Big Five personality traits using a Convolutional Neural Network (CNN). Training was done firstly using unsupervised bilingual word embeddings from an English-Chinese parallel corpus, and use these trained word representations as input to their CNN. This enables their model to yield relatively high cross-lingual and multilingual performance on Chinese texts, after training on the English dataset for example. They also trained monolingual Chinese embeddings from a large Chinese text corpus and then trained their CNN model on a Chinese dataset consisting of conversational dialogue labelled with personality. Their results achieved an average F-score of 66.1 in multilingual task compared to 63.3 F-score in cross-lingual, and 63.2 F-score in the monolingual performance.

Yu & Konstantin (2017) applied deep learning methods to automatically learn suitable data representation for the personality recognition task using Facebook status updates data. They investigated several neural network architectures such as fully-connected (FC) networks, convolutional networks (CNN) and recurrent networks (RNN) on the *myPersonality* shared task and compared them with some shallow learning algorithms. Experiments showed that CNN with average pooling is better than both the RNN and FC. Convolutional architecture with average pooling achieved the best results of 60.0±6.5%. Information about users and what they expressed through status updates are such important assets for research in the field of behavioural learning and human personality.

Considering the rich semantic information in user-generated texts, Xue *et.al., (*2018), proposes a deep learning-based approach for personality recognition of online text posted by social network users. They first utilize a hierarchical deep neural network composed of a newly designed Attribute Recurrent Convolution Neural Network (AttRCNN) structure and a variant of the Inception structure to learn the deep semantic features of each user's text posts. Then they concatenate the deep semantic features with the statistical linguistic features obtained directly from the text posts, and fed them into traditional regression algorithms to predict the real-valued Big Five personality scores. Experimental results show that the deep semantic feature vectors learned from their proposed neural network are more effective. The approach that utilizes the concatenation of their deep semantic features and the statistical linguistic features as the input of the gradient boosting regression algorithm achieves the lowest average prediction error among all the approaches they tested.

Souril *et.al.*, (2018), presented a hypothesis that users by similar personality are expected to display mutual behavioural patterns when cooperating through social networks. With the goal of personality recognition in terms of analysing user activity within Facebook, they collected information about the personality traits of users and their profiles on Facebook. The participants of their study are 100 volunteers of Facebook users. They asked

the participants to respond to the NEO personality questionnaire in a period of 1 month in May 2012. At the end of this questionnaire, there was a link that asked the participants to permit the application to access their profiles. Based on all the collected data, classifiers were learned using different data mining techniques to recognize user personality by their profile and without filling out any questionnaire. By comparing classifiers' results, the boosting-decision tree was their proposed model with 82.2% accuracy which is more accurate than previous studies that were able to foresee personality according to the variables in their profiles in five factors for using it as a model for recognizing personality.

 Adi *et.al*. (2018) presented an optimization technique for automatic personality recognition based on Twitter in Bahasa, the mother tongue of Indonesians. They evaluated a series of techniques implementing hyperparameter tuning, feature selection, and sampling to improve the machine learning algorithms used. The personality prediction system is built on machine learning algorithms. Three machine learning algorithms was used in their study, namely Stochastic Gradient Descent (SGD), and two ensemble learning algorithms, Gradient Boosting, and stacking (super learner). By implementing the series of optimization techniques, the evaluation results show huge improvement by achieving 1.0 Receiver Operating Characteristic Area Under Curve (ROC_AUC) score with SGD and Super Learner.

An & Rivka (2018) proposed two deep learning structures for the task of personality recognition using acoustic-prosodic, psycholinguistic, and lexical features, and presented empirical results of several experimental configurations, including a cross-corpus condition to evaluate robustness. Their models outperform state-of-the art on the well-known myPersonality corpus, and also set a new state-of-the-art performance on the more difficult CXD corpus. Yang & Zi-Rui (2019) built a recommender system for recommending computer games to players according to their identified personality traits. They first applied text mining processes on some textual contents related to the players to identify their personality traits using the Five Factor Model. The same personality recognition process was also applied on contents related to games. The games with similar personality traits to the players were then recommended. Experiment was performed on 63 players and 2050 games with data collected from Steam and obtained good result.

Considering the personality recognition as a multi-label classification problem, Wang *et.al.* (2019) analyses the impact of imbalanced data distribution, positive and negative sample overlap on the machine learning classification model. The classification model is based on the data resampling technique, which can improve the classification accuracy. These problems can be solved once the data are effectively resampled. They presented a personality prediction method based on particle swarm optimization (PSO) and synthetic minority oversampling technique + Tomek Link (SMOTETomek) resampling (PSO-SMOTETomek). Apart from effective SMOTETomek resampling of data samples, the model is able to execute PSO feature optimization for each set of feature combinations. Analysis reveals that PSO-SMOTETomek method is efficient under a small dataset, and accuracy of personality recognition is improved by around 10% which is better than those of similar studies.

Han *et.al.* (2020), presented a novel interpretable personality recognition model based on a personality lexicon. First, they use word embedding techniques and prior-knowledge lexicons to automatically construct a Chinese semantic lexicon suitable for personality analysis. Based on this personality lexicon, they analyse the correlations between personality traits and semantic categories of words, and extract the semantic features of users' microblogs to construct personality recognition models using classification algorithms. Extensive experiments were conducted to demonstrate that the proposed model achieved significantly better performance compared to previous approaches. Several approached have been considered in the literature to predict users' personality traits especially on social medial platform. More research is required to explore the attributes of other approaches to ascertain their performance level. Hence, the need for this study to explore the intrinsic attributes of MRBPNN to model complex relationships in Facebook data.

## 3  PROCESSING FACEBOOK DATA FOR PERSONALITY RECOGNITION

MRBPNN technique was used to analyses Facebook data for personality recognition. A sample of 300 user instances from Facebook (activity and demographic data) with 9,917 status updates was collected from MyPersonality project. The phases involved in the analysis includes: data generation, data collection, data transmission, data pre-processing, data storage, data analysis and decision making as shown in Figure 1.

**i. Data Generation-** The data used was generated from Facebook Page. The study of personality reflected in user's Facebook activities includes a wide range of features. Some of the most intuitively predicted indices are the statistical data for user's activities (e.g., number of likes, statuses, groups, tags, and events). Demographic characteristics such as: age and gender, were accounted for since their effect is known to manifest in the context under investigation. Egocentric network parameters such as density, brokerage, and betweenness, represent the number of friends. These parameters provided additional insight into user's social behaviour and are instrumental in assessing personality along several dimensions.

**ii. Data Collection -** Application Programming Interfaces (APIs) are instructions and tools for managing interactions between different software, and can be used to automatically extract social media data. Facebook provide APIs for free but impose restrictions on the amount or type of data that can be accessed. A sample of 300 user instances from Facebook (activity and demographic data) with 9,917 status updates used in our study was provided by MyPersonality project (http://mypersonality.org/wiki) with the aid of CrowdTangle API. In this phase, the data is obtained from the data sources using the log files
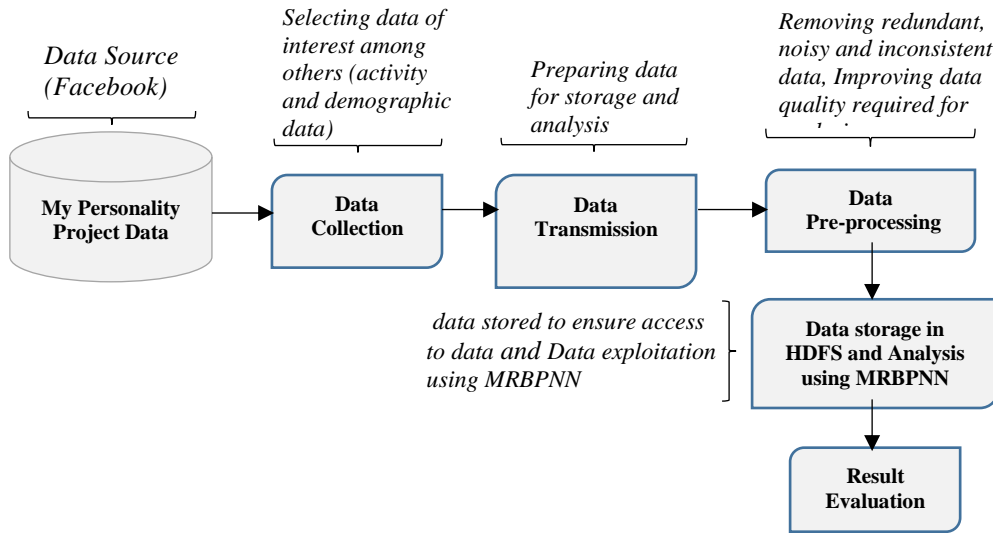
Fig. 1: The analysis of Facebook data for Personality Recognition

### iii. Data Transmission

In this phase, data was transferred from the source to the data centre and further transferred within the data centre for processing and analysis. Apart from storage of data, data centre also help in collecting, organizing and managing data.

### iv.   Data Pre-Processing

The data was pre-processed to improve the data quality required for analysis. This is done by removing redundant, noisy and inconsistent data. This helps to improve the accuracy of the analysis and reduce the storage expenses. Hadoop reduces the 9,917 features to 77 for each of the personality traits.

### 3.1 THE DESIGN OF MRBPNN ALGORITHM

MRBPNN targets the scenario in which Back Propagation Neural Network (BPNN) has a large volume of testing data to be classified (Facebook data in this case).

Consider $S$ and $s_i$ , $s_i = \{a_1, a_2, a_3, \ldots a_n\}$, $s_i \in S$, Where $S$ denotes a dataset; $s_i$ denotes an instance; *in* denotes the length of $s_i$; it also determines the number of inputs to the neural network;  the inputs are capsulated by a format of ($instance_k$, $target_k$, type);       $target_k$ represent the desirable output of $instance_k$ of a training instance; **type** field has two values, "train" and "test," which marks the type of $instance_k$; if "test" value is set, $target_k$ field should be left empty. Files which contain instances are saved into Hadoop Distributed File System (HDFS) initially. Each file contains all the training instances and a portion of the testing instances. Therefore, the file number $n$ determines the number of mappers to be used as shown in figure 2. The file content is the input to the MRBPNN.

When the algorithm starts, each mapper initializes a neural network. As a result, there will be $n$ neural networks in the cluster. Moreover, all the neural networks have exactly the same structure and parameters. Each mapper reads data in the form of ($instance_k$, $target_k$, type) from a file and parses the data records. If the value of type field is "train," $instance_k$ is input into the input layer of the neural network.  Every input of a neuron has a weight $\omega_{ij}$, where $i$ and $j$ represent the source and

destination of the input.  The number of neurons in the hidden layer is determined by users. Similarly, the number of neurons in the output layer is determined by the number of classifications.

Each neuron also maintains an optional parameter $\theta_j$ which is actually a bias for varying the activity of the $j^{th}$ neuron in a layer. Let $o_{j'}$  denote the output from a previous neuron and let  $o_J$ denote the output of this layer. The input $I_j$ of the neurons located in both the hidden and output layer can be represented by

$$I_j = \sum_i \omega_{ij} o_{j'} + \theta_j \qquad (1)$$

The output of a neuron is usually computed by the sigmoid function, so the output $o_J$ can be computed by

$$o_J = \frac{1}{1+e^{-I_j}} \qquad (2)$$

Completion of the feed forward process is by (1) and (2), where the neural network in each mapper starts the back-propagation process. Let  $Err_j$ represent the error-sensitivity and let $t_j$ represent the desirable output of neuron in the output layer; thus,

$$Err_j = o_j(1 - o_j)(t_j - o_j) \qquad (3)$$

Let $Err_k$ represent the error-sensitivity of one neuron in the last layer and let $\omega_{kj}$ represent its weight; thus, $Err_j$ of a neuron in the other layers can be computed using

$$Err_j = o_j(1 - o_j)\sum_k Err_k \omega_{kj} \qquad (4)$$

After $Err_j$   is computed, the weights and biases of each neuron are tuned in back-propagation process using

$$\Delta\omega_{ij} = Err_j o_j \qquad (5)$$
$$\omega_{ij} = \omega_{ij} + \Delta\omega_{ij} \qquad (6)$$
$$\Delta\theta_j = Err_j \qquad (7)$$
$$\theta_j = \theta_j + \Delta\theta_j \qquad (8)$$

The neural network inputs $instance_{k+1}$., repeat the feed forward and back-propagation process until all the instances which are labeled as "train" are processed and the error is satisfied. Each mapper starts classifying

instances labeled as "test" by running the feed forward process. As each mapper only classifies a portion of the entire testing dataset, the efficiency is improved. At last, each mapper outputs intermediate output in the form of ($instance_k$, $o_{jm}$), where $instance_k$ is the key and $o_{jm}$ represents the output of the $m$th mapper. The reducer starts collecting and merging all the outputs of the mappers. Finally, the reducer outputs ($instance_k$, $o_{jm}$) into HDFS. In this case, $o_{jm}$ represents the final classification result of $instance_k$. Figure 2 shows the architecture of MRBPNN.

The performance of MRBPNN on the analysis of Facebook data for personality recognition was evaluated based on precision, true positive rate/recall (TP), false positive rate (FP), and F-measure using confusion matrix. These terms were calculated using:

$$Precision = \frac{TP}{TP + FP} \qquad (9)$$

$$TruePositive\ Rate/Recall = \frac{TP}{TP+FN} \qquad (10)$$

$$FalsePositiveRate = \frac{FP}{TN+FP} \qquad (11)$$

$$F - Measure = 2 * \left(\frac{Precision * Recall}{TN + FP}\right) \qquad (12)$$

The MRBPNN technique in the analysis of Facebook data was done in MATLAB R2012a environment. To validate the MRBPNN models, a set of experiments was performed to investigate how accurate the model is at predicting personality traits. The experiment uses five factor model. The five-factor model of personality is a hierarchical organization of personality traits in terms of five basic dimensions: Extraversion (EXT), Agreeableness (AGR), Conscientiousness (CON), Neuroticism (NEU), and Openness to Experience (OPN).

## 4 RESULT AND DISCUSSION

The five-factor model was analysed with MRBPNN with respect to evaluation of the classification model using TP, FP, precision, accuracy and F-measure. A total of about 9,917 statuses was used in this work. The application of Hadoop reduces the 9917 features to 77 and presented to

MRBPNN for personality prediction. Each of the personality traits uses the 77 features i.e. OPE - 77; NEU – 77; CON – 77; EXT –77; AGR – 77.
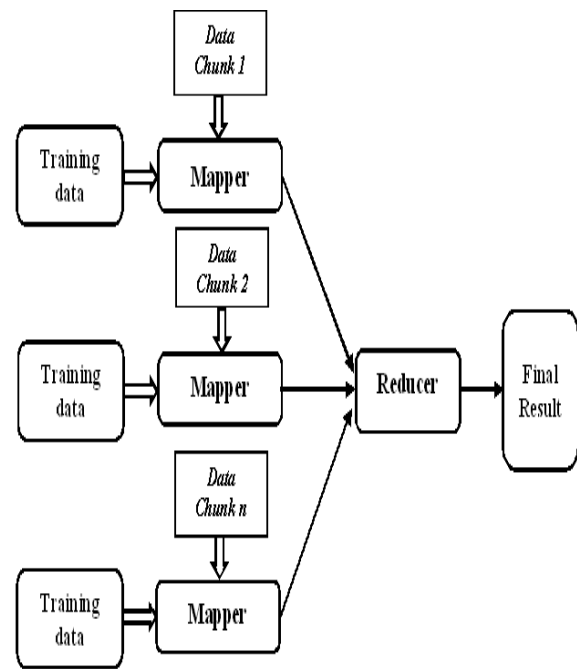


Fig. 2: Architecture of MapReduce Back Propagation Neural Network (Liu et. al., 2015)

The experimental result for the analysis of Facebook data from my personality project base on five measure of personality structure with respect to the aforementioned performance metrics are presented in Tables 1 and 2. The obtained result reveals that MRBPNN model is computationally efficient in terms of precision, FP, TP, accuracy and F-measure when compared with Back Propagation Neural Network (BPNN) and support vector machine (SVM). The result also shown that BPNN is more computational efficient than SVM using the same parameters.

Table 1.  Comparison of the Algorithm using EXT, NEU and AGR

| Trait/Measure | EXT | | | NEU | | | AGR | | |
|---|---|---|---|---|---|---|---|---|---|
| | MRBPNN | BPNN | SVM | MRBPNN | BPNN | SVM | MRBPNN | BPNN | SVM |
| TP Rate (%) | 94.5946 | 89.1892 | 86.4865 | 91.8919 | 86.4865 | 83.7838 | 91.8919 | 83.7838 | 81.0811 |
| FP Rate (%) | 8.10811 | 13.5135 | 16.2162 | 10.8108 | 16.2162 | 18.9189 | 10.8108 | 13.5135 | 16.2162 |
| Precision (%) | 92.1053 | 86.8421 | 84.2105 | 89.4737 | 84.2105 | 81.5789 | 89.4737 | 86.8421 | 84.2105 |
| Accuracy (%) | 91.3304 | 87.8378 | 85.1351 | 90.4315 | 85.1351 | 82.4324 | 89.1342 | 87.8378 | 85.1351 |
| F-Measure (%) | 93.3334 | 87.3371 | 84.6703 | 90.6667 | 84.6703 | 82.0054 | 90.6667 | 87.3371 | 84.6703 |

Table 2. Comparison of the Algorithm using CON and OPN

| Trait/Measure | CON | | | OPN | | |
|---|---|---|---|---|---|---|
| | MRBPNN | BPNN | SVM | MRBPNN | BPNN | SVM |
| TP Rate (%) | 94.5946 | 89.1892 | 86.4865 | 97.2973 | 91.8919 | 89.1892 |
| FP Rate (%) | 5.40541 | 13.5135 | 16.2162 | 2.7027 | 8.10811 | 10.8108 |
| Precision (%) | 94.5946 | 86.8421 | 84.2105 | 97.2973 | 91.8919 | 89.1892 |
| Accuracy (%) | 91.4031 | 87.8378 | 85.1351 | 93.4503 | 91.8919 | 89.1892 |
| F-Measure (%) | 94.5946 | 87.3371 | 84.6703 | 97.2973 | 91.8919 | 89.1892 |

# 5 CONCLUSION

Analysing big data does not only help to determine the hidden facts and figures, but also rank the data with respect to the important information it provides. Identification of personality-related indicators encrypted in Facebook profiles and activities are of special concern in most research efforts. This research modelled user personality based on set of features extracted from the Facebook data using Map-Reduce Back Propagation Neural Network (MRBPNN).

The essential features of map-reduce back propagation neural network were modelled in personality recognition, to determine its effectiveness on big data analysis using my personality project Facebook dataset. About 9,917 Facebook status updates were used with the MRBPNN model. The performance of the MRBPNN classification model was evaluated in terms of five basic personality dimensions: Extraversion (EXT), Agreeableness (AGR), Conscientiousness (CON), Neuroticism (NEU), and Openness to Experience (OPN) using True positive, False Positive, accuracy, precision and F-measure as metrics at the threshold value of 0.32. The MRBPNN model was bench marked with BPNN and SVM. The experimental results reveal that MRBPNN model has accuracy of 91.40%, 93.89%, 91.33%, 90.43% and 89.13% CON, OPN, EXT, NEU and AGR respectively for personality recognition which is more computationally efficient than Back Propagation Neural Network (BPNN) and Support Vector Machine (SVM).

In view of this, personality recognition based on MRBPNN would produce a more reliable prediction system for various personality traits with data having very large instances. Also, MRBPNN based personality recognition system can be used to prepare a personality-oriented interface to make user most receptive. This idea can be used in many fields including marketing and advertising. Personality traits predicted with the application of social media can be used for building better and dynamic recommender system.

# REFERENCES

Adi, G., Michael, T., Veronic,a O., & Derwin, S. (2018). Optimization for Automatic Personality Recognition on Twitter in Bahasa Indonesia. In proceedings of 3rd International Conference on Computer Science and Computational Intelligence, 135, 473-480.

An, G., & Rivka, L. (2018). Lexical and Acoustic Deep Learning Model for Personality Recognition. In proceedings of Interspeech 2018, Hyderabad, 1761-1765

Bachrach, Y., Kohli, P., Graepel, T., Stillwell, D., Kosinski, M. (2012) Personality and patterns of Facebook usage. Proceedings of the ACM Web Science Conference, 36–44.

Back, M., Juliane, S., Simine, V., Sam, G., Stefan, S., Boris, E., & Samuel, G. (2010). Facebook Profiles Reflect Actual Personality, Not Self-Idealization. Journal of *Psychological Science* 21(3), 372-374.

Han, S., Hailiang, H., & Yuqing, T. (2020). Knowledge of words: An Interpretable Approach for Personality Recognition from Social Media. Journal of Knowledge-Based Systems, 194, 1-25

Kalani, I., & Fernando, T. (2016). Persons' Personality Traits Recognition using Machine Learning Algorithms and Image Processing Techniques. International Journal of Advances in Computer Science, 5(1), 40-44.

Kalghatgi M., Manjula, R., & Nandini, S. (2015). Social-Network-Sourced Big Data Analytics for Personality Prediction: A Review. International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE), 14(2), 382-388.

Liu, Y., Yang, J., Huang, Y., Xu, L., Li, S., & Qi, M. (2015). MapReduce based parallel neural networks in enabling large scale machine learning. Journal of *Computational intelligence and neuroscience*, 2015, 1-13.

Liu, Y., Jiajia, W., & Yuanchun, J. (2016). PT-LDA: A Latent Variable Model to Predict Personality Traits of Social Network Users. Journal of Neurocomputing, 210, 155-163.

Liu, Z., Hongyan, L., & Gaoshan, M. (2010) MapReduce-based Backpropagation Neural Network over Large Scale Mobile Data. 2010 Sixth International Conference on Natural Computation (ICNC 2010), 1726-1730

Mairesse, F., Matthias, M., Marilyn, W. & Matthias, M. (2007). Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. *In Journal of Artificial intelligence Research*, 30, 457–500.

Newsfeed.org (2019). Available online at: https://newsfeed.org/facebooks-daily-active-user-count-grew-to-2-7-billion-in-q2-2019/ (accessed May 8, 2020).

Shoro, A., & Soomro, T.(2015). Big Data Analysis: Ap Spark Perspective. *Global Journal of Computer Science and Technology: C Software & Data Engineering.* 15(1), 7-14.

Siddique, F., & Pascale, F. (2017) . Bilingual Word Embeddings for Cross-Lingual Personality Recognition Using Convolutional Neural Nets. In proceedings *INTERSPEECH 2017*, Stockholm, Sweden, 3271-3275.

Souri, A., Shafigheh, H., & Amir, M. (2018). Personality Classification Based on Profiles of Social Networks' Users and The Five-Factor Model of Personality. Journal of Human Centric Computing and Information Science, 24, 1- 15.

Statista.com (2020). Available online at: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ (accessed May 8, 2020).

Sulochana, P. (2016). Social Data Analysis Using Big-Data Analytic Technologies APACHE FLUME, HDFS, HIVE. *Journal of Computer Science and Engineering.* 2(5), 1-3.

Tripathi, A. (2010). Personality prediction with social behavior by analyzing social media data – a survey. *Applied Computer Science*, 3-5.

Wang, Z., Chunhua, Wu., Kangfeng, Z., Xinxin, N., & Xiujuan, W. (2019). SMOTETomek-Based Resampling for Personality Recognition. IEEE Access, 20 , 1- 13

Xue, D., Lifa, W., Zheng, H., Shize, G., Liang, G. , Zhiyong, W., Xiaofeng, Z. & Jianshan, S. (2018). Deep learning-based personality recognition from text posts of online social networks. Springer Nature Journal of Applied Intelligence, 48 , 4232–4246.

Yang, H., & Zi-Rui, H. (2019). Mining Personality Traits from Social Messages for Game Recommender Systems. Journal of Knowledge Based Systems, 165, 157-168.

Yu, J., & Konstantin, M. (2017). Deep Learning Based Personality Recognition from Facebook Status Updates. In proceedings of IEEE 8th International Conference on Awareness Science and Technology, 383-387.