

Intelligent Personality Analysis on Indicators in IoT-MMBD-Enabled Environment



Rohit Rastogi, D. K. Chaturvedi, Santosh Satya, Navneet Arora, Piyush Trivedi, Akshay Kr. Singh, Amit Kr. Sharma and Ambuj Singh

Abstract Psychologists seek to measure personality to analyze the human behavior through a number of methods. As the personality of an individual affects all aspects of a person's performances, even how he reacts to situations in his social life, academics, job, or personal life. The purpose of this research article is to enlighten the use of personality detection test in an individual's personal, academics, career, or social life, and also provide possible methods to perform personality detection test. One of the possible solutions is to detect the personality, and the study is based on the individual's sense of humor. Throughout the twentieth century, psychologists show an outgoing interest in study of individual's sense of humor. Since individual differences in humor and their relation to psychological well-being can be used to detect the particular personality traits. Machine learning has been used for personality

R. Rastogi (✉) · A. Kr. Singh · A. Kr. Sharma · A. Singh
Department of Computer Science and Engineering, ABESEC, Ghaziabad, India
e-mail: rohit.rastogi@abes.ac.in

A. Kr. Singh
e-mail: akshay.17bcs1075@abes.ac.in

A. Kr. Sharma
e-mail: amit.17bcs1074@abes.ac.in

A. Singh
e-mail: ambuj.17bcs1078@abes.ac.in

D. K. Chaturvedi
Department of Electrical Engineering, DEI, Agra, India
e-mail: dkc.foe@gmail.com

S. Satya
Department of Rural Development, IIT-Delhi, Delhi, India
e-mail: ssatya@rdat.iitd.ernet.in

N. Arora
Department of Mechanical Engineering, IIT-Roorkee, Roorkee, India
e-mail: navneetrorkee@gmail.com

P. Trivedi
Centre of Scientific Spirituality, DSVV, Hardwar, India
e-mail: piyush.trivedi@dsvv.ac.in

© Springer Nature Singapore Pte Ltd. 2020
S. Tanwar et al. (eds.), *Multimedia Big Data Computing for IoT Applications*, Intelligent Systems Reference Library 163,
https://doi.org/10.1007/978-981-13-8759-3_7

detection involves the development and initial validation of questionnaire, which assesses four dimensions relating to individual differences in the uses of humor. Which are Self-enhancing (humor used to enhance self), Affiliative (humor used to enhance the relationship with other), Aggressive (humor used to enhance the self at the expense of others), and Self-defeating (the humor used to enhance relationships at the expense of self). Machine learning is gaining importance, nowadays, as it enables computers to perform self-learning without being programmed for a specific task. Psychologists seek to measure personality through different methods. Nowadays, the human being is so much complex that it is difficult to estimate the personality of an individual manually. The purpose of this chapter is to enlighten the use of personality detection test in academics, job placement, group interaction, and self-reflection. This book chapter provides the use of multimedia and IOT to detect the personality and to analyze the different human behaviors. It also includes the concept of big data for the storage and processing the data, which will be generated while analyzing the personality through IOT. We have used one of the supervised learnings called regression. Algorithms like Linear Regression, Multiple Linear Regression, Decision Tree, and Random Forest are used here for building the model for personality detection test. Among the different algorithms used in the project, Linear Regression and Multiple Linear Regression are proved to be the best so they can be used to implement the prediction of personality of individuals. Decision tree regression model has achieved minimum accuracy as in comparison to others so it is not the model, which can be used for training our model for determining the personality traits of individuals.

Keywords Sense of humor · Personality · Affiliative · Self-enhancing · Self-defeating · Aggressive · Python · Anaconda · Machine learning · Regression · Linear regression · Multiple linear regression · Decision tree · Random forest

1 Introduction

“Personality” has been taken from the word “Persona”. “Persona” is the characterization of a person which is different from his real-life character. Personality can be defined as the sum of characteristics, behavior, and qualities due to which a person is remarked as different or unique. The personality of a person can be described as the habitual behavior and emotional pattern of the person, which evolves from the biological and environmental factors.

The personality of a person can influence the social, personal, and professional life of an individual, therefore, we can use the personality detection test followed by some tasks to develop our personality so as to improve ourselves for better social, personal, and professional life. Personality psychology is the study of the psychology of personality, which analyses the differences in behavior of an individual.

Many approaches have been taken to studying personality, including learning and trait-based theories, biological, cognitive-based theories as well as psychodynamic and humanistic approaches. The possible method we are going to use in this research

article is taking a quiz from individuals and concluding their personality traits on the basis of result. The questionnaire used in this research article is based on the difference of sense of humor of person. The sense of humor of a person is the tendency of an individual to provoke laughter, but the sense of humor or the type of the humor the person uses can accentuate the personality trait of an individual. For example, the person can use the humor for sarcasm or to appreciate anyone or to get attention. Therefore, the sense of humor of a person or their goal for the usage of humor accentuates the personality trait of the individual. Throughout twentieth century, psychologists showed an interest in the study of individual differences in their sense of humor [1].

In the past two decades, researchers interested in detection personality of an individual since the personality of a person can influence the individual’s social, professional as well as personal life so to help them to develop their personality traits.

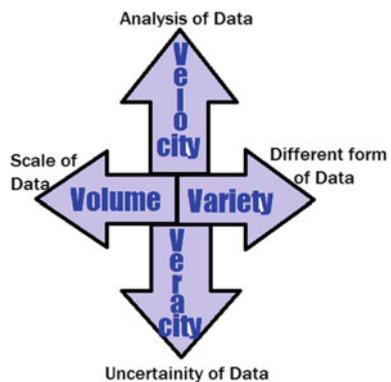
1.1 Big Data and IOT

Big data is the collection of data set, which is being generated at a tremendous rate around the world. These data can be structured or unstructured. These data are so large and complex that it is difficult to process them by using traditional data processing application. So, to overcome the processing and storage difficulty of big data, an open-source Hadoop is introduced.

Hadoop is an open-source distributed processing framework that is used to store a tremendous amount of data, i.e., big data and also for their processing. Big data has different characteristics which are defined using 4 V’s (Fig. 1) [2].

Internet of Things (abbreviated as IOT) is a system or network, which connects all physical objects to the internet through routers or network devices to collect and share the data without manual intervention.

Fig. 1 The characteristics of big data [8]



IOT provides a common platform and language for all physical devices to dump their data and to communicate with each other. Data emitted from various sensors are securely provided to the IOT platform and is integrated. Now, necessary and valuable information is extracted. Finally, results are shared with other devices to improve the efficiency and for better user experience automation.

1.2 Cloud Using Fog Computing with Hadoop Framework [3]

In today's world as data is growing at an exponential rate, and the need to store data is also increasingly parallel may be as a memory or as a record. Five–10 years back data is being stored in the hard disks of computers or in the smart phones. But due to the increase in number of profiles of individuals, there was a parallel increase in the data store. This causes to the insufficiency of the storage space, thus forcing more and more people to store their data onto the cloud. The main features of cloud computing are flexibility, scalability, efficiency, and multi-portability. Although cloud provides an excellent option to store the day but still, there are some loopholes in the security which is restricting the users from using cloud. Therefore, instead of securing the data only by authentication credentials like user name and password, an approach of using fog computing that is concoction of user's profile mapping using various behavior parameters and decoy data technology that is having a fake file of every file format, which will be used to launch a disinformation attack in case the user gets detected as an intruder came into being in order to maintain confidentiality of data. Thus, by getting large data set, our accuracy on personality in detecting human behavior will surely get increased.

1.3 Guiding Principles of Mist Computing [4]

Network [5] must provide information but not simply data. The network should deliver only information that has been requested and only when it has been requested. Dynamic creation of a system is based on information needs with end devices working together using a subscriber provider model. Devices must be situation aware, they must adapt to the information needs and the network configuration. We should not have static binding's rules for device and data providers. The devices must dynamically discover the data providers and execute the application.

Traditional wireless routing protocols are not suitable for routing in mist computing. Here, the routing protocols support device-to-device connectivity. It has been observed that suboptimal routing paths increase the bandwidth requirements of the network. Also, in mist, any node should be able to connect to any gateway eliminating dependence on a specific gateway. In a huge network, sometimes gateway failures force addition of a new gateway, the nodes near that gateway should have the ability to connect dynamically to such a node and a new route between itself

and the gateway efficiently established so that the crisis of information unavailability is solved within the tolerance time [6].

As we can observe in above representation, the mist nodes are responsible to process the data which has to be handed over to the IOT devices, which include the sensors and actuators with a physical connection among themselves. The mist node also monitors the quality of link parameters. At the gateway level, the functionalities are loading updated application rules and tuning application parameters, monitoring the health of local nodes, execution of computationally intensive services. Finally, at the cloud level, new applications can be deployed and applications can be coordinated along with service quality monitoring and monitoring health of the running applications [7].

1.4 Edge Computing [8]

Almost every use case and every connected device focused on by the Industrial Internet Consortium (IIC) require some sort of compute capability at its source, at the edge. Multiple sources define edge computing as “cloud computing systems that perform data processing at the edge of the network, near the source of the data”. While this is certainly true, it only scratches the surface of the immense power and remarkable capabilities that edge computing applications and architectures can provide to solve industrial internet users toughest challenges. But, as is typical with any powerful technology, innovative architectures and new terminology are needed to facilitate implementation, bringing increased complexity with it.

Interviews in corporate, during hiring procedure and ever for selecting the person in companies, for example, in agile software development methodology, it is very important to find the right person for the particular job. Personality tests in government organizations, As the jobs like them demands different characteristics and personality of an individual, for example, the ORQ (officer-like qualities are needed in defense some special characteristics are needed in IAS officer, that can differ from the characteristics needed for an IPS officer so in order to find right candidate the personality detection can be a great help.

It can help the aspirants of such jobs which need a good personality. So, they can prepare by working on their personality as if they will know the problem than they can easily find the solution.

This method can be used by the user to develop their personality. This method can be used in companies also to give appropriate work to appropriate employee according to their personality (nowadays, it is used in agile methodology of software development where the attitude of an employee plays a very important role). This program can be used in selection in defense, as the main requirement in defense is attitude and personality. This program can be used by the aspirants also to prepare for particular jobs or exams.

2 Effect of Personality in Person's Life

“Personality” has been taken from the word “Persona”. “Persona” is the characterization of a person which is different from his real life character. Personality can be defined as the sum of characteristics, behavior, and qualities due to which a person is remarked as different or unique. The personality of a person can be described as the habitual behavior and emotional pattern of the person which evolves from the biological and environmental factors.

The personality of a person can influence the social, personal, and professional life of an individual, therefore, we can use the personality detection test followed by some tasks to develop our personality so as to improve ourselves for better social, personal and professional life. Personality psychology is the study of the psychology of personality which analyses the differences in behavior of an individual [9].

The personality of a person can influence their social life. Personality of an individual influence his behavior, thought process, and actions, which affects the social life of a person. Personality trait such as aggressive can have a negative impact on the individual's life whereas the personality trait such as openness to experience can have the positive impact on individual's life. Today is the world of digitalization, people like to have a interactive and happy civic life or like to pretend in that way, which can be analyzed from their social networking accounts. People use humor to hide their feelings, they like to pretend and hide all the sentiments in humor.

The personality of a person can affect their personal life, today, everyone is so much busy in their careers and pretend to be happy in social networking sites that they forget to spend time with their closed ones. The personality of an individual directly influences the priorities of an individual.

How an individual behaves in different situations affects the relationships of an individual, personality of an individual can either eases the person's life of a person or can make it much difficult for that person.

The personality of a person can affect their professional life, in a workplace environment, the attitude of an individual is important. Someone's personality can be cheerful and upbeat. These are the person who fills their surrounding with happiness and has always positive thinking. Likewise, no one wants to work with those who have negative attitude as they make the environment unhealthy. In contrast, people with a positive attitude bring enthusiasm in workplace. Therefore, personality is very important in the professional life of an individual [10].

Therefore, we can use the personality detection test to examine our weakness and strengths so that we can improve our personality to make our social, personal and professional life better.

Various studies by researchers and psychologists have done to detect the personality traits of a person through learning and trait-based theories, biological, cognitive-based theories, and humanistic approaches. One possible method for personality detection test is to develop a questionnaire based on the basis of the answers given by the participants, and the personality trait of the individual can be predicted.

Various personality traits are defined by the human characteristics. It can be Optimistic, Realistic, Leader, Listener, Writer, Enthusiastic, Attentive, Responsible, Smart Kind, etc. [11].

3 Data Description

For data preparation, (Table 1), dataset is taken from “Martin, R. A., Puhlik-Doris, P., Larsen, G., Gray, J., & Weir, K. (2003). Individual differences in uses of humor and their relation to psychological well-being: Development of the Humor Styles Questionnaire. *Journal of Research in Personality*, 37(1), 48–75”.

Machine Learning is a study which performs learning without being explicitly programmed. So, if we want our program to predict the personality, we can run different machine learning algorithms with data of some previously performed personality test (Experience) and if learning is successfully done then, it will help in performing better prediction. We have used one of the supervised learning algorithms called regression. Regression is a modeling and analyzing technique, which computes the relationship between target (dependent variables) and predictors (independent variables). Regression allows comparing the effects of algorithms on variables of dataset. To make predictions, various kinds of regression techniques are available:

- **Linear Regression**—One of the most widely used techniques. It establishes the relation between independent variables (one or more) and dependent variable using a best-fit regression line. It is represented by an equation

Table 1 Column metadata

Features	Feature description	Data type
32 set of questions	These were statements rated on a five-point scale where 1 = Never or very rarely true, 2 = Rarely true, 3 = Sometimes true, 4 = Often true, 5 = Very often or always true (−1 = did not select an answer)	Numeric
Age	People of 14–87 years of age group	Numeric
Gender	Gender defines here the comparison of personality traits (1 = male, 2 = female, 3 = other)	Numeric
Accuracy	It is given by the users revealing that how much predictions are correct according to them. Scale (0–100)	Numeric
Affiliative	Who establish connections easily with others? Scale (1–5)	Numeric
Self-enhancing	Who makes people feel positive about themselves? Scale (1–5)	Numeric
Aggressive	It depicts the emotional distress and person who is likely ready to attack or confront. Scale (1–5)	Numeric
Self-defeating	Who works against himself/herself and unable to achieve the desired result. Scale (1–5)	Numeric

$$Y = b + aX + e \quad (1)$$

where b is an intercept, a is slope of line and e is error.

- **Multiple Linear Regression**—It is the most common form of linear regression analysis. It helps to establish the relationship when dependent variable is not only dependent on any one independent variable but on multiple independent variables. It is represented by an equation

$$Y = a_0x_0 + a_1x_1 + a_2x_2 + \dots + a_kx_k + e \quad (2)$$

where Y is response by k predictor variable $x_1, x_2, x_3 \dots x_k$; $a_0, a_1, a_2 \dots a_k$ are regression coefficients and e is the error.

It consists of five methods of building models, namely, All in, Backward Elimination, Forward Selection, Bidirectional Elimination, and Score Comparison.

- **Decision Tree Regression**—Decision tree algorithms are nonparametric supervised learning algorithms that are used for building regression and classification models, through this algorithm, we try to train our model to predict values of target variable by learning decision rule inferred from data features. Decision trees are able to manage both numeric and categorical data and they can also handle multi-output problems.
- **Random Forest Regression**—Random forest is like a forest of decision trees, i.e., it takes decision based on various decision trees. The final model of this technique is developed by the base models of various decision trees. It uses averaging for improving the predictive accuracy and controlling the over fitting. The equation used in this technique: $h_x = k_0x + k_1x + k_2x + k_3x + \dots$ here, h_x is final model and $k_0x, k_1x, k_2x, k_3x \dots$, are base models.

4 Previous Work Study

In a survey of Personality Computing [12], according to Alessandro Vienciarelli, Gelareh Mohammadi researches, personality can be explained by a few measurable individual characteristics. We can use various technologies (i.e., which can deal with human personalities) for personality computing approaches, i.e., understanding, prediction and synthesis of the human behavior This paper is a survey of such approaches and it provides the three main problems indicated in literature, i.e., (1) Automatic Recognition of personality, (2) Automatic Perception of personality, and (3) Automatic synthesis of Personality.

In results of Personality Recognition on Social Media with Label Distribution Learning [13], the main aim of the experiment was to efficiently, reliably, and validly recognize an individual's personality. Di Xue, Zheng Hong, and Shize Guo have stated the traditional ways of personality assessment are interviews, quiz prepared by the psychologists but these are expensive and not much practical in the social

media domain. It proposes the method of big five Personality Recognitions (PRs) with a new machine learning paradigm named label distribution learning. Out of 994 features, 113 features were extracted from active profiles and microblogs. Nine nontrivial conventional machine learning algorithms and eight LDL algorithms are used to train the personality traits prediction models. Results show that two of the proposed LDL approaches outperform in predictive ability.

In the development of Agile Person Identification through Personality Test and k-NN Classification Technique [14] Software methodology is a planning of the development of software; agile methodology is one of the most efficient software development methods. Agile method requires the cooperation and understanding between the employees, so this is important for the software project manager to assign the right work to the right people. Rintaspon Bhannarai, Chartchai Doungsa-ard have revealed the fact that the researchers provide five personality traits to predict suitable people for the agile method. A predicting method is done by using k-NN (k-nearest neighbor) classification technique. The k-NN techniques classify the unknown data from the known data by comparing the distance calculated between them. The most common method to calculate the distance between data is using Euclidian Distance. Using the participants, which involve software developers, testers, managers, etc., the pilot study is used to explore this technique that k-NN technique can be explored to predict the correct people for agile method.

In view to check individual differences in sense of humor and their relation with psychological well-being [15], the main aim of this survey was to identify the most common personality traits in males and females and which traits dominate in male and female. Here, the analysis was done by the quiz which was developed on the basis of different sense of humors of a person. After collecting the data from candidates, the analysis was done on the basis of some particular personality traits. Rod A. Martin, Patricia Puhlik-Doris, Gwen Larsen, Jeanette Gray, and Kelly Weir are the authors of related work and focused on the difference in individual's sense of humor and on the basis of that observation, they characterized the individual in terms of four personality traits that are aggressive, affiliative, self-enhancing, and self-defeating. The objective of the experiment was to analyze the proportionality of these personality traits in male and female.

To understand the product functions and operations, the personality detection test will predict the personality of an individual on the basis of four personality traits that are aggressive, affiliative, self-enhancing, self-defeating. It anticipates the personality of an individual by giving the dominance of each personality trait (aggressive, affiliative, self-enhancing, and self-defeating) in the range of 0–5. If the particular personality trait is 0 that means it is least in the particular person and if it reaches near 5 that means it is a very dominant personality trait in an individual.

The personality development test will first collect the input from the user/participant, when the user will perform the quiz. The answers collected will be used as the data, which will be passed to the multiple machine learning algorithms. The machine learning algorithms will predict the result on the basis of training set memory constraints—to run a python programs, 2 GB memory is required.

Operations that will be done by user are 1. The user/participant will perform the quiz, 2. On the basis of the result, they will try to develop their personality as they will be aware about their strengths and weaknesses to accomplish a healthy and successful life.

5 Supervised Learning with Regression Analysis Using Python Language

Python is a free and open source, simple and interpreted, object-oriented, and extensible language. Python uses NumPy and Pandas, Scikit-learn, Model Evaluation and Validation, Training and Testing, and Metrics for evaluation in the process of data analysis and training. It serves several benefits as most appropriate for machine learning as open source–open software license with GPL compatible. It is proved highly versatile with Scikit-learn machine learning library. It uses Pandas for data munging, NumPy for data representation, matplotlib for visualization, and Django for web application integration. It is faster and easier to develop prototypes. Python language is supported by for deep learning framework by Google (<https://www.tensorflow.org/>), and is used in Hadoop streaming in many industries.

Supervised learning [16] is used to train the model with some examples of a pair of input and output. Supervised learning algorithms use the training set, which is a set of data used to train the model. On the basis of the training set of data and learning of model, the result for other inputs can be inferred. The testing data set is used to verify the performance of the model or the accuracy of model. The test data set is some part of initial data set as the output will already be available so that we will have to compare the original output with inferred output to check the accuracy of model (Figs. 2 and 3).

Regression analysis and its Gantt Chart (Fig. 4a, b) are a set of statistical processes, which are used for estimating the relationship among variables. It provides many techniques for modeling and analyzing several variables. Regression analysis mainly focuses on the relationship between dependent variable and one or more independent variable (or we can call independent variable as predictors). It mainly focuses to understand how the typical value of the criterion variable (dependent variable) changes when any one of the predictors is varied and other predictors are fixed. Prediction and forecasting are the most common application of regression analysis [17]. The following are the various regression techniques:

Linear Regression, Multiple Regression, Decision tree Regression, and Random Forest Regression.

The following steps are taken in the process of regression analysis:

1. **Raw Dataset:** Once the raw dataset is collected and the data is refined for the analysis. Then, the data is partitioned into train dataset and test dataset. The data set contains the column of input variables and output variables, which is used to train the model for predicting the result using regression analysis techniques or

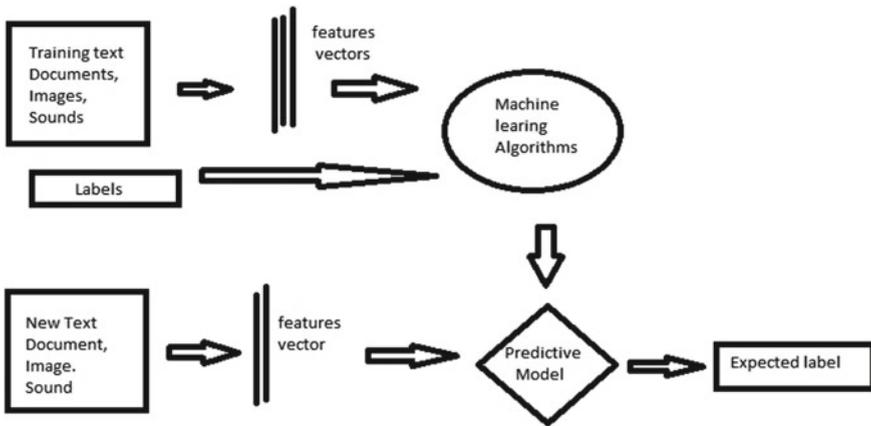


Fig. 2 Methodology of supervised learning [22]

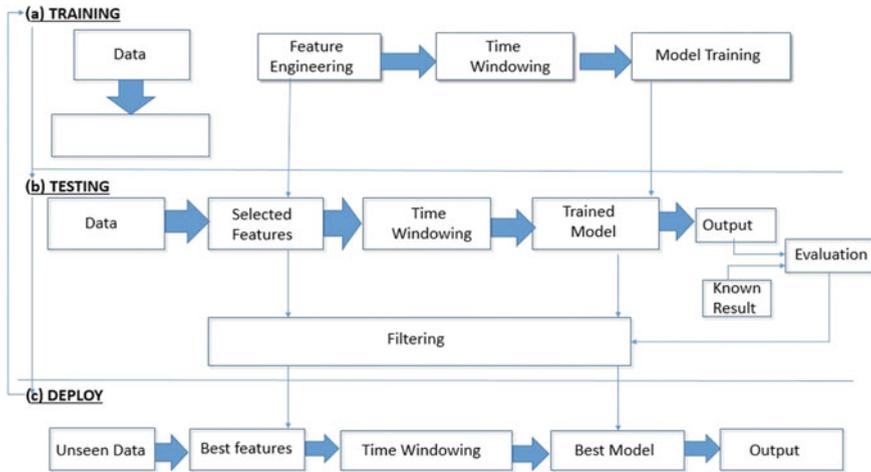
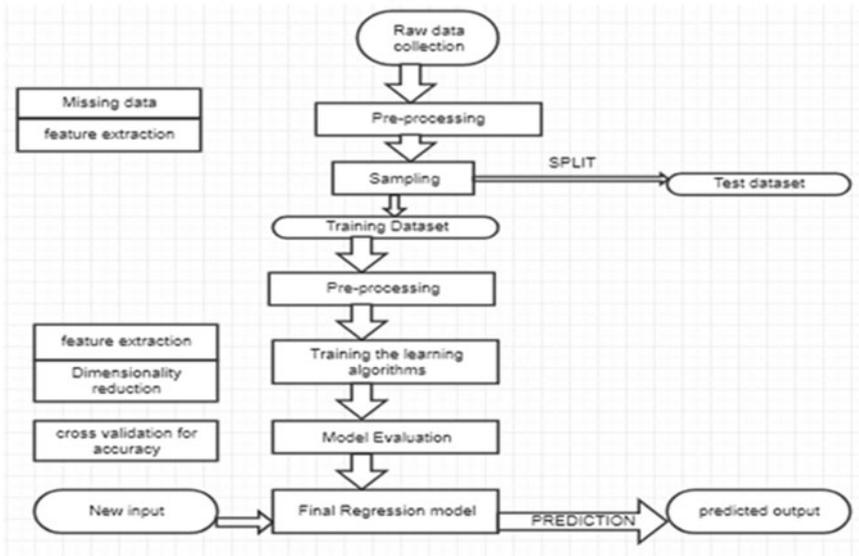


Fig. 3 Iterative process of constructing and applying ML-based prediction models [23]

regression algorithms. Here, the entire dataset is into 80:20 ratio, where training is done on 80% data and cross-validation is done on 20% data.

2. Train the Model: Here, data processing is done. Train data is used to train different models. The best model is chosen among them. Data processing involves removal of nonnumeric data columns, NA, and NAN terms from the data frame. It is done either by replacing them with 0 or replacing the attribute value with its mean or median value. Then, individual models are built using different algorithms from the packages like Caret, rpart, etc., from R and Scikit, sklearn from Python.
3. Get Best Model: After training the model using different regression algorithms, the comparison between those algorithms is done by finding the accuracy of each

(a)



(b)

	Task	Start	End	Dur	%	2017		2018		
						Q4	Q1	Q2	Q3	Q4
	Project on Personality Detection	5/11/17	18/10/18	240		[Gantt bar spanning Q4 2017 to Q4 2018]				
1	Data Study and Research	5/11/17	2/3/18	80		[Gantt bar from 5/11/17 to 2/3/18]				
2	Fetching Dataset	15/1/18	23/3/18	48		[Gantt bar from 15/1/18 to 23/3/18]				
3	Algorithm Research	15/2/18	15/4/18	41		[Gantt bar from 15/2/18 to 15/4/18]				
4	Implementation	16/4/18	15/6/18	44		[Gantt bar from 16/4/18 to 15/6/18]				
5	Performance Analysis and Reporting	16/6/18	18/10/18	86		[Gantt bar from 16/6/18 to 18/10/18]				

Fig. 4 a Flowchart of regression analysis. b Gantt chart of activity

model using test data set for validation and calculating the accuracies of each model. After comparing the accuracies of each model and limitations, the best model is selected for further predictions.

4. Get prediction of the new dataset: Once the most accurate model is selected and trained, the new data set can be applied to the model as input and the output can be predicted using trained model and machine learning algorithms.

6 Study Plot

The dataset consists of the Humor Style Questionnaire [18]. Dimensions of data are 39 attributes and 1071 instances. There are no missing values in the dataset. Now, while exploring the dataset

- Step 1** Apply data ETL operations
i.e., Apply data feature extraction, Data Transformation, Loading of dataset. [Loop Begins] Till all the dataset is completed.
- Step 2** Estimation of summary of each attribute including max, min, mean, count value and also some percentiles.
- Step 3** Check for any null value.
- Step 4** Check the data type of various attributes in dataset.
[End of Loop].

For data visualization first, we start with univariate plots that is, the plot of each variable to get an idea of the distribution (Fig. 5).

7 Algorithm(s) Implementation Results—Graphs and Tables

The graphs above (Figs. 6, 7, 8, and 9) show the range of different personalities among individuals. The use of machine language IOT has helped to detect the different personalities among the people. For detection of each personality, separate graphs are used, hence the data obtained are large and these tremendous amounts of data can be handled by the help of big data only.

The next step is data preprocessing where we initially declare dependent and independent variables. Splitting of the dataset into Training set and Test set. Dataset is split in ratio of 80:20. In other words, 80% of the data is kept in training set and 20% data is kept in test set. Now, we have a training data in X_train and Y_train and test data in X_test and Y_test. We are building some ML models by establishing the correlation between dependent variable and independent variable and once the correlation estimated then, we will test the model using test data (X_test and Y_test) that whether it can understand the correlation between the training set on test set.

Prediction is an action performed on the test data to forecast the values that up to which extent model has learned from the trained data. Predicted values help in estimating the degree that how much accurate model has learned and able to predict the values (Figs. 6, 7, 8, and 9).

Building of models enables us to know that which model is best for learning to perform further computations. As different regression models like Linear Regression (by Eq. 1), Multiple Linear Regression, Decision Tree Regression, and Random Forest Regression. Machine learning regressors usually support a single dependent variable. So, here we create multi-output regressors for Multiple Linear Regression

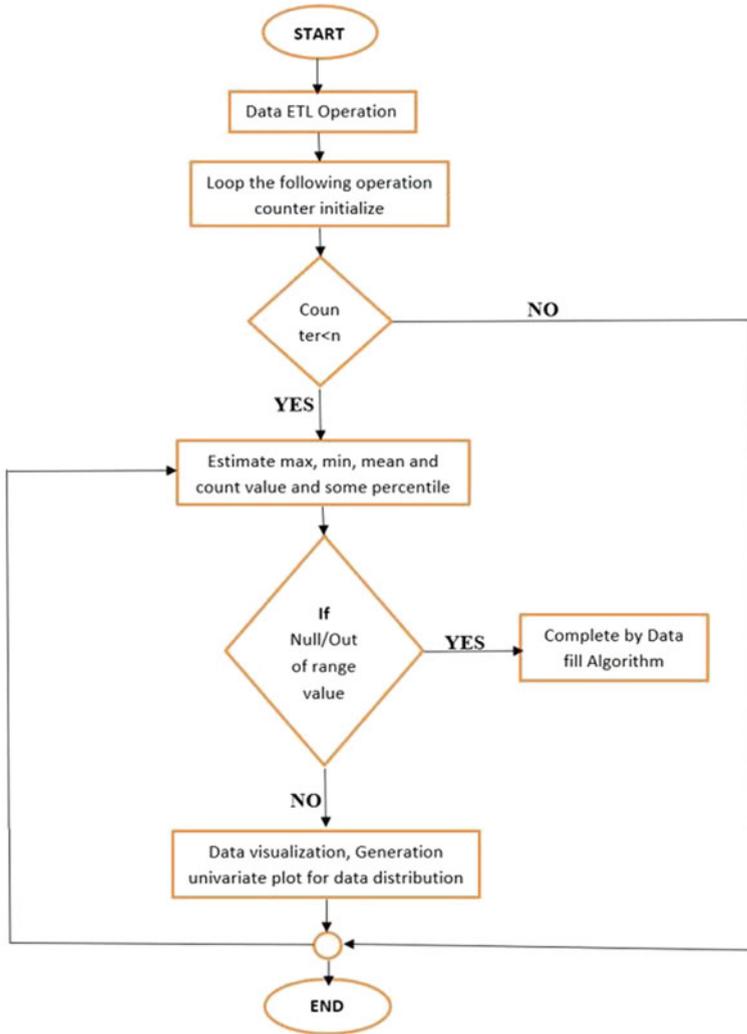


Fig. 5 The flowchart of algorithm applied

(by Eq. 2), Decision Tree Regression, and Random Forest Regression. This strategy consists of fitting one regressor for each target. Prediction and accuracy of multi output regressor for each model is computed. The graphs below (Figs. 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, and 22) gives the visualization of predicted and tested values and shows that predicted values are how much close to tested values. Since accuracy is the most important parameter to decide that which regression technique is best and the accuracy will be achieved only if these large amounts of data, i.e., big data are processed and stored properly (Figs. 23, 24, 25, and 26).

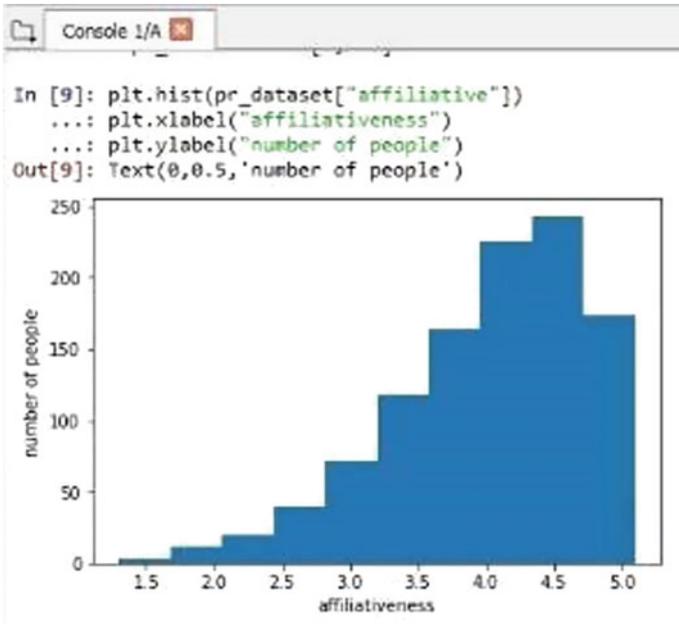


Fig. 6 It shows range of affiliative personality among individuals

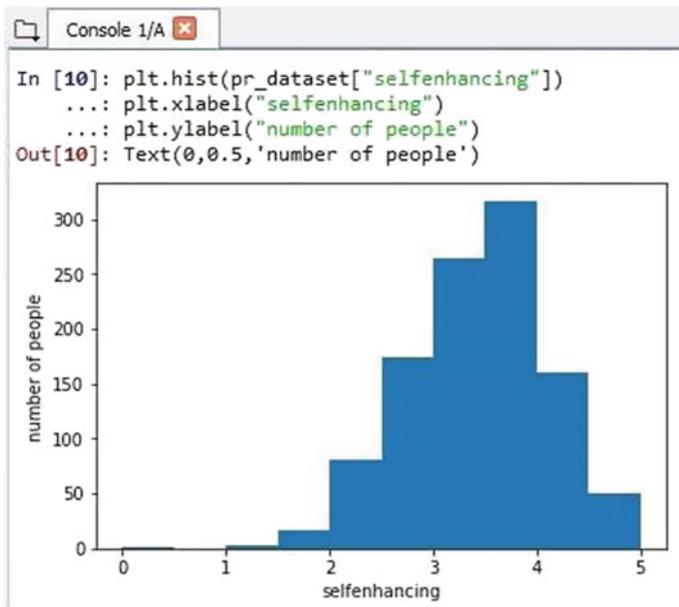


Fig. 7 It shows the range of self-enhancing personality among individuals



Fig. 8 It shows the range of aggressiveness personality among individuals

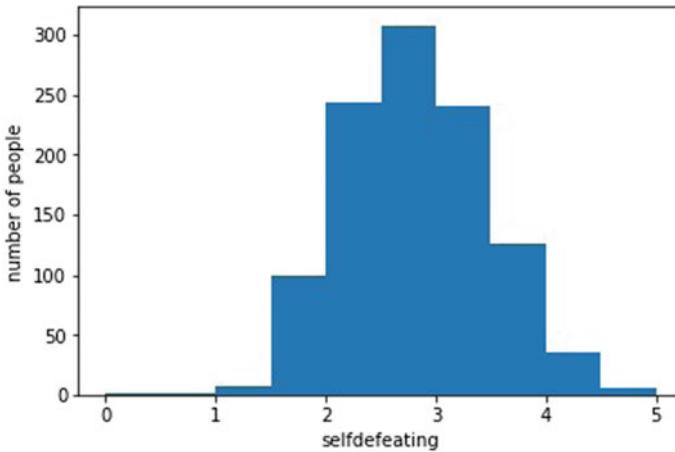


Fig. 9 It shows range of self-defeating personality among individuals

Decision Tree Regression2 and Decision Tree Regression4 achieve almost same accuracy that is, near to 70%.

Least accurate is Decision Tree Regression3 of 28.75% accuracy. Average accuracy of Decision Tree Regression is 64.175%. This method doesn't give the best result. Many times, IOT device collects the data but the data are very complex so it is not able to give good result by this regressions.

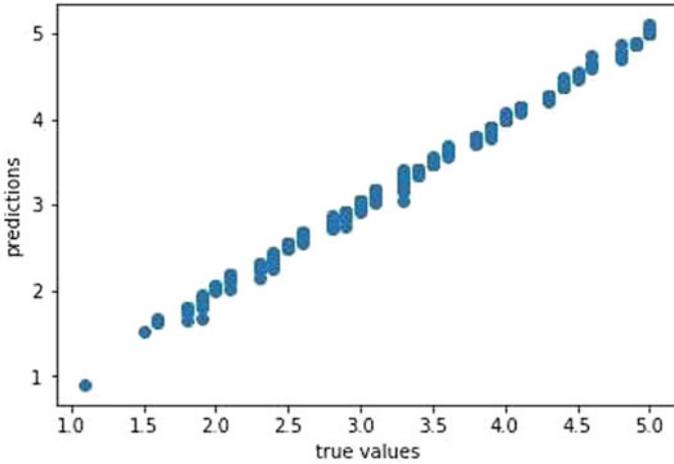


Fig. 10 This graph shows the relationship between the tested and predicted values computed from Linear Regression model. As we can see in the graph, blue color dots are coinciding with each other and are not scattered in the whole plane because the predicted values are very close to the tested values. Accuracy achieved is 99.56% with mean square regression loss is 0.00156390272662

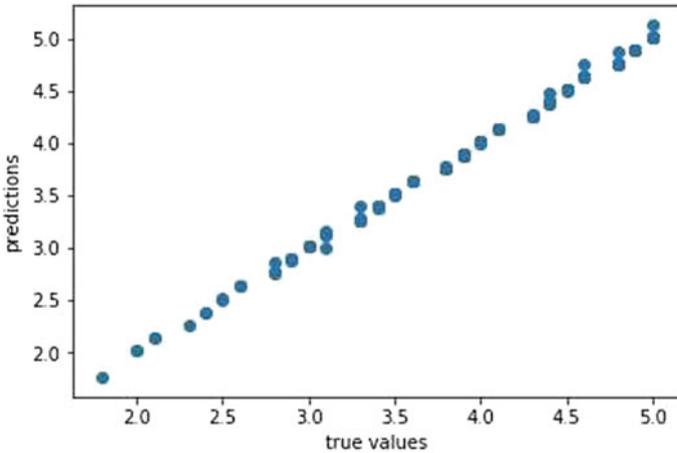


Fig. 11 This graph shows the relationship between the tested and predicted values of Multiple Linear Regression1. The dots are lying very close to each other and are also not scattered in whole plane so it shows that predicted values are very much accurate with accuracy 99.79%. Mean square regression loss is 0.00111408573375

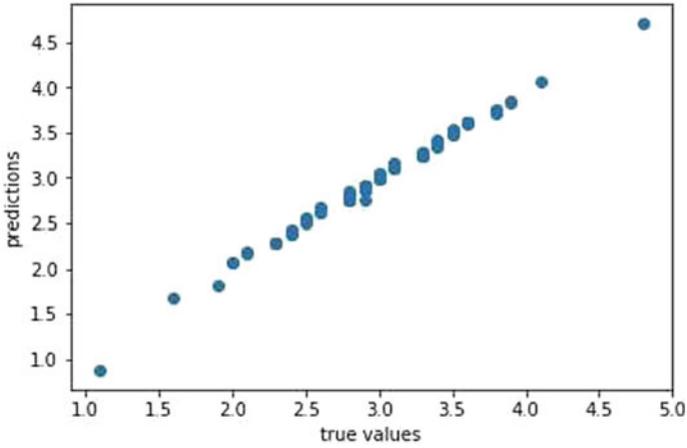


Fig. 12 Visualization of predicted and tested values of Multiple Linear Regression2. This model also predicts the accurate results and this is clearly visible through scattering of points linearly in a plane. Accuracy is 99.51% and mean square regression loss is 0.00169502757472

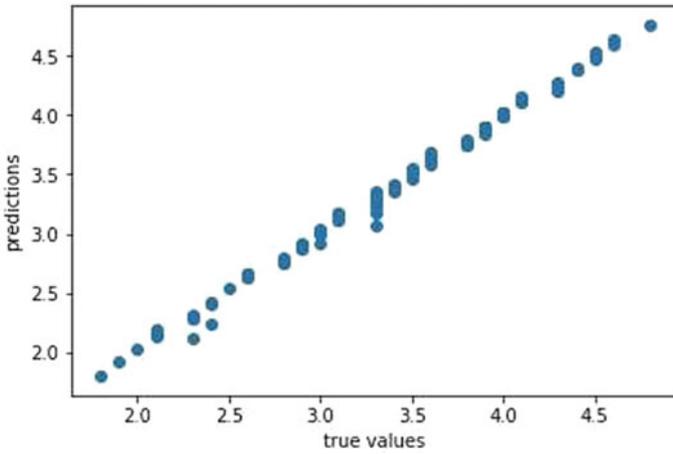


Fig. 13 Visualization of predicted and tested values of Multiple Linear Regression3. This model also predicts the accurate result but in comparison to Multiple Linear Regression1 and Multiple Linear Regression2m there is minor difference in accuracy. Accuracy is 99.14% and mean square regression loss 0.00159577024357

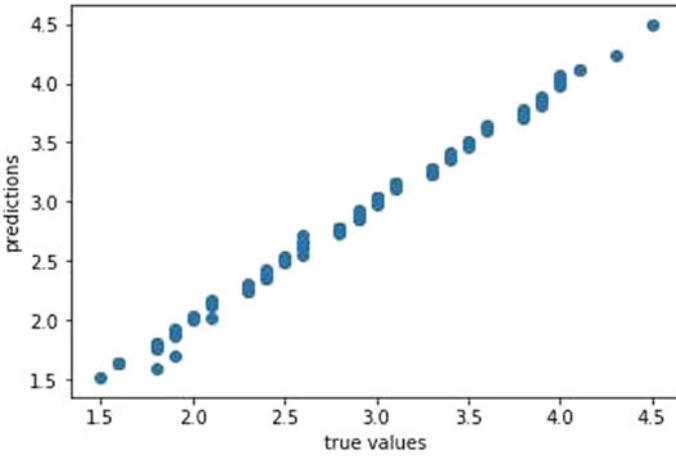


Fig. 14 Visualization of predicted and tested values of Multiple Linear Regression4. This model predicts the accurate results with accuracy of 99.56% and mean square regression loss is 0.00159151039372

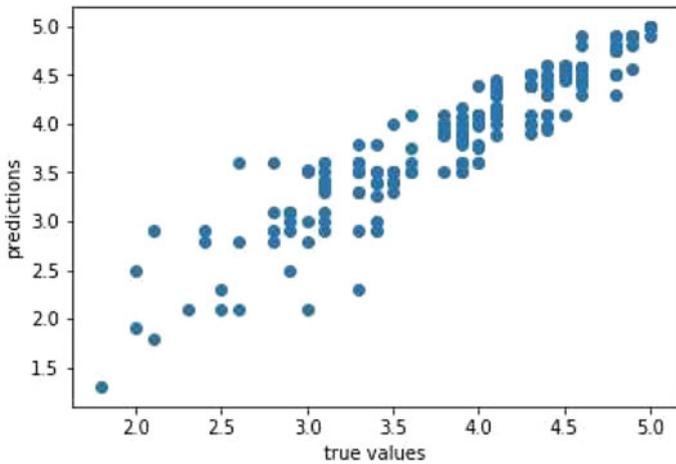


Fig. 15 Visualization of predicted and tested values of Decision Tree Regression1. As we can see in the graphm all points are scattered in a plane because the prediction is not much accurate as in case of Linear regression and Multiple Linear Regression. Scattering shows that predicted and test values are not lying much close to each other. Accuracy is 87.52% and mean square regression loss is 0.0675178855488

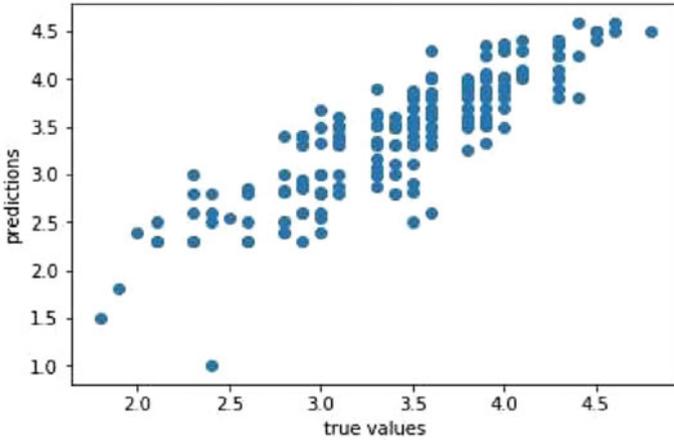


Fig. 16 Visualization of predicted and tested values of Decision Tree Regression2. In this graph, the points are much scattered initially that is, lying at some distance. It is a good model but not accurate model as in comparison to Decision Tree Regression1. Accuracy is 71.48% and mean square regression loss is 0.0984915683583

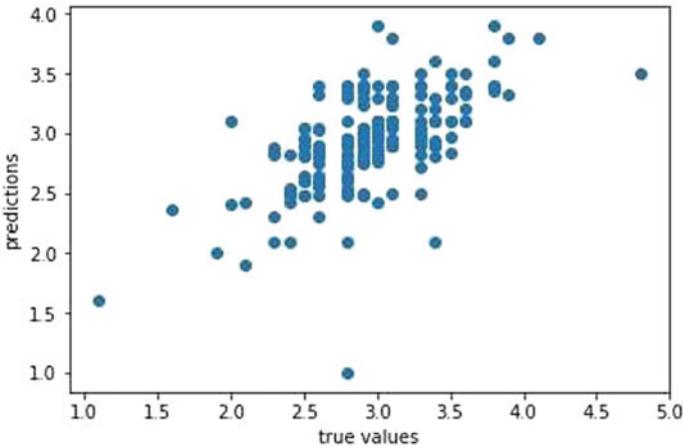


Fig. 17 Visualization of predicted and tested values of Decision Tree Regression3. This model is least accurate as we can see in the graph that points are clustered at various places and are not scattered linearly. This is model has very less accuracy. Accuracy is 28.75% and mean square regression loss 0.132380568167

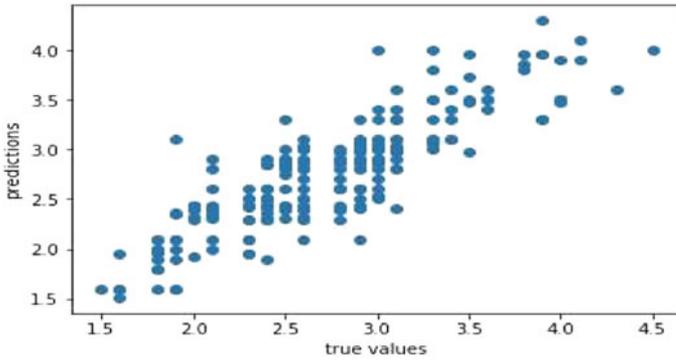


Fig. 18 Visualization of predicted and tested values of Decision Tree Regression4. This is far better than Decision Tree Regression3 as we can see that points are scattered in the linear form. This is very much similar to Decision Tree Regression2 in terms of prediction. Accuracy is 72.84% and mean square regression loss is 0.0982561775804

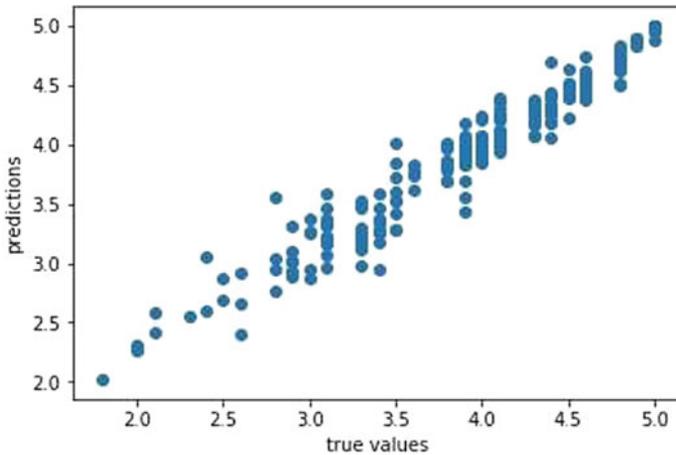


Fig. 19 Visualization of predicted and tested values of Random ForestRegression1. This is accurate model as in comparison to Decision Tree Regression where average accuracy is 65.1475%. All points are scattered linearly in a plane. So, better predictions are obtained. Accuracy is 94.16% and mean square regression loss is 0.0315995487491

8 Performance Evaluation, Learning and Outcomes, Product Perspective

Among the different algorithms used in the project, Linear Regression and Multiple Linear Regression are proved to be the best. Both of them are the best algorithm that can be used to implement the prediction of personality of individuals. On the

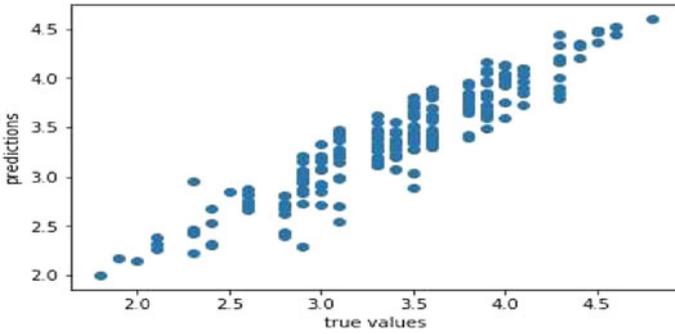


Fig. 20 Visualization of predicted and tested values of Random Forest Regression2. This model predicts the results but with less accuracy of 88.0% in comparison to Random Forest Regression1. Mean square regression loss is 0.0315995487491

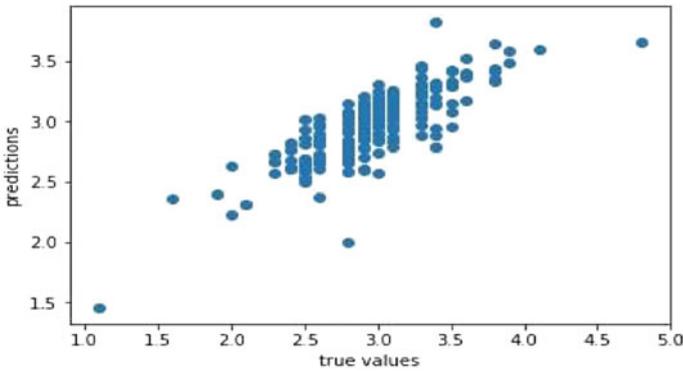


Fig. 21 Visualization of predicted and tested values of Random Forest Regression3. This is model has comparatively less accuracy to other Random Forest Regressors. Accuracy is 64.84% and mean square regression loss 0.065326862103

basis of different datasets used, different algorithms are required to be implemented accordingly. This is because the calculated root mean square error value was the minimum, i.e., 0.00156 and 0.0016.

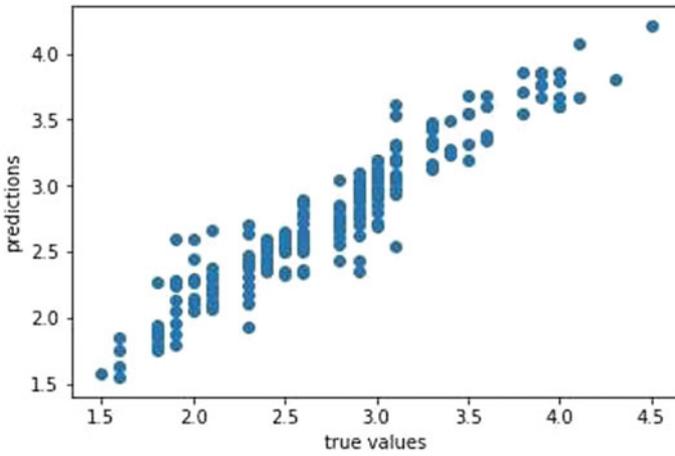


Fig. 22 Visualization of predicted and tested values of Random Forest Regression4. This model is performing predictions in same manner as Random Forest Regression2. Graph is almost the same for both the regressors. Accuracy is 89.35% and mean square regression loss 0.0385353877922

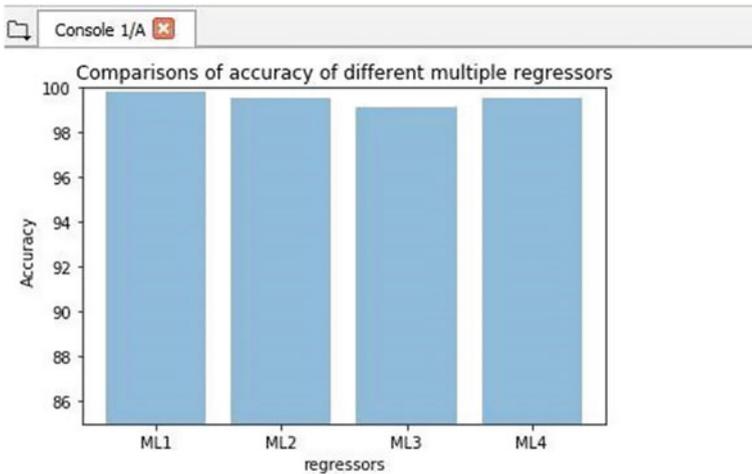


Fig. 23 Comparison of accuracy of different multiple regressions. This graph represents the comparison among the accuracies of four regressors of Multiple Linear Regression. All the regressors are computing almost the same results as there is not much difference in their accuracies. But the highest accuracy is of Multiple Linear Regression1 that is, 99.79%. Average accuracy of multiple Linear Regression is 99.5%. By this accurate result, the model is very impressive and this method compiles all the data that are collected by IOT device, and gives good result

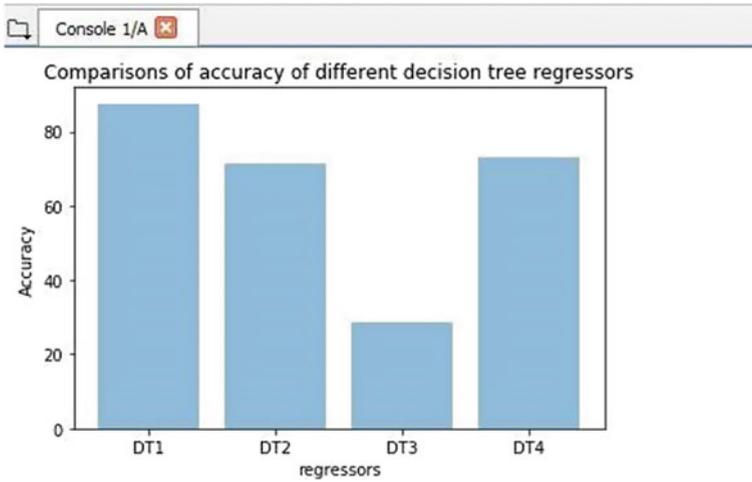


Fig. 24 Comparison of accuracy of different decision tree regressions. As it is clearly visible through the graph that Decision Tree Regression1 is having the highest accuracy that is, 87.52% greatest among all the regressors

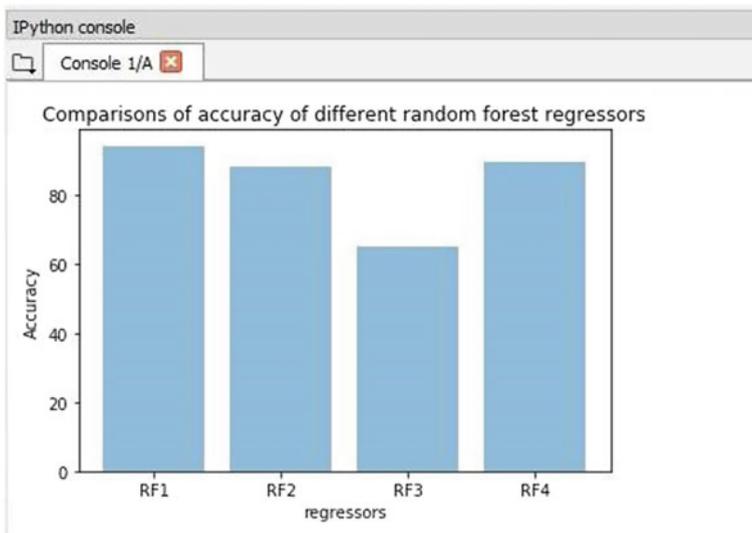


Fig. 25 Comparison of accuracy of different Random Forest Regressions. As it is clearly visible through the graph that Random Forest Regression1 is having the highest accuracy among all the regressors that is, 94.16%. Random Forest Regression2 and Random Forest Regression4 achieve almost the same accuracy but Regression3 with accuracy of 64.84%. The comparison between all the regressors that we have got in this figure using the concept of IOT. Since the data processed during regression are very large, so the compared data of different regressors are in tremendous amount which can be processed and stored using big data

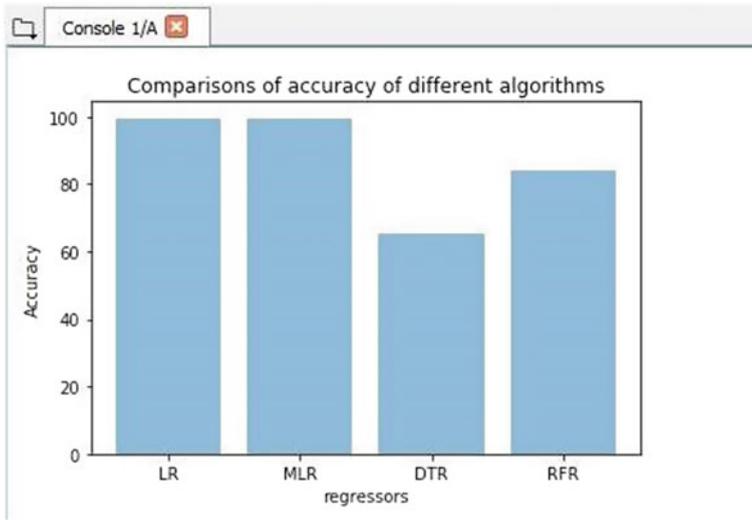


Fig. 26 Comparison Graph of Accuracies of Algorithms. This graph represents the performance of all the models in predicting the values in form of accuracy. As graph depicts that Linear Regression and Multiple Linear Regression attained almost the same accuracy. Random Forest Regression has the accuracy of 84.0875%. Least accuracy is estimated by Decision Tree Regression that is, 64.175%. Accuracy is very important because accuracy gives minimum error. Prediction of personality is accurately achieved by multiple linear regressions as it can handle the large amount of data, i.e., big data perfectly

As we can see in the graph below that decision tree regression model has achieved minimum accuracy as in comparison to others so it is not the model, which can be used for training our model for determining the personality traits of individuals. For training our model, decision tree is not efficient as more the number of decisions, less the expected value of the accuracy.

Accuracy Percentage

Linear Regression = 99.56%. Multiple Linear Regression = 99.5%.

Decision Tree Regression = 64.175%. Random Forest Regression = 84.0875%. Learned Python, work under pressure and teamwork. Among the different regression algorithms, Linear Regression and Multiple Linear Regression are best in order to perform the personality detection test. Linear Regression will be used for performing the computations as it performs prediction with 99.56% accuracy.

The framework is based on a report named "Individual differences in sense of humor and their relation with psychological well-being". In this, the author will develop a questionnaire. These questionnaires will be used as a source for predicting the personality of an individual in terms of four personality traits (aggressive, affiliative, self-enhancing, and self-defeating). Every participant or user will have to give answers and on the basis of that answers, the range of each personality trait in the range of 0–5 will be calculated.

9 Future Directions, Novelty, and Complexity Analysis

Since the personality of an individual can adversely affect the professional attitude of a person. All the MNCs and even our defense organization select candidate for job on the basis of their personality, therefore, they use some procedure to predict the personality of an individual, which can be done through machine learning and take IOT data and model result to select the perfect candidate. This experimental study can be further modified to be used in these organizations to get more efficient prediction.

Different disciplines and backgrounds of computer science engineering with psychology have been merged to perform the work. Specifically, the author aims to predict the personality traits of an individual as the personal and professional life of an individual.

The analysis is performed on the personality detection to get the domination of some particular personality traits in male or female [19]. The prediction of personality has been done through k-nearest neighbor classification technique [20]. Personality detection is less efficient and costly manually so the author has tried to do these jobs by machine learning. Regression techniques used for the prediction are linear regression, decision tree, and random forest. An individual can use the predicted result to develop their personality and improve their personal and professional life.

The work has been done by merging mathematical algorithms and the psychology and behavior of human being. Psychologists seek to measure the personality by analyzing human behavior. This process of analyzing the human behavior can be done mathematically. The author has used the basic analysis of human behavior to prepare the questionnaire machine learning algorithms to train the model using the collected dataset. The author has converted the subjective concept into some quantitative approach. This analysis can be proved very much beneficial to improve the personality to accomplish healthy and successful life.

The personality is something, which describes your actions and reactions toward in any situations or how you will react in particular circumstances. Prediction of personality traits is not much accurate when performed manually and it is a complex process too. Using machine learning subjective work can now be performed using mathematical models. These models are trained with experiences and could easily perform a personality test, which is much more accurate and fast as in comparison to manual task.

The human behavior is very different to analyze as human mind along with human nature is very complex. So, we need some technology, i.e., IOT which may analyze the person's personality from very near. The calculated attributes are subjective and vary from person to person. Also, it is so complex phenomenon that it can be treated as NP-complete program. The run-time complexity will be high in the program as the human behavior and thinking are very complex to convert it into some quantitative calculations.

10 Recommendations, Limitations, Future Scope, and Possible Applications

The given program or method is used to predict the personality of an individual in terms of four personality traits (aggressive, affiliative, self-enhancing, and self-defeating) where we need some technology to handle the big data because all four personality traits generate abundant data. For processing and storage of data, we use data processing application software. As the personality of an individual can directly influence the social, personal, and professional life of an individual. The result of this technology helps in analyzing the personality traits of an individual and with the help of this result, the user can develop his or her personality as they will be aware of their weaknesses and strengths. IOT gives every information about the personality of a person and may predict that what will be the need of person and constantly watch their behavior, executes the data by method of multiple regression and with time IOT makes them smarter.

The method used by the authors is a new way for computing personality detection test. Machine learning is something which makes the human work easier and gives more efficient output. As it is a well-known fact that decision-making processes of human beings are easily affected by the activities and up and downs of human life. Process of personality test is automated in order to reduce the human error and it's done by IOT device, which always connects the person with Internet. This is one of the reasons that this method is much accurate and preferable in order to predict the personality of an individual because personality data is very large and by using Big Data concept processing of this type of large data has become easy and it gives an accurate result. These tests are better to understand the perception of an individual for self-reflection and understanding. The result of this technology is helpful to decide the one's personality and take important decisions in the field of job placement, group interaction, drive to learn, and team spirit.

The framework can work for short number of parameters. The personality of an individual is predicted on the basis of answers collected from the user. The questionnaire is developed on the basis of difference in someone's sense of humor and usage of humor, but there are so many factors which can influence the personality, which is determined by IOT therefore to obtain the accurate and more precise result, maximum factors should be covered during the development of questionnaire and these data are processed using big data and gives more precise result.

The model used for personality detection is Linear Regression. But Linear Regression is prone to overfitting such that regression begins to model noise in the data along with the relationship of the variables. This is pretty extrapolation and not necessarily accurate. Decision Tree Regression is not estimating the good results so we can't use this for learning and prediction of personality traits.

Explore yourself from a completely different perspective. To find a successful career and improve social and personal life by understanding different personality traits.

But this is possible when connectivity will improve, data collection will manage and differentiate for different parameter. Internet of Things will be smarter than human in the future but for protection, we need some organization to monitor all data process and identify unethical action.

Collection of data will be larger in the future so if anyone manipulates, then IOT will not show good result.

To overcome this vulnerability, we need to make some hardware and software to secure the result and behavior of modal.

Interviews in corporate, during hiring procedure and ever for selecting person in companies, for example, in agile software development methodology, it is very important to find the right person for the particular job.

Personality tests in government organizations, As the jobs like them demands different characteristics and personality of and individual for example the ORQ (officer like qualities are needed in defense some special characteristics are needed in IAS officer, that can differ from the characteristics needed for an IPS officer so in order to find right candidate the personality detection can be a great help.

Can help to the aspirants of such jobs which need good personality. And same data transfer to other company, organization or government bodies to select good candidate. So here role of big data will be increase. So, they can prepare by working on their personality as, if they will know the problem than they can easily find the solution [21].

This method can be used by the user to develop their personality. This method can be used in companies also to give appropriate work to appropriate employee according to their personality (owadays, it is used in agile methodology of software development where the attitude of an employee plays a very important role). This program can be used in selection in defense, as the main requirement in defense is attitude and personality. This program can be used by the aspirants also to prepare for particular jobs or exams.

11 Motivation

Personality is something that explains the behavior of an individual in different circumstances. If the people will be aware of their personality than they can improve their personality, if they want as if we don't know the disease than how can we find the accurate medicine to cure the disease. Not only for an individual personality can detection be used by companies to hire the right person for the right job. As each job require demands, some unique characteristics to find the person who holds them can be found by the procedure of personality detection.

Machine learning is getting popular day by day and enable the computers to perform computations without being explicitly programmed. Science has made our life simpler and proved to be boon for solving many complex problems. Many researches in personality ranges from the study of biological systems to genetic codes in order to study social, cultural basis of thought, emotions, feelings, and behavior. Issues of

personality appeal to everyone and there are many interesting web sites that can be visited both inside and outside of academia. So, this is basically an approach that can be used by any organization during the personality test conducted during the selection process of any employee. A personality test can give the different characteristics or features which were not known to us.

Today, everyone wants a better life but instead of focusing on their weakness and strength, they criticize others and system. Therefore, we need to improve ourselves, and this test can be proved useful for personality development of an individual. Interviews in corporate, during hiring procedure and ever for selecting person in companies, for example, in agile software development methodology, it is very important to find the right person for the particular job.

Personality tests in government organizations, As the jobs like them demands different characteristics and personality of and individual for example the ORQ (officer like qualities are needed in defense some special characteristics are needed in IAS officer, that can differ from the characteristics needed for an IPS officer so in order to find right candidate the personality detection can be a great help.

It can help to the aspirants of such jobs which need good personality. So, they can prepare by working on their personality as, if they will know the problem than they can easily find the solution.

12 Conclusion

The proposed work will be an effort to utilize the machine intelligent techniques like Python programming on Spyder framework. Human beings are complex to understand and their personality traits are changed according to the changes in them. So, we will try to learn more and more about them.

To predict the personality traits of an individual here, the author has implemented a method that involves the supervised learning. Supervised learning has a subset of algorithms called regressions. These regressors have provided us with the models which are trained for learning. Models learned from experiences that is, 80% of train data and the remaining 20% are used to test the accuracy of prediction. Among the four models, Linear Regression and Multiple Linear Regression have attained the best prediction accuracy of 95% about.

Decision Tree Regression has a prediction accuracy of 64.175% which is very less. So, this model can't be used for predicting the personality traits using dataset available with us. Random Forest Regression attained the accuracy of 84.0875%, which is better than Decision Tree Regression it can also be used as a secondary model to predict the personality traits.

The difference in an individual's sense of humor can influence the personality trait of an individual. Therefore, the sense of humor and usage of humor can be used to predict the personality of an individual. The personality of an individual affects their life, therefore, it is very important to develop their personality and anyone can improve their personality only when they are aware about their weakness and

strengths. So, to know the weakness and strengths of an individual can be known by analyzing the personality traits of an individual.

The proposed work is a successful mathematical framework to calculate or predict human personality traits. Which can help the user to develop their personality to achieve a healthy and successful life?

This method can be used by the user to develop their personality. This method can be used in companies also to give appropriate work to appropriate employee according to their personality (Nowadays, it is used in agile methodology of software development where the attitude of an employee plays a very important role). This program can be used in selection in defense, as the main requirement in defense is attitude and personality. This program can be used by the aspirants also to prepare for particular jobs or exams.

Acknowledgements The authors are obliged for the guidance of ABES Engineering College staffs and faculties. Mr. Shubham Sidana, Mr. Abhishek Goyal, Mr. Mayank Gupta of TCS, and evaluation team of experimental presentation to understand the concept well and for showing the path ahead. The acknowledgement is to all those forces which inspired us to work hard and to learn something and to make a difference in society.

References

1. A. Kumari, S. Tanwar, S. Tyagi, N. Kumar, Fog computing for healthcare 4.0 environment: opportunities and challenges. *Comput. Electr. Eng.* **72**, 1–13 (2018)
2. A. Kumari, S. Tanwar, S. Tyagi, N. Kumar, M. Maasberg, K.K.R. Choo, Multimedia big data computing and Internet of Things applications: a taxonomy and process model. *J. Netw. Comput. Appl.* **124**, 169–195 (2018)
3. K.H.K. Reddy, H. Das, D.S. Roy, *A Data Aware Scheme for Scheduling Big-Data Applications with SAVANNA Hadoop*. Futures of Network (CRC Press, 2017)
4. H. Das, A.K. Jena, P.K. Rath, B. Muduli, S.R. Das, Grid computing-based performance analysis of power system: a graph theoretic approach, in *Intelligent Computing, Communication and Devices* (Springer, New Delhi, 2015), pp. 259–266
5. B.B. Mishra, S. Dehuri, B.K. Panigrahi, A.K. Nayak, B.S.P. Mishra, H. Das, *Computational Intelligence in Sensor Networks*. Studies in Computational Intelligence, vol. 776 (Springer, 2018)
6. S. Tanwar, P. Patel, K. Patel, S. Tyagi, N. Kumar, M.S. Obaidat, An advanced internet of thing based security alert system for smart home, in *International Conference on Computer, Information and Telecommunication Systems (IEEE CITS-2017)*, Dalian University, Dalian, China, 21–23 July 2017, pp. 25–29
7. A. Srivastava, S.K. Singh, S. Tanwar, S. Tyagi, Suitability of big data analytics in indian banking sector to increase revenue and profitability, in *3rd International Conference on Advances in Computing, Communication & Automation (ICACCA-2017)*, Tula Institute, Dehradun, UK, pp. 1–6
8. I. Kar, R.R. Parida, H. Das, Energy aware scheduling using genetic algorithm in cloud data centers, in *International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, March 2016 (IEEE, 2016), pp. 3545–3550
9. J. Villena-Roman, J. Garcia-Morera, C. Moreno-Garcia, L. Ferrer-Urefia, S. Lana-Serrano, J. Carlos Gonzalez-Cristobal, A. Waterski, E. Martinez-Camara, M.A. Umbreras, N1.T. Martin-Valdivia, L. Alfonso Urefia-Lopez, *TASS-Workshop on Sentiment Analysis at SEPLN, Workshop on Sentiment Analysis*. Sociedad Espanola para el Procesamiento del Lenguaje (2012)

10. M. De Juan-Espinosa, *Personalidad Artificial: Hacia Una Simulación De Las Diferencias De Personalidad En Situaciones De Interacción* (Universidad Autónoma de Madrid, Madrid, 2007)
11. A. Bandura, R.H. Walters, A. Riviere, *Aprendizaje social y desarrollo de la personalidad*. Alianza Editorial Sa (2007)
12. N.D. Liver good, *Strategic Personality Simulation: A New Strategic Concept*. Center for Strategic Leadership, US Army War College, vol. 3 (1995), pp. 60–93
13. L.J. Francis, L.B. Brown, R. Philipchalk, The development of an abbreviated form of the Revised Eysenck Personality Questionnaire (EPQR-A): its use among students in England, Canada, the USA and Australia. *Pers. Individ. Differ.* **13**(4), 443–449 (1992)
14. G. Young, Mapping mayhem: the geography of crime. *Computer Age*. Google Scholar **171**(107) (2003)
15. D.K. Rossmo, *Geographic Profiling*, vol. 1 (CRC Press, 1999). ISBN-0849381290
16. P. Sarkhel, H. Das, L.K. Vashishtha, Task-scheduling algorithms in cloud environment, in *Computational Intelligence in Data Mining* (Springer, Singapore, 2017), pp. 553–562
17. H. Das, A.K. Jena, J. Nayak, B. Naik, H.S. Behera, A novel PSO based back propagation learning-MLP (PSO-BP-MLP) for classification, in *Computational Intelligence in Data Mining*, vol. 2 (Springer, New Delhi, 2015), pp. 461–471
18. T. Polzehl, S. Möller, F. Metzke, Automatically assessing acoustic manifestations of personality in speech, in *Spoken Language Technology Workshop (SLT)*, December 2010 (IEEE, 2010), pp. 7–12
19. L.R. Goldberg, Language and individual differences: the search for universals in personality lexicons. *Rev. Pers. Soc. Psychol.* **2**(1), 141–165 (2007)
20. A.V. Ivanov, G. Riccardi, A.J. Sporka, J. Franc, Recognition of personality traits from human spoken conversations, in *Twelfth Annual Conference of the International Speech Communication Association* (2011)
21. A.J. Gill, J. Oberlander, Taking care of the linguistic features of extraversion, in *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 24, no. 24, January 2002
22. M. De Juan, *Personalidad y Criminalidad*, Apuntes de Psicología Criminológica. la asignatura Psicología Criminológica, Universidad Autónoma de Madrid (No publicado) (2005)
23. P.T. Costa, R.R. McCrea, *Revised neo personality inventory (NEO PI-R) and neo five-factor inventory (NEO-FFI)* (Psychological Assessment Resources, 2012)
24. H.J. Eysenek, H.J. Eysenck, *Dimensions of Personality*, vol. 5 (Transaction Publishers, 1950), pp. 398–423
25. F. Mairesse, M. Walker, Automatic recognition of personality in conversation, in *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers. Association for Computational Linguistics*, June 2006, pp. 85–88
26. J.W.P.M.E. Francis, R.J. Booth, *Linguistic Inquiry and Word Count*. Technical Report (Southern Methodist University, Dallas, TX, 1993)
27. A. Ruiz-Falcó, Instituto de Astrofísica de Andalucía, CSIC. Entiempo real Lenguajes de Descripción Hardware: Conceptos y Perspectivas (126), 35 (2009)
28. G.V. Caprara, C. Barbaranelli, L. Borgogni, M. Perugini, The “Big Five Questionnaire”: a new questionnaire to assess the five-factor model. *Pers. Individ. Differ.* **15**(3), 281–288 (2011)