ISSN No:-2456-2165

Personality Classification with Data Mining

Gaurav R. Savant (M.E. P.T. C.S.E.(, Dr. G. R. Bamnote)Professor)
Department of Computer science and engineering PRMITR, Badnera

Abstract:- Personality of a person decide whether he can play the role of leader, influence people around, master in communication skills, do collaborative work, able to do negotiateiation in business and handle stress. Personality describes human features that decides how people deal with world around. personal behaviour features can be used for deciding a person's personality based on his/her personality traits. Narcissist person is having features like seeking attention, over self-importance, lack of empathy for other people around. Using big five personality traits, whether a person is narcissist can be determine. By knowing the level of narcissism with the combination of five personality traits personality types like modest, semi modest, confident, somewhat overconfident, overconfident can be predicted with APC. Data consists of answers to 40 different questions, along with their scores, which are used to evaluate Narcissistic Personalities. Later when user answers a questionnaire related with big five personality traits, ensemble learner does the prediction and system will show the personality of user. This learning can classify/predict user personality based on past classifications.This system can be use in organizations, agencies where they recruitcandidates depending upon their personality features along with technical knowledge. Thispaperproposes a system which brings out the personality of an candidate. Personality classification refers to the psychological classification of various styles of individual. This project deals with the areas wherever it determines the characteristics of someone. It is often useful to classify person mistreatment temperament classification mistreatment information mining approach. During this paper, aim is to alter the personality prediction of the users by taking a mental test using 40 questions related to five personality traits. The system uses ensemble classification algorithmic rule. The analysis is finished vast knowledge of information in data set and is been compared with the user input. This paper in the main focuses on multi classification.

Keywords:- *Personality classification, Narcissism, APC.*

I. INTRODUCTION

To identify the personality of a person by observing their nature is an old technique. This was a manual process to predict the nature of the person. Data mining is prominently used today by companies with a strong consumer focus- retail, financial, communication, and marketing organizations. Various ways used to analyse the data are surveys, interviews, questionnaires, classroom activities, shopping website data, social network data about the user experiences and problems they face. These conventional approach requires more time.This system reveals information portraying the personality of the user, based on the answers given by user that are relating to five personality traits. System will compare data stored in the databasewith the results and automatically predict the user's personality. Based on the personality traits of the user, system will provide classes as modest, semi modest, confident, somewhat overconfident, overconfident that describes the user's personality. Personality decides how a person interact with the outer world. To reveal personality of a person by analysing the behaviour of him is a conventional technique. As this is manual method of personality prediction takes lot of time and resources. Analysing personality based on one's nature was a difficult task and much human effort will be required to do manual analysis. Also, this did not provide correct results while analysing the personality of a user from their nature and behaviour. Since analysis was done manually, as humans prone to be prejudice and generally see the things accordingly affects the correctness and thus decreases the accuracy.

II. LITERATURE SURVEY

Aleksandar Kartelj etc. al.]1[said that reliable approaches can be used to classify the personality in various new researches by applying the concept of Automated Personality Classification. Firstly, examined all the possible solutions and what all improvements can be made to the existing problems of Automated Personality Classification. Then considered the extension of the Automated Personality Classification]APC[problem such as the Dynamic APC and how to remove inconsistency in textual data. This entire research was carried out in the context of social networks and related data mining mechanisms**.** Fazel Keshtkar et. al.]2[said that aims of developing methods for modelling student behavior based on data such as online conversations, discussions in class, etc. However, methods like Intelligent Tutoring System)ITS(and Educational Data Mining)EDM(used an individual's behavior and personality for analysis purpose. Thus, a system is developed which can be adjusted by the user and analyze student's behavior during their interaction as well**.** Nurbiha A Shukora etc. al.]7[have given the concept of Online learning which became highly popular because of technological advancement that made it possible to have discussions even from a distance. Most studies that have been conducted report how effective online learning has helped students to improve their learning power while assessing the learning process simultaneously. This kind of discussion can be possible only by applying data mining technique wherein one can assess the different experiences of students which they filled online on the basis of their log files. However, it is suggested by the results that students should put more hard work to became an excellent online learner.]17[A comprehensive investigation of a company's ideal clients is known as a customer personality analysis. Customer personality analysis enables a company to adapt its product depending on the preferences of its target customers from various customer categories.the main

ISSN No:-2456-2165

motive for the paper is to find the accuracy of the prediction of the personality of the customer who is shopping and improve the research out there using the ensemble technique.

III. SYSTEM DESIGN

To solve the issues of the existing system a personality classification system is designed in which some data mining techniques ae used and machine learning algorithms are implemented to classify the personalities of user. It is achieved by using ensemble classifier of different algorithms like KNN, Logistic regression, Deep forest, MLP and Support Vector Machine. By using the previousdata, new techniques can be applied to identify the personality, so that it solves the issues with the existing system. In this proposed system, the user has to answer 40 questions which are relating five personality traits. Here it converts textual answers to features,1 for 'yes' and 2 for 'no'. Based on the obtained results system predict the personality labels asmodest, semi modest, confident, somewhat overconfident, over confident.This system can be usedin many sectors like interview, recruitment process, government sectors, psychometric tests. Once a user personality is revealed then he/she can be hired in any organizationwhere they are allocated with their personality type jobs. The answers chosen by the candidate in the personality questionnairereveal type of the personality the person is having. Once the person knows the personality features of him, he can choose the carrier optionsbest suitable for him. Ensemble classifier is used to improve the test accuracy score instead of a single classifier.Ensemble learning helps to improvise machine learning predictions by combining several models. This approach builds better predictive model compared to a single model.

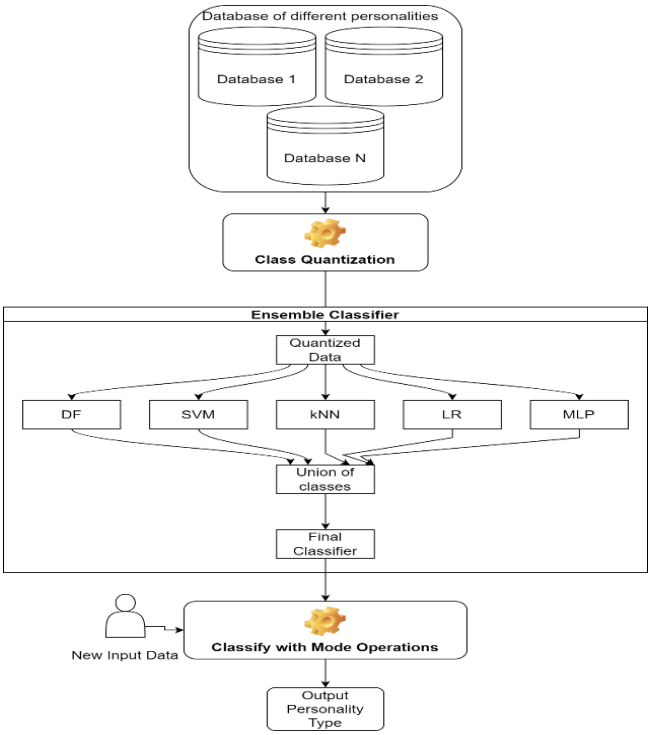


Fig. 1: Algorithm implantation for APC

IV. SYSTEM ARCHITECTURE

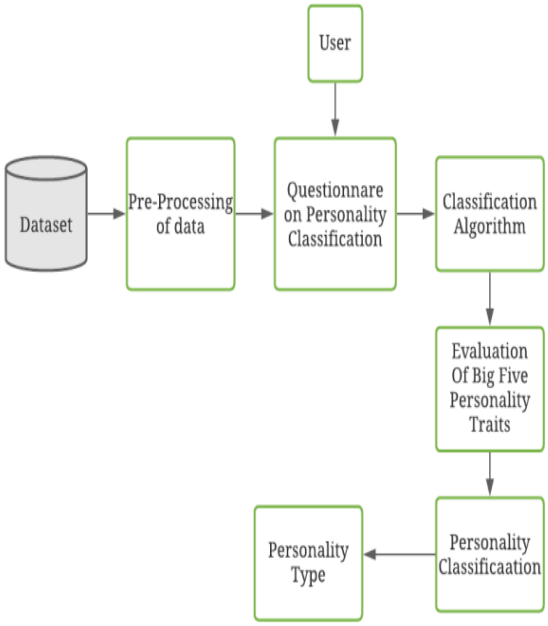


Fig. 2: System Design for Personality Prediction System

The following are the research objectives of the work that focuses on Personality prediction.

- The objective is classifying personalities and to analyse them based on the big five model with a given data set using classification algorithms and advanced data mining concepts.
- Using and exhibiting the data mining concepts and automate personality classification using python data science libraries
- To design a system to improve performance of Multi Class classification in Personality Prediction Analysis.

V. IMPLEMENTATION AND RESULT

Following are the four modules which are divided as for every project it is a necessary to divide the collection of source files required into individual functional unit**.** Every module need to be built independently, tested and debugged**.**

- Data collection
- Attribute selection
- Pre-processing of data
- Prediction of personality

It is always necessary for any prediction system to collect data and take decision about training dataset and test dataset**.** Attributes in dataset are age, gender, answers to 40 questions in **'**yes**'** and **'**no**'**, personality label based on results**. '**elapse**'** attribute removed as it is not contributing towards classification**.**For better accuracy created the ensemble model by combining support vector machine, KNN, logistic regression, MLP and deep forest**.** The accuracy of all the models alone was lesser ascompared to the ensemble model**.**

ISSN No:-2456-2165

The personalities are divided into 5 classes,
- Modest
- Semi Modest
- Confident
- Somewhat overconfident
- Overconfident

These classes were evaluated by combining the score from all questions, and then quantizing it to the scale of 1 to 5
- For example, if the score is 10, then class will be 5*10/40 = 1
- If score is 25, then class will be 5*25/40 = 3

Training was done for Deep Forest, k Nearest Neighbor, Logistic Regression, Multilayer Perceptron and Support Vector Classifiers.

Each classifier used the following configurations**:**
Deep Forest**:** Estimators = 100, Depth = 2
k Nearest Neighbor, Number of Neighbors =1
Logistic Regression, Solver=Limited-memory Broyden Fletcher–Goldfarb–Shanno Model, Total Iterations=1000
Multilayer Perceptron, Number of hidden layers = 100
Support Vector Classifier, Error Tolerance = 10^{-5}

The classifiers were trained, and their classified outputs were combined using Union Based Ensemble Learning Model.

This Model works via the following process,
- Combine all classes from all classifiers
- Identify unique classes
- Check which classes are common between classifiers, and discard them for better results
- Use remaining classes for evaluation of Precision, Recall, fMeasure, and Confusion Matrix for the combined classifiers.

For any new input, classify it via all classifiers, and use a Mode operation to get the final classification result to identify different personality types

PySimpleGUI is used to create user interface**.**

Fig. 3: GUI for personality detection

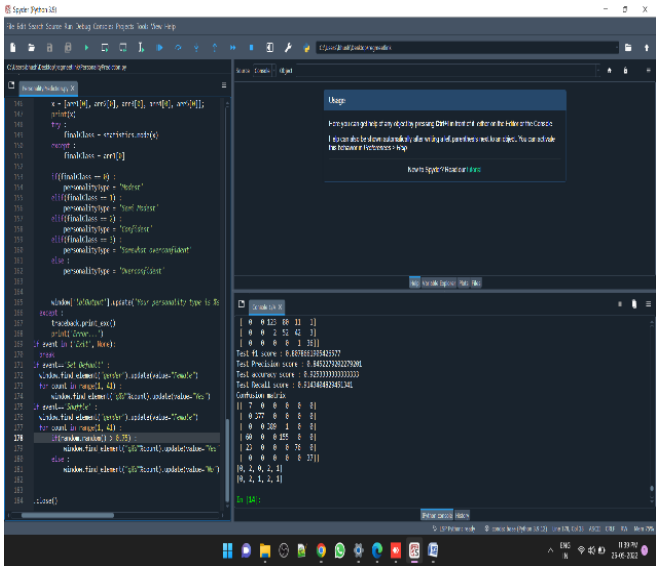
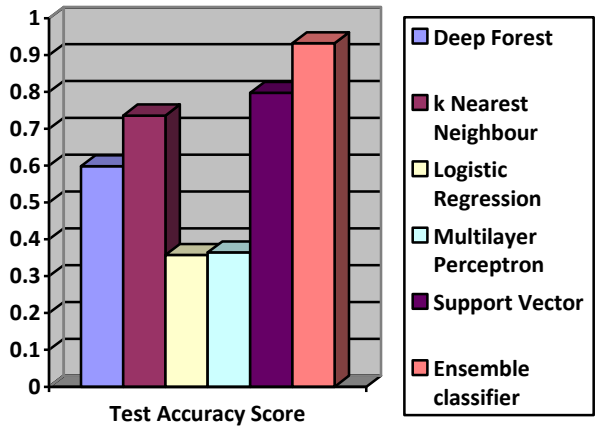


Fig. 4: Test F1, Test recall, Test Accuracy, Test precision score for ensemble classifier

Classifier	Test Accuracy Score
Deep Forest	0.598222222222222
k Nearest Neighbour	0.736
Logistic Regression	0.357333333333333
Multilayer Perceptron	0.364444444444444
Support Vector	0.797333333333333
Ensemble classifier	0.931555555555555

Table 1: Classifiers with test accuracy score



Graph 1: Classifiers with test accuracy score

VI. CONCLUSION

Research in prediction and analysis of human being is in great demand these days. Predicting the personality of the candidatebythis system have made things easy in various fields like recruitment process, medical counselling, and likewise. Personality prediction using questionnairehelps to find out the behavioroulfeatures of the candidates taking the survey. This paper focuses on providing a accurate system forpersonality detection from questionnaire.

ISSN No:-2456-2165

REFERENCES

[1.] Aleksandar Kartelj, Vladimir Filipović, Veljko Milutinović, Novel approaches to automated personality classification: Ideas and theirpotentials.

[2.] Fazel Keshtkar, Candice Burkett, Haiying Li and Arthur C. Graesser,Using Data Mining Techniques to Detect the Personality of Players in an Educational Game

[3.] G. Kumar and P. K. Bhatia, "Comparative Analysis of Software Engineering Models from Traditional to Modern Methodologies," in 2014 Fourth International Conference on Advanced Computing & Communication Technologies, 2014, pp. 189–196.

[4.] D. Leffingwell, Agile Software Requirements Lean Requirements Practices for Teams, Programs, and the Enterprise, 1st ed., MA: Addison Wesley, 2011.

[5.] "Manifesto for Agile Software Development."]Online[.Available: http://www.Agilemanifesto.org/.]Accessed: 20-Sep-2015[.

[6.] "Principles behind the Agile Manifesto."]Online[.Available:http://Agilemanifesto.org/principles.html.]Accessed: 20-Sep-2015[.

[7.] Nurbiha A Shukora , Zaidatun Tasira, Henny Vander Meijden)2015(. An Examination of Online Learning Effectiveness using Data Mining, Science Direct - Procedia - Social and Behavioural Sciences 172, 555 – 562

[8.] D. Bishop and A. Deokar, "Toward an Understanding of Preference for Agile Software Development Methods from a Personality Theory Perspective," in 2014 47th

[9.] L. Yan, Z. Mingyuan, and Y. Yongbo, "Risk Correlation Analysis Based on Information Management," in 2010 3rd International Conference on Information Management, Innovation Management and Industrial Engineering, 2010, vol. 4, pp. 27.

[10.] H. Pretorius and H. Zaaiman, "Why use communication training as enterprise-wide project risk mitigation tool?," Enterprise Systems Conference)ES(, 2013. pp. 1–6,.

[11.] "Personality and Values."]Online[. Available: http://saylordotorg.github.io/text_principles-of-management-v1.1/s06-0 2-personality-and-values.html.]Accessed: 20-Sep-2015[.

[12.] S. Zhu and L. Wang, "Research on software undergraduates training countermeasures based on the competency model," in 2011 6th International Conference on Computer Science & Education)ICCSE(, 2011, pp. 804–807.

[13.] T. Kanij, R. Merkel, and J. Grundy, "An empirical study of the effects of personality on software testing," in 2013 26th International Conference on Software Engineering Education and Training)CSEE&T(, 2013, pp. 239–248.

[14.] R. Kaplan and D. Saccuzzo, "Psychological Testing: Principles, Applications, and Issues," 2012, pp. 7–9.

[15.] S. John, O. P., & Srivastava, "Big Five Inventory)BFI(," Handb. Personal. Theory Res., vol. 2, pp. 102–138, 1999.

[16.] "Rosenberg Self-Esteem Scale)SES(- Statistics Solutions."]Online[. Available: http://www.statisticssolutions.com/ rosenberg-self-esteem-scale-ses/.]Accessed: 28-Sep-2015[.

[17.] Madarapu Soumica, Chamarthi Somasekhar Varma, Bobbili Siva Rama Krishna, 2021, Customer Personality Prediction using the Ensemble Technique, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY)IJERT(Volume 10, Issue 12)December 2021(,