PERSONALITY CLASSIFICATION USING DATA MINING

[1]Salomi Pawar, [2]Sakshi Panchal, [3]Madhavi Kad
[1]Student, [2]Student, [3]Student, [4]Assistant Professor
Department of Information Technology,
Ajeenkya DY Patil University, Pune, India.

Abstract—Personality of a person decides whether he can play the role of leader, influence people around, mastering communication skills, do collaborative work, able to do negotiation is business and handle stress. This project deals with the areas wherever it determines the characteristics of someone based on the frequent patterns observed. Personality classification refers to the psychological classification of different types of individuals. The analysis is done using vast set of data in dataset and is being compared with the user input. In this project, the classification of personalities will be done on the basis of these specific characteristics; conscientiousness, openness, extroversion, agreeableness, neuroticism. Researchers have utilised social media data for auto predicting personality. However, it is confusing and complex to mine the social media data as the data can be noisy. The paper proposes machine learning techniques using Random Forest, Logistic Regression, Decision Tree, Support Vector Machine, KNN. The process of implementation and obtaining of the result will include certain steps like- Data collection, Attribute selection, Preprocessing of data, Prediction of personality. The type of personality classification and prediction can be used in certain fields like business intelligence, marketing and psychology. Research in prediction and analysis of human being is in great demand these days. Predicting the personality of candidates by this system has made things simple in varied fields like recruitment procedure, medical counseling and likewise. Personality prediction using the questionnaire helps to find out the behavioral features of the individuals taking the survey.

Index terms—personality classification and prediction, Random forest, SVM, decision tree, logistic regression, KNN, frequent patterns.

1.INTRODUCTION

Personality is defined as the characteristics that make a person unique in terms of their way of thoughts, emotions, behaviors, habits and interests while influencing how one make decisions in life. Personality classification is one of the most studied topics in recent years**.** Personality is a combination of an individual's characteristics features that determines how he/she will behave in various situations. Personality influences every choice a person makes, from books to clothing to music and movies.
Personality also influences its interaction with the outside world and its environment. Personality can be used inthe hiring process, career counselling, health counseling, etc. It can also be used as an additional feature.The routine method of self-
estimation is time consuming and limited in scale.
Also, this manual breakdown couldn't provide precise conclusions while analysing the personality of a user from their nature and conduct. Since the analysis is suited manually, it affects the exactness of the conclusions because people are biased.
Data mining techniques are therefore used to study and analyse data and then identify any hidden patterns or information from a large dataset. This process is used to create my user profile, which is used to train models to predict the future behaviour of other users.
Motivational influences and human behaviour are the best predictors in personality that will predict an individual's work performance. People's experiences which are emotionally important with circumstances, can also be influenced by personality. This approach reflects a person's character. Based on the above mentioned parameters they defined the personality of a person as a set of attributes that describes a likelihood on the uniqueness of behaviour, feeling and thoughts of the person. These attributes of a person change through time and situations. In a easy term, we can admire personality as a blend of characteristics and standards that built an

candidates's one in many character. There are many different personalities used to define a person. Here we are using these big five personality traits openness, conscientiousness, neuroticism, extraversion and agreeableness.

People currently deliver their thoughts and feelings through social media platforms. The posts can be in so numerous ways, similar as using an image, URL link, and music. People's personality also can be fetched using social media. The personality of people has proved to be useful in predicting job satisfaction, professional and romantic relationship success. In the process of recruiting the rightful candidates, companies currently tend to examine the candidates' social media profiles to know the personality of the applicants for a selected job.

The big five come from the statistical analysis of responses to personality details. Using a technique called factor analysis researchers can look at the responses of people to hundreds of personality items and ask the question" what is the best was to summarize an individual?". This has been done with numerous samples from all over the world and the overall conclusion is that, while there seem to be limitless personality variables, five stand out from the rest in accordance to explaining a lot of a individual answers to questions about their personality: extraversion, neuroticism, agreeableness, conscientiousness and openness to experience. The big- five aren't associated with any particular test, a variety of measures have been developed to measure them. This test uses the Big- Five Factor Markers from the International Personality Item Pool, evolved by Goldberg(1992).

Our Big Five Personality Traits:
• Openness: As the word suggests, This quality features characteristics such as openness and imagination and curiosity.

• Conscientiousness: Conscientiousness talks about a high amount of thoughtfulness, a goal- oriented attitude and good decision-makers.

• Extraversion: Extraversion also means extroversion is identified by excitement, talkativeness and assertiveness.

• Agreeableness: Agreeableness refers to features such as trust, affection and social behaviour of an individual.

• Neuroticism: Neuroticism includes attributes like sadness, moodiness and sudden burst of emotions.

As these five dimensions cover almost all avenues needed to know someone it is the right method that forms the basis of a person's overall personality.

OBJECTIVE:

The main objective of this project is to overview the data mining algorithms which are used to predict the personality of the user. In this paper, we focus on an online dataset and then personality would be predicted accordingly based on the Big Five Personality traits. In this way, we can filter candidates applying for certain positions in the organization. Therefore, it would save the resources of the organization and they would also solicit only those applicants which would be most suitable for the job.

2. RELATED WORK:

Kalghatgi presented a Neural Network Approach based on the Big Five Test to predict the personality of individuals depending on tweets published on Twitter by extracting meta-attributes from tweets. Which are used to analyze ones social behaviour. The authors followed a four-step process which is Data Collection from tweets, Preprocessing, Transformation and Classification. Although neural networks are used to predict personality there are limitations such as countering fake information, automatic analysis of tweets and relying on just Twitter is not enough to predict an individual's personality but only user behaviour and trends.

Allan Robey propounded a system to degrade the cargo on the Human Resource department of companies by owning two sides association and seeker initiated. The authors affirm that the offered system will be more effective to shortlist CVs from a large pool making sure that the ranking is fair and legal. The main difference between the being system and the proposed system is that rather of just surveying the CVs, the authors propose to conduct an aptitude test and a personality test for personality vaticination. Personality Prediction System from Facebook user for numerous times, Facebook has been using Personality Prediction Systems that can prognosticate a user's personality automatically from their Facebook functions. Facebook uses the Big Five Personality model that directly predicts a user's personality grounded on someone's personality traits. Several traits can be discovered using this model similar as extraversion, meticulousness, neuroticism, affability and openness. In this study, the experimenters used two collections of datasets to prognosticate the user's personality. The first dataset is samples data from the my Personality design, and the alternate dataset is data that was generated manually.

3. PROPOSED SYSTEMS:

1. Dataset:

As homemade data collection is time-consuming, we analysed an online check data containing further than 10,000 rows. The collected data was in csv and exceed format.

2. Methodology:

The ideal of our paper is to prognosticate the personality of an existent grounded on the parameters openness, extraversion affability, neuroticism and meticulousness. Grounded on the check options, there are certain statements that are solitarily towards the below mentioned parameters like "I'm the life of the party" easily depicts an extroversion personality.

3. Workflow:
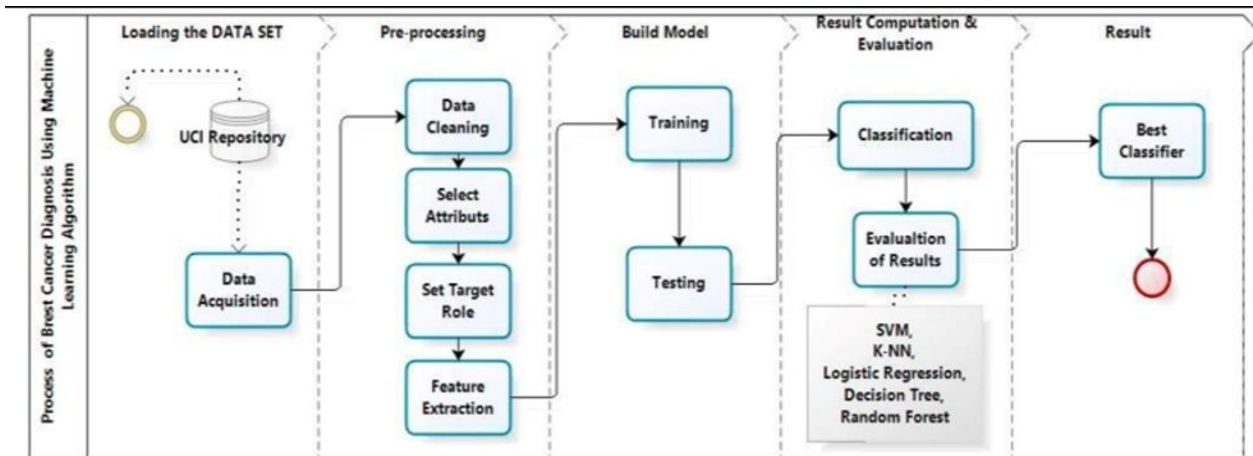
4. Model Development:

Dataset: We collected the data from Kaggle repository (https://www.kaggle.com/datasets/tunguz/big-five-personality-test) In this study, the dataset contains over 10000 rows of data with multiple columns,

Exploratory data analysis: We conducted exploratory data analysis to get visual representation for further disquisition through a bar plot. The number of statements per personality was examined to gain the intuitive idea of judgment structure for each personality. In this step, we used 'Seaborn' which is a Python data visualization library and 'Matplotlib' which is a Python 2D plotting library for data visualisation and correlation of the personality types.

Agreeable Personality : AGR1

Conscientious Personality: CSN1
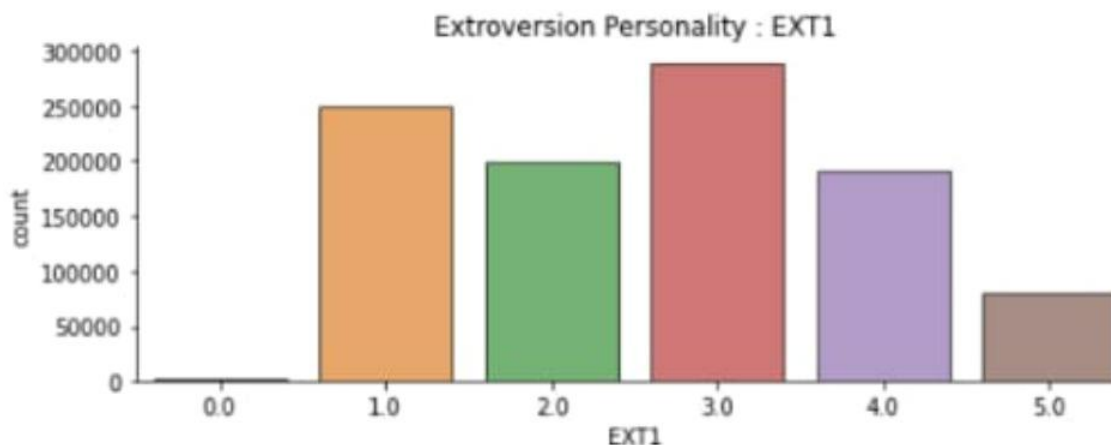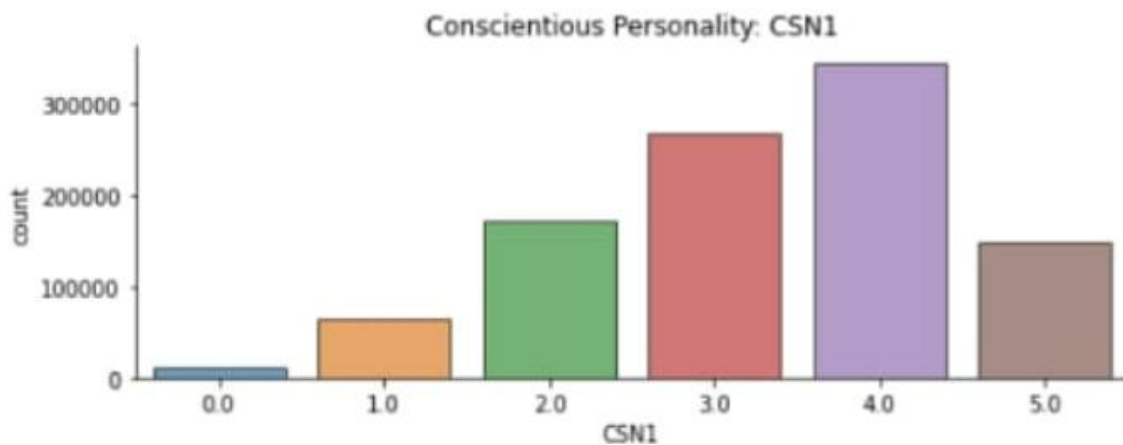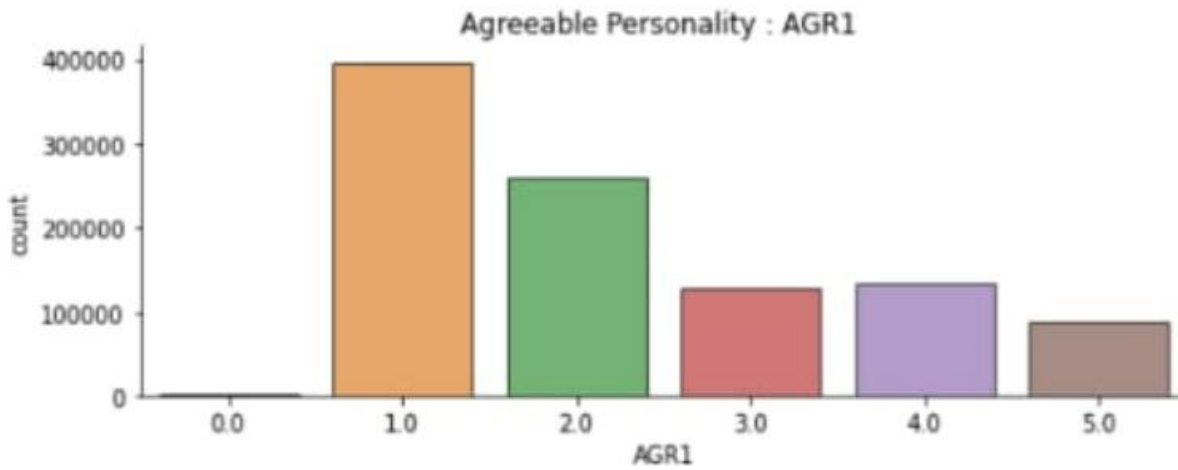
4.1 Model Testing and Training:

Using the sklearn library, we've used 70 of our data for training purposes and 30 for testing the results. For prognosticating the personality of a prospective seeker, we've used colourful machine literacy and Data Mining algorithms like Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM) and KNN.

- Logistic Regression

It's an algorithm similar to Linear Regression, except it predicts whether commodity is true or false. It's a popular algorithm for working bracket problems like double Bracket (Pass/Fail, Rain/No Rain).

- Decision Tree

A decision tree is one of the most important tools of supervised algorithms used for both bracket and regression tasks. It builds a flowchart-suchlike tree structure where each internal knot denotes a test on a trait, each branch represents an outgrowth of the test, and each splint knot terminal knot) holds a class marker. It's constructed by recursively splitting the training data into subsets grounded on the values of the attributes until a stopping criterion is met, similar as the maximum depth of the tree or the minimal number of samples needed to resolve a knot.

- kNN

kNN stands for k-nearest neighbours, a supervised machine learning algorithm capable of solving both regression and bracket problems. Intuitively we can think of the proverb Birds of the same feather flock together as similar to kNN. The algorithm assumes that similar data points usually occur in close proximity. The algorithm assumes that analogous data points generally do in close propinquity.

- SVM

Support vector machine is a supervised machine learning algorithm used to deal with data for bracket and regression analysis. The thing of SVM is to find a hyperplane in N dimensional space (N-number of features) that can fluently classify the data points.

- Random Forest

Random forest is another ensemble technique used for bracket and regression tasks. It uses multiple decision trees to produce the result. Bagging or bootstrap aggregation are used to train the random forest algorithm's forest.

4.2 Model Evaluation:

After feeding the data to the models, we managed to gain the loftiest delicacy of these two algorithms Random Forest algorithm and Decision Tree algorithm, they give the stylish delicacy followed by SVM, KNN and Logistic Regression.
As anticipated, Random Forest has the least mean squared error, which measures the normal of the forecourt of the difference between factual and estimated values.
The results of this analysis can also be used to make advancements to the model or to inform farther exploration on sentiment analysis of social media content.

5. CONCLUSION AND FUTURE SCOPE:

In this paper, we have used different types of powerful Machine Learning Algorithms such as Logistic Regression, Naive Bayes, Random Forest, SVM and KNN for Personality prediction. Using pyresparser, spaCy and PhraseMatcher we were able to predict the personalities of various individuals. The results indicate Random Forest has the maximum delicacy of0.71 still due to lack of available data, the delicacy is significantly lower than it was anticipated. The proposed system can be used by numerous companies in order to streamline the reclamation process by considering the personality of implicit campaigners.

Upcoming work can also be beseemed to add the effectiveness and performance of the proposed system in arrangement to foretell the personality more directly.

The predictive analysis of personality assessments provides an overview of a candidate's behavioral tendencies and that allows recruiters to really understand if a candidate will, in fact, be a top performer and if he will fit the culture of the company. It helps people learn about their personality types and attributes. It reveals the candidate's personality traits, such as work preferences, motivation, strengths and weaknesses and attitude.

Personality tests can reveal whether or not the candidate fits the position and the team in more than just capability and skills. It displays whether they are capable of thinking on their feet, how they approach problem-solving, and whether they display leadership skills when necessary or under pressure.

6. REFERENCES:

A Demetriou, L. Kyriakides, and C. Avraamidou, The missing link in the relations between intelligence and personality in Journal of Research in Personality,vol. 37 issue 6,December 2003, pp 547- 581.

M. Kalghatgi, M Ramannavar, and Dr. N. S. Sidnal, Neural Network approach to personality prediction based on the Big-Five Model in IJIRAE, vol2 issue 8, August 2015, pp 56-63.

A..Robey, K. Shukla, K. Agarwal, K. Joshi, Professor S. Joshi Personality prediction system through CV Analysis, in IRJET vol 6, issue 02, February2019.

J. Zubeda, M. Shaheen, G. Narsayya Godavari, and S. Naseem Resume Ranking using NLP and Machine Learning, unpublished

Md Tanzim Reza, and Md. Sakib Zaman, Analyzing CV/Resume using natural language processing and machine learning, unpublished.

https://towardsdatascience.com/

https://spacy.io/

https://www.kaggle.com/