Personality Prediction with social media using Machine Learning

Dr Swetha P[1], Sunil B[2], Viresh[2], Vivek S Shyavi[2]

[1]*Asst Professor, Department of Computer Science and Engineering, Global Academy of Technology*
[2]*Student, Department of Computer Science and Engineering, Global Academy of Technology*

Abstract: Predicting Personality with a much more accuracy that can be very useful for the betterment of the society. And many of the researchers were conducted a use of the given data for the purpose of organizational development, social marketing network, individually recognized recommendations and health care systematic. And with an introduction the popularity of the social media networking websites or online platforms like Twitter, Facebook and LinkedIn researchers were conducted based on the given data from public are available through the online social networking platforms and social behaviors aspects that could help to the friends and followers for the much prediction of personality. In the paper we proposed a system that mainly aims to automate the personality prediction assessment process by training a model using Myers Briggs Type Indicator (MBTI) personality prediction analysis. Online social media platforms networks provide more and more opportunity from knowledge perspective discovery and the data from public with respect to their view. The Structured information of the social media networking content can be provided with the knowledge to predict the main and important of the personality traits. People nowadays share some of their thoughts, information, feelings more specifically their personality traits, on social media platforms more often they are the sources of data generation. Many researchers faced problems and many challenges and still they facing and working towards on and in the future, there is plenty of opportunity for many research to predict it with the better accuracy and good results.

Keywords: ***Tweets, Comments, Myers Briggs Type Indicator (MBTI System)***

I. INTRODUCTION

Personality Prediction using machine learning with social the behavior of the persons can be predicted by analyzing with the social media data or from tweets. And we can predict the personality, behavior of a person with the social media data. Prediction of personality with better accuracy could be very useful for the betterment of society. Personality is the way through which a person responds to individuals response situation with his feelings. It is the new way of combining the data of the characteristics that make an individual behavior as unique in his own way. Hence there is need of something through which we can increase the number of many people involved in the given survey and to make the process more reliable and accurate. Personality prediction using machine learning with social behavior of a particular person can be analyzed by using social media data. The main Activities performed on Twitter, Social media platforms Facebook can provide concrete, with the enormous of data set. There can be a variety of content and data processing available here and it can be used for multiple analyses. Through the online social media networking platforms bring both the advantage and which can dangers to society, which need to be analyzed in its way to examined carefully as needed. The Myers-Briggs type indicator device. The Myers Briggs Type Indicator (MBTI system) is a personality predicting type system that system that divides people into 16 components of personality types based on four axes:

Extroversion(E) – Introversion(I)

Thinking(T) – Feeling(F)

Judging(J) – Perceiving(P)

Sensing(S) – Intuition(N)

With a brief discussion which includes the many open type issues for the further research in the respected areas of online social media platform networking sites-based personality prediction predicting conclusion. Researchers faced many problems and challenges which has still lot of scope in the future research to the prediction of the personality of a individual person with better accuracy. Online social media networks provide unprecedented opportunity from knowledge researching and data detecting at that point of view. Social media networking will very rich content in the form of text, video and image etc. Algorithm used for prediction plays very important to get an accurate result. With the recent data research, the researchers have used regression method analysis with the data assessment of personality traits to predict personality.

II. LITERATUR SURVEY

DI XUE, ZHENG HONG in 2017 worked on the Personality prediction for the data on social network platform for the data with the label distributing the data for learning. Researchers used the new way of prediction techniques in the machine learning model named label distribution learning in the field and it may be respected with the purpose. Personality prediction will be a psychological construct which will aim to explain the new wide variety

with which a human's behavior and the persons predictability way of behaving the social media platforms in terms of a stable and standard way of personality prediction must be important and a psychological construct which may account for an individually available differences in the way of prediction perspective. It can be a important reason to recognize and reliable in an effective way of personality prediction which will able to recognize the persons personality where a prediction can be done it with the social media platforms.

Hetal Vora, Mamta Bhamare in 2020 worked on personality prediction with the social media data as a text, according to the researchers the shallower, it will be implemented and can be done using natural language techniques for the prediction and it can be implemented and predicted easily as possible. The new technique implemented with the predictability model should be the regular way to check the various dataset. The proper work which increases the size of the datasets and to improve and enhance the feature extraction and the prediction purpose to improve the accuracy better way to predict the personality. With the multivariate should be predictable and can be predicted with the dataset and regression model for all the traits at the great level of prediction. The more advantageous way that could be used to provide and recommend the new way of services and it may help to predict the social media data and marketing strategies in all other ways apart from text, comments they used the audio, video usage in social media for the prediction of personality.

S. Adali and J. Golbeck, in 2012 worked on Predicting Personality with Social Behavior, researchers were present the series for understand the human behavior on the social media network twitter, which shows the correlations between the measures and personality traits and it demonstrate the usage of personality prediction which uses the behavioral features which was equivalent to predict the personality successfully for the prediction using the features. In the way which we can predict the individual persons behavior that can be analyzed with the many levels of research that analyzes to which it can be predicted personality can b predicted and analyzed for specific roles of a person takes.

Jenifer Golbeck, Cristina Robles, Micon Edmondson and Karen Turner in 2011 they worked on the prediction of personality from the twitter data and the users big five model can be analyzed with researchers model which can be implemented and predicted from the data of the information which shares the data on twitter. In the proper way of prediction of the subjects and completed a personality prediction test and through the data and it can be collected in a proper way from their profiles. A data as feature can be implemented with the machine learning algorithms Gaussian processes to predict the accuracy with better way.

K. N. P. Kumar et.al in 2019 worked on Personality traits classification on twitter, created has created a framework for distinguishing personality traits based on twitter text that includes language features and a combination of numerical code based on numbers and word embedding. They used the MBTI database to train and validate classification algorithms. To differentiate, they used the machine-learning algorithm XGBoost and SVM. The weighted total vector representation of each suitable document was made using the TF-IDF vectorizer

J. Staiano, B. Lepri and N. Aharony in 2012 worked on Friends don't Lie : Inferring Personality Traits from Social Network Structure, according to researchers study contributes to a growing body of research integrating language and personality. In the results of the multiple analysis which will provide more support for the better accuracy and for the positive effectiveness of psychopathy in the given personnel decision-making method.

III.EXPERIMENTAL SETUP

Google Collaboratory is the free online cloud-based GPU's, CPU's and TPU's environment of type that allows us to train the model of machine learning. We make use of Google Collaboratory since we face Memory error because it uses the machine learning model for the personality of individual persons using machine learning algorithm on the large dataset. And it will help to execute and access the unlimited computational power with the usage of system.

3.1 HARDWARE REQUIREMENTS

- CPU: Intel 8th Gen i5 or i7
- RAM: 8GB or More
- OPERATING SYSTEM: Windows 10 And Above
- PROCESSOR: Intel Core v4 10 core processor, 2.2GHz with turbo boost up to 3.11GHz, 50 MB cache memory.
- Motherboard: ASRock EPC612D8A
- STORAGE: 120GB Storage Space

3.2 SOFTWARE REQUIREMENTS

- Programming Language: Python 3.7.0 and above
- Python Packages:
- Sklearn - 0.14.5
- NumPy - 1.14.6
- Pandas - 0.25.0
- Matplotlib - 2.2.3

IV.SYSTEM DESIGN

The main object of the design phase is to produce the overall design of the software. All the System design process is one of the main and most important phases of the software development processes. The main aim and purposes of the problem which specify the requirements documentation. It can be implemented in a first step in a solution to the main problem is the main design feature of the important project. Feature extraction of the design implementation of the system is the most and critical factor which affects the quality for the software. The important main aims to figure out the machine learning models so that it can be fulfils all the system requirements in an effective way. The model conceits of user, who feed the input data to the system where the data preprocessing, feature extraction weightage assessment for extracted features for personality prediction of the given model to train and extract the features and final feed into the trained model.

4.1 System Architecture:

An architecture of the system which is the most important conceptual way of model that predicts the personality of the persons behavior, the implemented way of the structured views of a system. The incremental architecture of a formal way of description is the representation of the system requirements. It helps to organize and represents the supports for the prediction of the reasoning structure and which behaves in the respected way of the system.

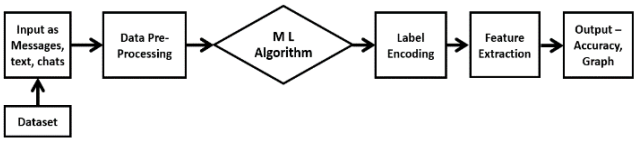


Fig 4.1.1 System Architecture

4.2 Data Flow Diagrams:

The basic way of the Data Flow Diagram is the main overview of the whole systems which process the important part of analyzing or modeled method of prediction. Designed way of concept where a personality prediction can be predicted in a way which shows the single level process as a system diagram, and to the external entities for the system relationships. Data flow diagram will further divide into different levels which shows the detailed data flow of level 0, level 1.

4.2.1 Data Flow Diagram Level 0:

Fig 4.2.1 Data Flow Diagram Level 0

4.2.2 Data Flow Diagram Level 1:

The data which has the flow diagram and notifies the single process for the main diagram which broke down with respect to the process flow of system flow diagram.

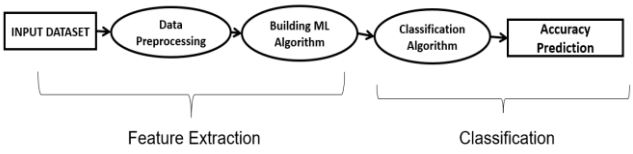


Fig 4.2.2 Data Flow Diagram Level 1

4.3 Use Case Diagram:

Unified modeling language a use case diagram which summarizes the details of systems users which also called the actors in use case diagram. And it interacts with the system to performs the action on it. Use case diagram is the main and an important diagram which shows the internal and the external factors for the prediction of model interactions among some users with different components usage. The most and important effective way of the usage of use case diagram it will help to represent in the system for the application or system interacts with people, and organizes the external system goals that helps to analyze the entities. The system diagram models the subsystem of an application. With a single use case diagram helps to captures the more functionality of the system.

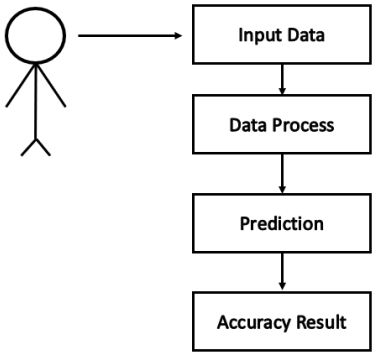


Fig 4.3.1 Use case diagram

V. METHODOLOGY

In this project, modules used are data pre-processing, building machine learning controls and classifying algorithms. We will be using algorithms of machine learning which can be used with tweets, messages, comments, text and many features.

A. DATA COLLECTION

The data which also contain various punctuations marks, emoticons which need to be cleaned and removed and

helps to predict the personality of the data in the dataset. In the dataset which were collected from the huge social behavior of the persons in the social media network and it can be implemented with the many informal way of words that could not be useful for the semantic way of prediction and identifying the persons behavior in a respective way of data prediction in one's personality. A data in the dataset contains the various punctuations and a lot of emoticons were cleaned with a rounding of many of the cleaning with all the redundant words and multiple words were cleaned in the regular expression in the processing library. The given dataset is cleaned with the related for the prediction which is more required and implementation for analyzing the personality types which were implemented by taking the MBTI system which uses the personality types in other way much of certain attributes in the dataset it can be unbalanced when we are training the dataset and removes the minority techniques which are used. The main aim of the given model is to organize the scenario as much as possible. The paper focuses on the data of predicting and designing the main parameters of the model in which it has less balance in the dataset to obtain the accuracy.

B. FEATURE EXTRACTION TECHNIQUE

Text with which can helpful and can be implemented as much as directly into the given machine learning model with given dataset.in the feature extraction we can extract data with the particular order to the individual persons behavioral types. From which we can create specific personality types with many of words which uses the personality traits. In order to make it more helpfully to the text and a certain weightage of data need to be classified and need to be processed to be added to the given words. Vectorizing it to process the data in a respective way to predict the personality. The data with unstructured format text which was further implemented with the data values based on the vectorizing method. Weightage values can be mathematical and documented unusual frequency. All the raw data extraction in the feature extracting can be implemented in the respective way of prediction of the persons behavior which was numerically implemented and analyzed and can be ignored with the data which can be analyzed and predicted with greater responsibility and knowledge personality.

C. CLASSIFICATION MODEL GENERATION

Random Forest Prediction

Prediction with the algorithm classification in the model of random forest prediction and improvement in prediction with the predictive accuracy of the dataset. The number of subset in the dataset with different attributes that can take average of given model.

The most popular way is the supervised learning technique in the given model for the random forest in the machine learning algorithm for the data that can be easily represented and implemented.

In the concept of resemblance which enables the process in the particular way to the learning techniques and can represented in the way through which the multiple classification which has the solution to perform and improve given machine learning model.

Logistic Regression Prediction

In prediction the logistic regression is the most effective way to the popular machine learning algorithms, it can be represented and implemented with given model for the better accuracy prediction. Under the supervised learning technique, it can be implemented and initialized to the categorical way of dependent variable using a given set for the improvement variables.

For the significant implementation of the given algorithm hence it can make the probabilities and to perform new data.in the dataset for the data using continues and discrete datasets.

Logistic regression predicts the required output in a categorical dependent variable. However, the outcome of the values in the data with the dependent variable either yes or no, 0 or 1, true or false etc

KNN neighbor prediction model

K nearest neighbor KNN algorithm can be implemented to store the data in the available dataset and it classifies the new data point based on the required way for the prediction and can be implemented with the content and appears with the easily classified data representation. Data appears then it classifies with aim to predict the accuracy and well-suited category by using KNN algorithm.

KNN algorithm with the training data that possess the way of representing new way of data prediction which can be stored at the new data point for the prediction of personality, then it can be classified with category which has same values. The Unwanted URLs removed, since URLs are just hyperlinks that do not contain any information, they need to be removed.

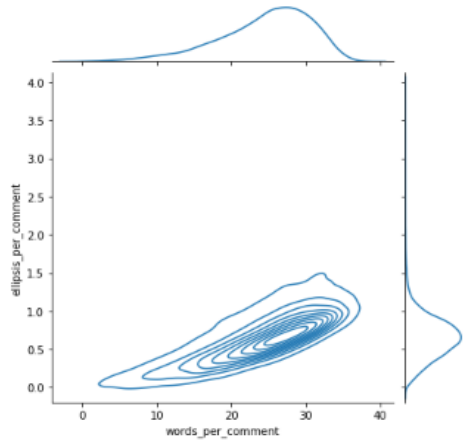
D. DATA CLEANING AND PREPROCESSING

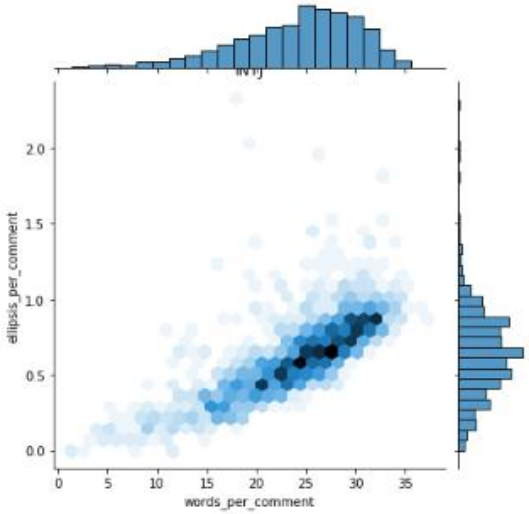
In data cleaning our dataset by removing unnecessary parts of data that would have no role in the prediction task. All the data can be Cleaned and organized with the data point and a raw data to make it suitable for machine learning model.

V. RESULTS

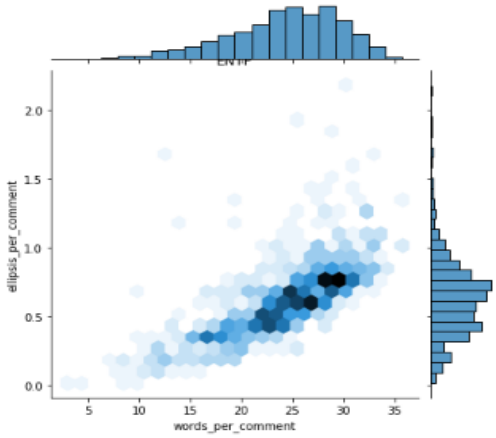
In the section we are demonstrating implementation on machine learning models on each feature extraction technique separately as follows:

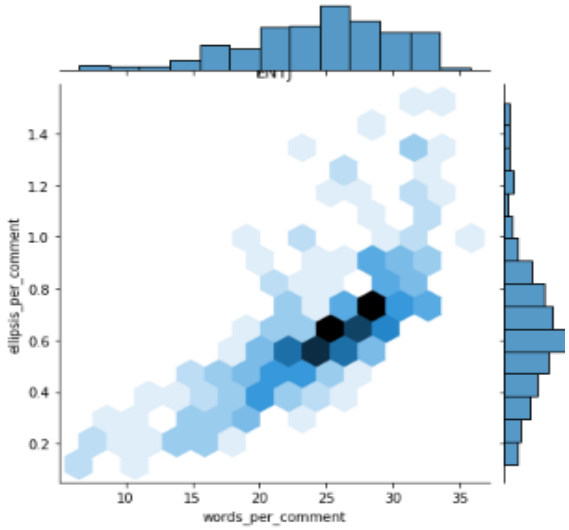
Creating new columns showing the amount of question marks per comment, exclamations or other types

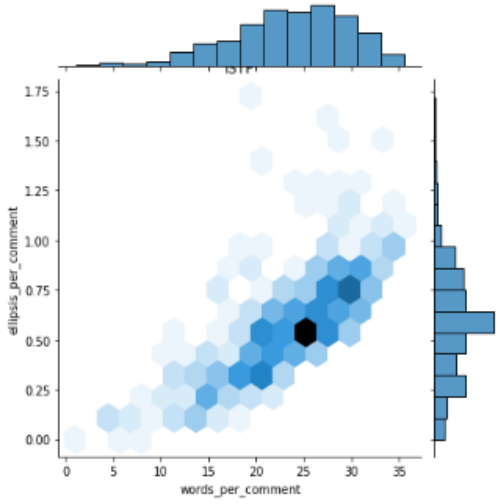


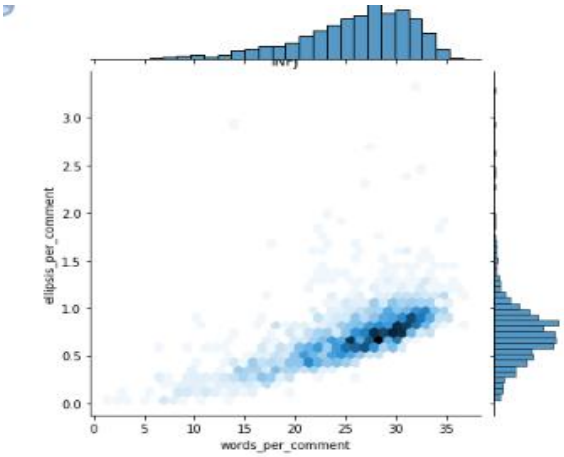


Here we focusing on the correlation variables for the different types of the personality types the Meyer Briggs outline in comparison to the word's comments and ellipses per comment type will yield the highest correlation value.

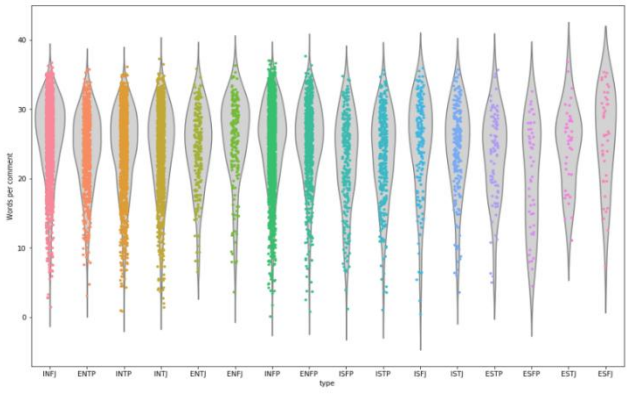




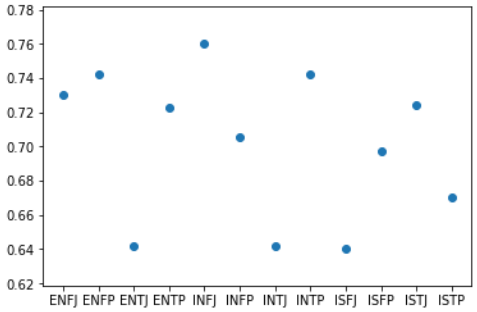




We have compared and analyzed all the above personality types.



.



It can be inferred from the above graphs that we can see different prediction results of different personality types.

VI. CONCLUSIONS

Personality prediction with which uses behavioral features was equivalently successful to prediction using social media network such as Twitter, Facebook etc. With the social behavior of the persons prediction in the online social media platforms for the different perspective prediction of personality for the implementation and usage of behavioral data aspects of social media network like a comment data, message content and type of comment with the people's behavior on the social media platforms towards the friends, followers' behavior can be implemented. Which has more useful thing in the future work which can be analyzed and implemented for the researchers in which we need to get the accurate data prediction. When the data in the dataset getting data with better accuracy with usage of different types of networks and process it to predict the personality. Researchers envisaged with the lot of data and can be able to predict the personality prediction automatically and can be implemented easily. For content based and linking data in the twitter data ,which can be predicted and analyzed quite accurately.

REFERENCES

1. Di Xue1 , Zheng Hong1 , Shize Guo2 , Liang Gao2 , Lifa Wu1 , Jinghua Zheng3 , And Nan Zhao4 "Personality Recognition on Social Media With Label Distribution Learning" IEEE Access on Personality Prediction on social media with LDL Volume No. 5 (2017):1348 – 13488

2. Hetal Vora , Mamta Bhamare , Dr. K. Ashok Kumar, 2020, "Personality Prediction from Social Media Text: An Overview", International Journal Of Engineering Research & Technology (Ijert) Volume 09, Issue 05 (May 2020),

3. S. Adali and J. Golbeck, "Predicting Personality with Social Behavior", Proc. the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012), pp. 302-309, 2012.

4. J. Golbeck, C. Robles, M. Edmondson, and K. Turner, "Predicting personality from twitter," in Proc. of the 3rd IEEE Intl. Conf. on Social Computing, 2011, pp. 149–156.

5. J. Staiano, B. Lepri and N. Aharony, "Friends don't Lie : Inferring Personality Traits from Social Network Structure", Proc. the 2012 ACM Conference on Ubiquitous Computing, pp. 321-33-, 2012

6. Myers, Isabel Briggs. "The Myers-Briggs Type Indicator: Manual (1962)." (1962)

7. J. Staiano, B. Lepri and N. Aharony, "Friends don't Lie : Inferring Personality Traits from Social Network Structure", Proc. the 2012 ACM Conference on Ubiquitous Computing, pp. 321-33-, 2012

8. Choong, En Jun, and Kasturi Dewi Varathan. "Predicting judging-perceiving of Myers-Briggs Type Indicator (MBTI) in online social forum." *PeerJ* 9 (2021): e11382.

9. T. Yo and K. Sasahara, "Inference of personal attributes from tweets using machine learning," 2017 IEEE International Conference on Big Data (Big Data), 2017.

10. B. Y. Pratama and R. Sarno, Prediction of Personality with the twitter text using KNN, Naive Bayes analysis and SVM, 2015 International Conference on Data and Software Engineering (ICoDSE), 2015