

Personality Prediction

A Project Work Synopsis

Submitted in the partial fulfilment for the award of the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING (AIML)

Submitted by:

| Sno. | Name | UID |
|------|--------------|-----------|
| 1 | Hitesh Kumar | 20BCS6157 |
| 3 | Priya Yadav | 20BCS6126 |

Under the Supervision of:

Mrs. Lata Gupta



**CHANDIGARH
UNIVERSITY**

Discover. Learn. Empower.

CHANDIGARH UNIVERSITY, GHARUAN, MOHALI -

140413,

PUNJAB

2020-24

ABSTRACT

Personality is an important aspect of one's perspective towards their life. It majorly impacts the decision-making and approach to solving the problem. Personality prediction to better accuracy could be very useful for society. There are many papers and research conducted on the usefulness of the data for various purposes like marketing, dating suggestions, organization development, personalized recommendations, and health care to name a few. Particular approaches differ concerning different machine learning algorithms, data sources, and feature sets. The goal of this project is to investigate the predictability of the personality traits of users based on different features and measures of the Big 5 model.

Timeline/Gantt Chart

| S.N | Strategies | 1 st week | 2 nd week | 3 rd week | 4 th week | 5 th week | 6 th week |
|-----|--------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| 1) | Problem Identification | | | | | | |
| 2) | Research & Analysis | | | | | | |
| 3) | Design | | | | | | |
| 4) | Coding | | | | | | |
| 5) | Implementation & testing | | | | | | |
| 6) | Project finalisation | | | | | | |
| 7) | Documentation | | | | | | |

Table of Contents

| | |
|-----------------------------------------------------|----|
| Title Page | 1 |
| Abstract | 2 |
| Timeline / Gantt Chart | 3 |
| | |
| 1. INTRODUCTION* | |
| 1.1 Problem Definition | 5 |
| 1.2 Project Overview/Specifications* (page-1 and 3) | 5 |
| 1.3 Hardware Specification | 6 |
| 1.4 Software Specification | 6 |
| | |
| 2. LITERATURE SURVEY | |
| 2.1 Existing System | 7 |
| 2.2 Proposed system | 7 |
| | |
| 3. Problem Formulation | 8 |
| | |
| 4. METHODOLOGY | 10 |
| 5. TENTATIVE CHAPTER PLAN FOR PROPOSED WORK | 12 |
| 6. APPENDICES | 13 |
| 7. REFERENCES | 14 |

1. INTRODUCTION

1.1 PROBLEM DEFINITION:

Knowing your personality helps you answer better the dilemma of how's and why's about yourself. You have a certain way and feel a certain way, but often it's hard to understand why especially when other people aren't thinking or feeling the same way. Getting insights into your personality helps in career growth by molding yourself accordingly. The main goal of this project is to know you better to excel in different aspects of life. A large portion of the information is accessible to humanitarian organizations that are not organized. Compiling questions for personality analysis and surveying the data collected for personality conclusions is a bit challenging.

1.2 PROJECT OVERVIEW:

Personality prediction is based on identifying the personality of an individual using machine learning algorithms and big 5 models. Personality plays a major role in one's personnel and professional life. Nowadays, many organizations have also started short listing the candidates based on their personality as this increase the efficiency of the work because the person is working in what he is good at than what he is forced to do. When the statistical analysis is applied to personality survey data, some person words used to describe the person and these words give a summary of the overall character or personality of the person accurately. Users can easily identify their personality from this system.

1.3 HARDWARE SPECIFICATIONS:

- Processor – 64-bit eight-core, 2.5GHz per core
- RAM – Minimum 4GB required
- Hard Disk – SSD or HDD minimum 40GB free space required
- Webcam – HD Webcam supporting 720p video recording

1.4 SOFTWARE SPECIFICATIONS:

- Edition - Windows 10 Home Single Language
- Version - 21H1
- Installed on - 16-08-2020
- OS build 19043.1526
- Experience Windows Feature Experience Pack 120.2212.4170.0
- Python installed – version 3.7 to 3.10
- Steamlit.
- Spyder.

Literature Survey

Arsa and Shubhangi, (2015) [1], aimed for developing handwritten- based personality and behaviour identification system. Supervised techniques Support vector machine and Artificial Neural network were employed into designing this model. They achieved their aim by securing the accuracy of 98.5%. In future, analysis will be performed for multiple lines.

Kedar et al.(2015) proposed [2] ,By using the supervised techniques ANN and Zernlike and Pseudo-Zernlike, Author builds a personality identification through handwriting analysis and graphology study, which is capable of determining personality of an individual by achieving the accuracy 90%. This system has its potential in personal recruitment in marketing, medicine, and counselling etc.

Ilmini and Fernando (2016)[3], build models to identify the personality traits from face image, identification of criminal behaviour in criminology etc. Implementation of Supervised technique ANN and SVM into the model achieved the accuracy of 98.5%. Large dataset can improve the accuracy of the classification. More study on psychology and improve feature extraction phase may improve final results.

Liu et al.(2016) [4], analyzed a broad range of interpretable image for personality features from twitter profile pictures using supervised techniques linear regression and root mean squared error. This system managed the accuracy of 89%. Further experiments needed on a dataset orders of magnitude larger than previous. To analyse more diverse set of photos that user post on social media.

Sagadevan et al. (2015) [5], recognized the personality of face book users from messages based on three Factor Personality (PEN) model. Techniques like supervised, stemming and part of speech tagging (POST) were employed into this system that managed to achieve the accuracy of 95%. In future, use of the higher negative words as cues to detect the psychoticism trait among face book users will be implemented.

Problem Formation

We are faced with many challenges which can be overcome by the advancing technology. One such challenge faced by every individual is getting to know their own self better. Knowing your personality helps you answer better the dilemma of how's and why's about yourself. You have a certain way and feel a certain way, but often it's hard to understand why especially when other people aren't thinking or feeling the same way. Getting insights into your personality helps in career growth by molding yourself accordingly. It's a vivid example of how technologies with a transformative potential like the web bring with them each opportunity and challenges; and it implies advanced equalization between basic rights and principles, as well as freedom of expression and also the defense of human dignity.

Using the Twitter dataset, one will get a charming mix of tweet contents and different connected information like hash tags, retweets, location, users, and a lot of others that pave the approach for perceptive analysis. The Twitter dataset consists of 31,962 tweets based on racism / sexist tweets. Twitter brings you vast information about what the world has to share about a topic whether it's movies, sentiments regarding U.S. elections, or the other trending topic like predicting who would win the FIFA tournament 2018. Operating with the Twitter dataset can assist you to perceive the challenges related to social media data processing and additionally find out about classifiers comprehensively. The foremost downside that, we tend to be acting on as a beginner is to make a model to classify tweets as positive or negative.

METHODOLOGY

Sentimental analysis of Twitter data reviews is a new area that needs more attention. The collected Twitter information has been developed from scratch to prepare for cleanliness. Also, using an element selection technique, important content is removed from the entire content. Third, the information bits are marked as physically true or negative tweets to configure the preparation package. Finally, preparation sets marked with individual highlights are provided as an aid to the classification to add additional information such as test sets.

Data Source

Determining the source of information is a critical task to advance the final investigation. Online media levels are broadly divided into three general classifications as sources of information a miniature writing for a blog website such as Twitter is ubiquitous due to its limited content and the public availability of information. The accompanying findings on the rate of Twitter development show that Twitter is being used as a site for review research.

Data Pre-processing

Information from Twitter is difficult. Collected Information Approximately, you must prepare or swap up raw data to execute a classification. Completed works include uniform packaging, hashtags and other Twitter documents (@, RT), emoticons, URLs, stopwords, slag word decompression and extended word stress.

Feature Extraction

Prefabricated datasets have different individual properties. In the component extraction technique, we separate the different angles, descriptive words, action words, and objects, and then identify these contexts as default or negative to indicate the intensity of the whole sentence.

Classifiers and Technique

Naïve Bayes Algorithm

The rule is known as when illustrious statistician Thomas Bayes WHO planned Bayesian theorem. This theorem assumes that each one the attributes square measure not absolutely freelance to every alternative. during this rule, chance for every attribute with regard to bound category level is calculated. The new

document {is categorized assessed is classed} victimization total of chances for every class. The classification framework is shortly mentioned as follows: Suppose we've D set of tuples and every tuple has attribute vector $X(x_1, x_2, x_3, \dots, x_n)$ of n dimensions. Let there square measure k variety of categories $C_1, C_2, C_3 \dots C_k$. The classifier predicts X belongs to C_i if

$$P(C_i | X) = P(C_i | X) \text{ for } 1 \leq i \leq k, i \neq j \quad (1)$$

We can calculate Posterior probability as:

$$P(C_i | X) = \frac{P(X | C_i) P(C_i)}{P(X)} \quad (2)$$

7. APPENDICES

Objectives:

Personality prediction takes one a step closer into knowing themselves better. It helps organization into finding a right fit for their required role.

Uses:

Personality prediction helps you to determine one's personality and his perspective towards his life. It helps organization to find cooperative employees that helps in maintaining the decorum of the office environment.

APPLICATIONS:

- Can be used to predict a person's personality with an accuracy of 85.81%.
- Used to identify the right candidate based on his personality and skill.
- Can be used to match martial profiles.

REFERENCES

- [1] Asra S., Shubhangi D. C., Personality Trait Identification Using Unconstrained Cursive and Mood Invariant Handwritten Text. *International Journal of Education and Management Engineering*, 2015, 5(5), 20
- [2] Kedar S., Nair V., Kulkarni S., Personality identification through handwriting analysis: a review. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, 2015, 5(1)
- [3] Ilimini K., Fernando T. G. I., Persons' personality traits recognition using machine learning algorithms and image processing techniques. *Advances in Computer Science: an International Journal*, 2016, 5(1), 40-44
- [4] Liu L., Preotiuc-Pietro D., Samani Z. R., Moghaddam M. E, Ungar L. H., Analyzing Personality through Social Media Profile Picture Choice. In *Tenth international AAAI conference on web and social media (ICWSM)*, 2016, 211-220.
- [5] Sagadevan S., Malim N. H. A. H., Husin M. H., Sentiment Valences for Automatic Personality Detection of Online Social Networks Users Using Three Factor Model. *Procedia Computer Science*, 2015, 72, 201-208
- [7] V. Kharde and S. Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques", *International Journal of Computer Applications*, vol. 139, p. 11, 2016.
- [8] Huma Parveen & Prof. Shikha Pandey "Sentiment Analysis on Twitter Data-set using Naive Bayes Algorithm", Dept. of Computer Science and Engineering Rungta College of Engineering and Technology Bhilai, India , 2016.
- [9] Rincy Jose & Varghese S Chooralil , "Prediction of Election Result by Enhanced Sentiment Analysis on Twitter Data using Classifier Ensemble Approach" ,Department of Computer Science and Engineering Rajagiri School of Engineering and technology Ernakulam, India , 2016.