Plagiarism Scan Report

Report Generation Date: May 26,2020 **Words:** 1611 **Characters:** 10074

Exclude URL :

7%	93%
Plagiarism	Unique

8	99
Plagiarized Sentences	Unique Sentences

Content Checked for Plagiarism

INDIRA GANDHI DELHI TECHNICAL UNIVERSITY FOR WOMEN MBTI PERSONALITY PREDICTION (BI Project Report) (Team 5) Submitted By Akanksha Singh- 01304092018 Arshiya- 04604092018 Bhavana sinha-02804092018 Harshita Dwivedi- 01904092018 Sapna Rai- 03004092018 Shivani Giri- 05404092018 Under the supervision of Mr. Rishabh Kaushal Assistant Professor Department of Information Technology i STUDENT UNDERTAKING Dated: ..... This is to undertake that the work titled MBTI Personality Prediction in this Minor Project Report as part of 4rd Semester in MCA (Information Technol- ogy) with specialization in during January { May, 2020 under the guidance of Rishabh Kaushal . The report has been written by us in my own words and not copied from elsewhere. This report was submitted to plagiarism detection software on 25-05-2020 and percentage 6% similarity found was 94% unique, similarity report attached as Appendix. Anything that appears in this report which is not my original has been duly and appropriately referred / cited / ac- knowledged. Any academic misconduct and dishonesty found now or in future in regard to above or any other matter pertaining to this report shall be solely and entirely my responsibility. In such a situation, I understand that a strict disciplinary action can be undertaken against me by the concerned authorities of the University now or in future and I shall abide by it. Student Signature Student Name Date-of-Submission, New Delhi ii DEPARTMENT OF INFORMATION TECHNOLOGY INDIRA GANDHI DELHI TECHNICAL UNIVERSITY FOR WOMEN KASHMERE GATE, DELHI - 110006 Dated: ..... CERTIFICATE This is to certify that the work titled MBTI personality prediction submitted by Team 5 in this project report as part of 4th Semester in MCA (Information Technology) with specialization in during January { May, 2020, done under my guidance and supervision. This work is her original work to the best of my knowledge and has not been submitted anywhere else for the award of any credits / degree whatsoever. The work is satisfactory for the award of Minor Project credits. Name and Signature of Faculty Advisor Designation Department of Information Technology Indira Gandhi Delhi Technical University for Women Name and Signature of Faculty Advisor Designation Department of Information Technology Indira Gandhi Delhi Technical University for Women iii ACKNOWLEDGEMENT It is high privilege for us to express our deep sense of gratitude to Assistant professor Rishabh Kaushal who helped us in the completion of the project. My special thanks to all Batch mate Seniors of Indira Gandhi Delhi Technical Uni- versity, New Delhi for helping us in the completion of project work and its report submission. Student Name iv MBTI Personality Prediction Akanksha Singh Arshiya Bhavana sinha Harshita Dwivedi Sapna Rai Shivani Giri May 2020 Abstract MBTI personality types can be predicted through many ways. The most commonly used methodology has been questionnaires that are time consuming and needs the focus of participant.This project will explore the area of predicting personalities without questionnaires. People are

increasingly using digital platforms like facebook, twitter ,etc. This gives us an opportunity to study if there's a way to predict their personality using these platforms. v 1 Introduction In the eld of psychology, personality is studied as it speaks volumes about how people behave in their life. As the world is advancing, people are using digital platforms like social medias to express themselves. Personality can be predicted using dierent models. One such model is the Myers-Briggs Type Indicator (MBTI) where personalities are divided into 16 dierent types. The MBTI di- vides the traits into four classes such as: Introversion (I) or Extroversion (E), Sensing (S) or Intuition (N), Thinking (T) or Feeling (F), and Judging (J) or Prospecting (P) . Eg: INFP. In this project, we want to study the correlation between the language of individuals used on their social medias and their re- spective personality traits. This will help us know that to what extent can we predict personality traits from various linguistic features. 1.1 Problem Statement & Objectives The MBTI tests use many multiple choice questions to determine the personality of an individual. But, this approach is time consuming and requires people to be focused enough to answer correctly. Thus, we think of minimizing the eorts of users and making it more ecient to predict a personality. We can thus model this as a classication problem. A successful implementation of such a classier would demonstrate a good connection between linguistic features and potential personality in general. This won't just help users know their personality but can also be used in psychoanalysis to help nd the relation between natural language and personality type. Predicting MBTI personality type using texts from social medias. Study the relation between natural language and MBTI personality. 1.2 Motivation If we happen to nd a correlation between natural language and MBTI per- sonalities, it can be a contribution towards psychoanalysis.People don't always realise how do they think like and a lot of what they post on social media may have a signicant relation with their true selves. This project will also help people know about themselves without solving a lot of questions that many a people nd tiresome and thus don't participate into nding their ownselves and if they participate then a lot of times they don't answer correctly. Employers can nd their employees using public information provided by employees if we achieve sucient accuracy in this project. 1.3 Scope and Limitation As we mentioned above, the project has scope in psychoanalysis however it has certain limitations. 1 The project is based on texts.It does not include images, URLs etc. People don't always express their original thoughts and writing texts isn't as interesting for everyone.They prefer sharing images, blogs, etc. written by another person than writing something about the topic on their own. This makes it dicult for our code to nd sucient data. If we include other factors like the images they share, the types of blogs they share, the content of articles they shared, the connects they have or prefer, their likes or dislikes. This may add value to our accuracy and provide sucient data. 2 2 Related Work Mihai Gavrilescu [1] and Champa H N [2] used deep feed forwards neural net- works for small datasets that are textual.This was proven to be successful in predicting personality.They used a 3 layered feed forward architecture on tex- tual data which is handwritten.Even though, this model that they used had handwritten features than just text, they proved it that MBTI personalities can be predicted using deep neural architectures. 3 3 Methodology 3.1 Dataset Description For this project, we used the Myers-Briggs Personality Type Dataset available on Kaggle 1.This data which is available on Kaggle was collected through the PersonalityCafe forum that provides a large number of people,their respective MBTI personality types and what they have posted. Details Count Number of instances(posts) 8,675 Number of unique attributes 16 Table 1: Details of the dataset. Every dataset also comprises of data attributes. Table 2 describes attributes of data. In case of supervised learning, clearly mention which attribute(s) would be considered as the labels. Data Attributes Brief Explanation Personality type Posts written by that personality type Table 2: Details of Data Attributes. 3.2 Description of attributes I - Introversion E - Extroversion S - Sensing N - Intuitive F - Feeling T - Thinking P - Prospecting J - Judging A combination of I or E, S or N , F or T and P or J gives us an MBTI personality. 3.3 Data Pre-processing 3.3.1 Data Proportion As shown in Figure 1 given below, the data is clearly not in proportion.The number of posts are higher for INFP than any other type. 1https://www.kaggle.com/datasnaek/mbti-type 4 Figure 1: Graphical Representation of available data. 3.3.2 Data Cleaning Since our project is strictly based on text, removal of URLs was necessary. We also removed all the NULL values. Next step was to remove common llers like \or" , \a", \the", etc. This we did using python's NLTK.In order to preserve the data, we replace the null values with the hyphen symbol. 3.3.3 Lemmatization We used imported WordNetLemmatizer from nltk.stem to lemmatize the text which means that infected forms of the same word are treated as one form of the root word(e.g. \running", \ran", \run" all become \run"). 3.3.4 Tokenization Tokenization is necessary. Here, we split the available text into words using python's Natural Language ToolKit (NLTK).We tokenized further nd the use- less words.To apply this, we needed bag of words. We have dened a set of useless words with nltk.stopwords to tokenize correct posts. 3.3.5 Bag of words We built bag of words by removing all the stopwords and punctuation marks in order to have only necessary data on which we can apply our machine learning algorithm. 5 3.3.6

Splitting Since each number of personality type has dierent number of posts,they must split accordingly. We have split the data into 2 parts. 80 percent is for training and 20 percent is for testing. 3.4 Proposed Methodology Figure 2: Flowchart Domain : Predictive analytics { Task of classication and prediction is the key 6 { Predicting/lling the information that is unknown { In our project, we are predicting the personality trait of a person Machine learning task : Supervised learning { Given data and associated target response, model is trained and then is used to predict the correct response for a new data. In our project, Data : post of user(post column in dataset) Target response : personality type of user (type column in dataset) New data : New user whose personality we want to predict Type of problem under supervised learning : Classication(multi-class classication) { Class : 16 personality type , a user is assigned one out of these class

Sources

Sidharth_Mtech_Thesis_CS_Batch14-16.doc | C Sharp (Programming...)

in such a situation, i understand that a strict disciplinary action can be undertaken by the con cerned authorities of the university and i shall abide by it. sidharth dhadra m t ech (cse).

https://www.scribd.com/document/318995310/Sidharth-Mtech-Thesis-CS-Batch14-16-doc

9%

1download_up63142_017607.pdf - PROJECT REPORT ON ...

In such a situation, I understand that a strict disciplinary action can be undertaken against me by the concerned authorities. Name: Alok Sharma Date:15 July, ...

https://www.coursehero.com/file/42400870/1download-up63142-017607pdf/

5%

Managerial Economics - 64, 68.docx - ECONOMIC CRISIS OF ...

This work is their original work to the best of my knowledge and has not been submitted anyw here else for the award of any credits / degree whatsoever.

https://www.coursehero.com/file/37062065/Managerial-Economics-64-68docx/

3%

Arun SHARMA | Dr | Indira Gandhi Delhi Technical University for...

arun sharma currently works at the department of information technology, indira gandhi del hi technical university for women. arun does research in big data, computer architecture an d software engineering. their current project is 'component-based systems'.

https://www.researchgate.net/profile/Dr_Arun_Sharma

3%

Indira Gandhi Delhi Technical University for Women(IGDTUW)

the university is recognised by university grants commission(ugc) of india under section 12b o f ugc act.taarangana- 2020, the annual cultural festival of indira gandhi delhi technical unive rsity for women. one week ict based short term course to be conducted in dept. of ece...

https://igdtuw.ac.in/

2%

What Does Your College Major Say About Your... | Daily Astrologie

namely, the test "measures" extraversion (e) or introversion (i), sensing (s) or intuition (n), thin king (t) or feeling (f), and judging (j) or perceiving (p). though you may not realize it, your perso nality type can influence your decision making, including what college major you choose.

https://blog.dailyastrologie.com/2019/10/01/%EF%BB%BFwhat-does-your-college-major-say-about-you r-personality/

1%

Фразовый глагол **RUN**, выражения и идиомы

Выражения с глаголом **run**, фразовый глагол **run**, основные значения глагола **run**, примеры с перево дом и комментариями.Глагол **run** – одно из многозначных, употребительных слов, используемых са

1%

мостоятельно или в составе выражений, фразовых глаголов.

https://langformula.ru/run/

Fine-Tune ERNIE 2.0 for Text Classification - Towards Data ...

Aug 1, 2019 - We have split the data into 2 parts in the 80:20 ratio. b) Create Test data. Place the data in the correct place. Create 2 folders.

1%

https://towardsdatascience.com/https-medium-com-gaganmanku96-fine-tune-ernie-2-0-for-text-classification-6f32bee9bf3c