

Club Recruitment Assistance using Personality Analysis

S Surya
20BCE1071

I Nilavan
20BCE1080

Sumegh S
GONUGADE
20BCE1100

G Nithish KANNA
20BCE1304

Vellore Institute of Technology, Chennai, India

ABSTRACT:

The goal of this project in Natural Language Processing (NLP) is to recommend the appropriate club department for an individual based on their personality traits. The project involves examining the language used in their social media posts and linking it to various personality traits. The NLP model will be taught using a dataset of position descriptions and personality traits to determine the most relevant department for a given student. To achieve this, we are using nlp techniques such as text preprocessing, feature extraction, and some machine learning algorithms will be used. Initially, the text data will be processed by removing stop words and special characters, and then tokenized and stored in a dataset. Subsequently, feature extraction will be employed to identify the crucial words and phrases in the position descriptions that are connected with specific personality traits. The machine learning algorithm like naïve Bayes, SVM will be trained on a labeled dataset of personality traits and position descriptions to anticipate the most fitting department for a given individual based on their personality traits. The model will be assessed using metrics like accuracy, precision, and recall. In summary, this project can aid club committees in selecting the most suitable department for a student's personality traits, resulting in a better fit and more suitable position within the club.

Keywords: NLP, Naïve bayes, SVM, Feature Extraction, Preprocessing, Labeling

1. INTRODUCTION:

The increasing number of clubs in colleges has created a challenge in the selection process for members and heads. With more options available, the competition to join a particular club has become tougher. Choosing suitable heads and members for various departments such as management, marketing, technical, finance, and cultural involves a meticulous assessment of applicants' skills and qualifications. Apart from academic requirements, the selection process also evaluates candidates' leadership abilities, communication skills, and dedication to the club's objectives. This process can be demanding and time-consuming for selection committees, potentially making it challenging to uphold a fair and transparent selection process.. To address this issue, we are introducing more comprehensive selection criteria with the help of natural language processing by providing additional support to the selection committees.

This document will also review existing research on personality trait analysis using NLP and deep learning techniques. And will provide an overview of the existing methods, their strengths, and their limitations. To achieve our goal, we will then collect data on the requirements and the personality traits of current members

in the department. This data will be used to create a baseline profile of the ideal candidate for the department. We will then analyze the descriptions of the candidates using NLP techniques to extract their personality traits which later mapped with the department. Next, we will use machine learning algorithms to compare the personality traits of the candidates with the baseline profile of the ideal candidate and as a output we will get the suitable department for the particular.

In summary, our NLP-based system will provide recruiters with a powerful tool to analyze candidates' personality traits and match them with the requirements of a particular department. Given approach can help clubs attract and retain talented individuals, and ultimately, contribute to a more vibrant and dynamic college community.

2. RELATED WORK:

In order to aid with our project, we widened our search and took in a variety of papers on personality type prediction. We found that personality analysis is done by various methods and for a wide range of purposes. The MBTI and NLP tools are applied to find the right combination of employees for a project team, with a given set of constraints [1]. The MBTI classification and machine learning models are applied in selecting ideal candidates for jobs in the IT industry [10]*. And the reverse is also accomplished given the personality of a person, the suitable career is suggested with the help of Deep Neural Network classifiers [13]*. With the help of our research we realized that many recent works on personality analysis extensively used Twitter dataset. The LIWC and SALLEE scores of tweets show heavy correlation with certain personality types [6]. The literature survey also helped us to know how to use a large dataset of texts in predicting personality types [12] and came to know that still machine learning algorithms are finding it difficult to perform well in classifying people into one of the MBTI types [4]*. But with binary classification such as classifying a person as extrovert or introvert they perform very well [2]. On the contrary Deep learning models perform very well, the LSTM model outperformed RNN and GRU [11]. The NLP architecture BERT found out among the twitter data biographies, statuses and liked tweets have high impact on MBTI personality types [14]. The use of emoticon count, mention count along with tf-idf score is also proven to be effective in personality classification [3]. We came across a data centric approach with NLP to predict the MBTI personality types [9]. And along with Twitter, Reddit also proves to be a significant data hub for personality prediction [15]. Most of the personality predictions rely on the correctness of the responses given by people to questions in the MBTI forum, but there is a possibility that the responses can be false. In order to tackle this, images were used and the task was automated with the help of Neural Networks [5]. There is also another classification of personality types like MBTI, which is the Big 5, but a comparative analysis revealed that MBTI trained algorithms show better performance [7]. Among these works which predicted the personality of real world people, an initiative was taken to predict the personality of fictional characters from the narrative description with the help of SVM and BERT [8]. The literature survey helped us to know what are the various techniques which can be used for personality classification. And it also made us realize the importance of personality prediction and automation of the same in future.

*: work is closely related to our project

3. BODY:

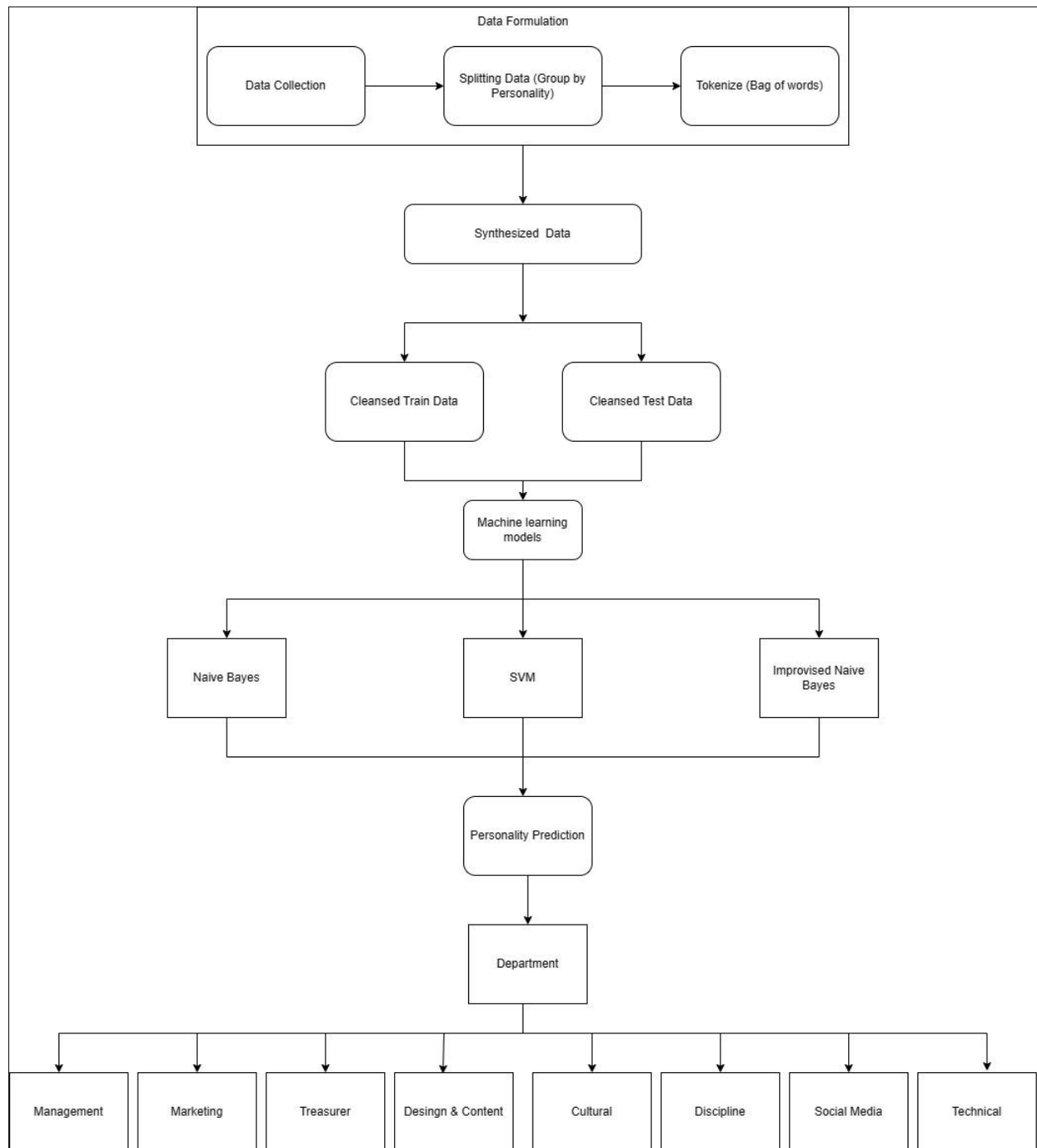
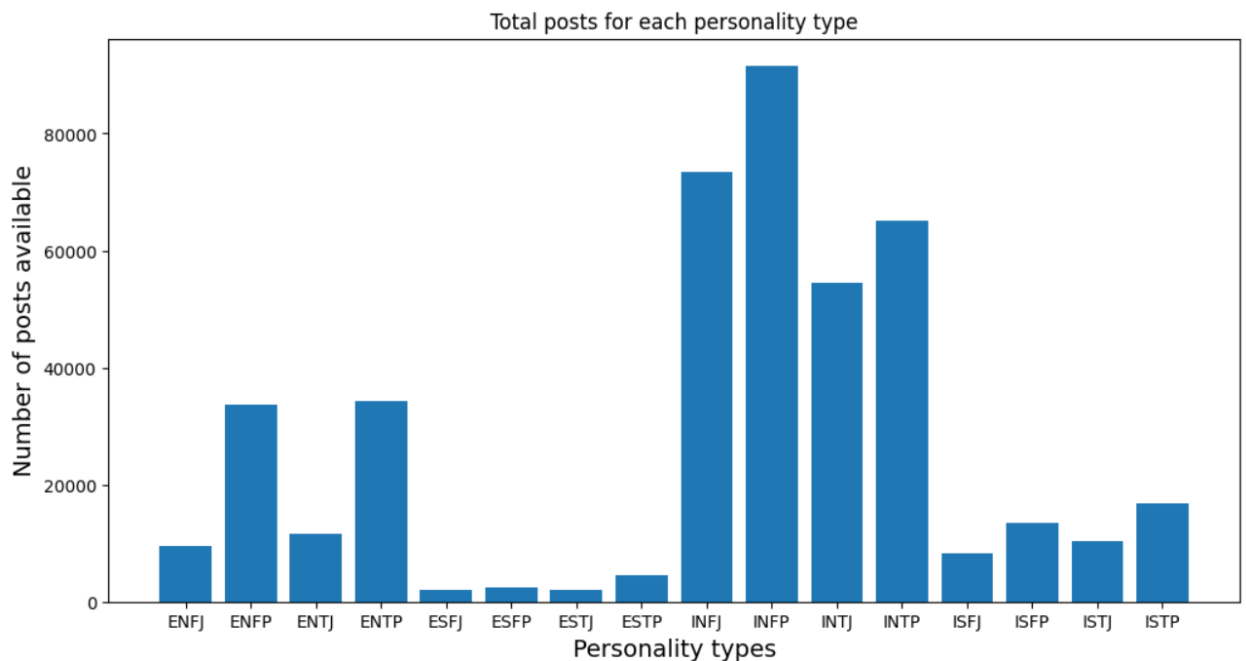


Figure 1 Proposed Methodology

3.1.Data Preprocessing

- Data Extraction: Data for this work is taken from Kaggle ([\(MBTI\) Myers-Briggs Personality Type Dataset | Kaggle](#)). The data consists of two features: Personality type and social media posts for 8000+ individuals.
- Data Organizing: The dataset consists of 50 posts per person clubbed together in a data frame. Each post was separated by '|||'. Further, data of each person is split and merged into one personality type. For the 16 different personality types in the dataset, posts are classified separately.



- Data Cleaning: Dataset contains no missing values. For each post individually, **build_bag_of_words** model is called to tokenize, which uses the NLP toolkit (nlk package) to remove the useless words. This function also removes punctuation.
- Data Splitting: Data is split into train and test data in 80:20 ratio.

3.2. Model Selection and Planning:

After Data-preprocessing, appropriate models for personality classification are selected:

3.2.1. Naive Bayes

NaiveBayesClassifier is used on the train data for training the model. The trained model gives an accuracy of 43.94% for train data and 34.76% for test data. So, this model performs poorly for this dataset.

3.2.2. SVM

LinearSVC is used on the train data for training the model. The trained model gives an accuracy of 57.9%. So, this model performs slightly better than the Naive Bayes model. For higher accuracy, an improvised Naive Bayes model is used.

3.2.3. Improved Naive Bayes

In this model, instead of selecting all 16 types of personalities as a unique feature, the dataset is explored and simplified.

The Myers Briggs Type Indicator (or MBTI for short) is a personality type system that divides everyone into 16 distinct personality types across 4 axis:

- Introversion (I) – Extroversion (E)
- Intuition (N) – Sensing (S)
- Thinking (T) – Feeling (F)
- Judging (J) – Perceiving (P)

This model will use this and create 4 classifiers to classify each person.

For each of the 4 classifiers, NaiveBayesClassifier is used and trained as 4 separate models.

3.3. Personality and Department Segregation

The tellmyMBTI() function takes three inputs: input, name, and traasits. The input argument is a list of strings, each string representing a question or prompt used to determine the user's MBTI (Myers-Briggs Type Indicator) personality type. The name argument is a string used to save the resulting plot as a PNG file with the given name. The traasits argument is an optional list that stores the user's MBTI personality type. The function then calls the MBTI() function on each item in the input list to determine the user's MBTI type. It then calculates the percentages of each of the four dichotomies (Introversion vs. Extraversion, Intuition vs. Sensing, Thinking vs. Feeling, and Judging vs. Perceiving) based on the user's responses. It determines the user's overall MBTI type by selecting the highest percentage from each dichotomy and concatenating the resulting four letters.

The dept() function takes a personality trait (as a string) as input and returns a string indicating which department the personality is suited for.

- Management Department: INFP, ESTP, ENFJ
- Marketing Department: INFJ, ENTJ, ENTP, ENFP, ESFJ
- Treasurer: ESTJ
- Design and Content Department: INTJ

- Cultural Department: ISTP, ESFP
- Discipline Department: ISTJ, ISFJ
- Social Media Department: ISFP
- Technical Department: INTP

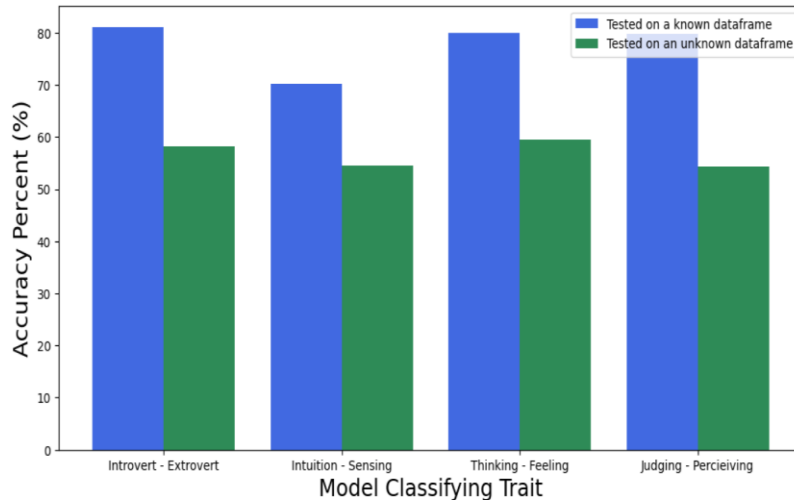
4. RESULTS AND DISCUSSIONS

For the SVM model used, metrics obtained are:

| | precision | recall | f1-score | support |
|------|-----------|--------|----------|---------|
| INFJ | 0.60 | 0.25 | 0.35 | 48 |
| ENTP | 0.52 | 0.38 | 0.44 | 143 |
| INTP | 0.59 | 0.38 | 0.46 | 53 |
| INTJ | 0.56 | 0.36 | 0.44 | 151 |
| ENTJ | 0.29 | 0.22 | 0.25 | 9 |
| INFP | 0.00 | 0.00 | 0.00 | 8 |
| ENFP | 0.00 | 0.00 | 0.00 | 9 |
| ISFP | 0.50 | 0.18 | 0.27 | 22 |
| ENFJ | 0.56 | 0.63 | 0.59 | 337 |
| ISTP | 0.58 | 0.70 | 0.64 | 398 |
| ISFJ | 0.58 | 0.53 | 0.55 | 230 |
| ISTJ | 0.61 | 0.65 | 0.63 | 285 |
| ESTP | 0.58 | 0.41 | 0.48 | 34 |
| ESFP | 0.28 | 0.42 | 0.33 | 45 |
| ESTJ | 0.38 | 0.55 | 0.45 | 40 |
| ESFJ | 0.47 | 0.61 | 0.53 | 75 |

For the split train and test data, the obtained results for the improvised Naive Bayes model are:

| | Introvert - Extrovert | Intuition - Sensing | Thinking - Feeling | Judging - Percieving |
|-------|-----------------------|---------------------|--------------------|----------------------|
| train | 81.124440 | 70.145242 | 80.034569 | 79.793411 |
| test | 58.204693 | 54.462623 | 59.413152 | 54.405496 |

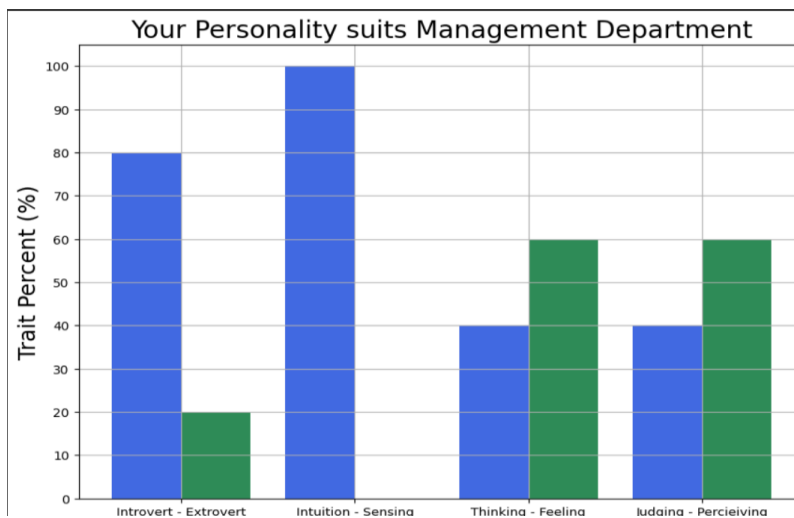


These accuracies obtained from the improvised Naive Bayes model are certainly better than the other two models.

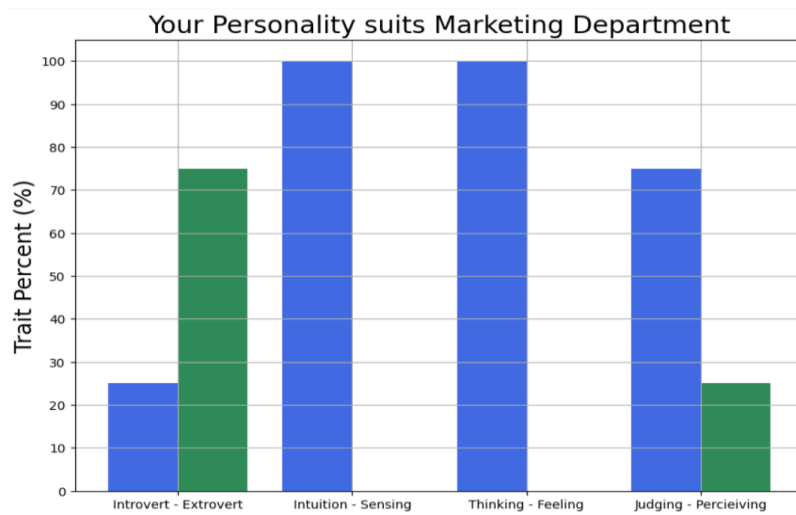
Hence, this model will be used for further processes (club selection), where each candidate's social media (Quora) posts are given to the model to predict the department that he/she is suited in that club.

We performed analysis for different persons with their Quora post and the results obtained are as follows:

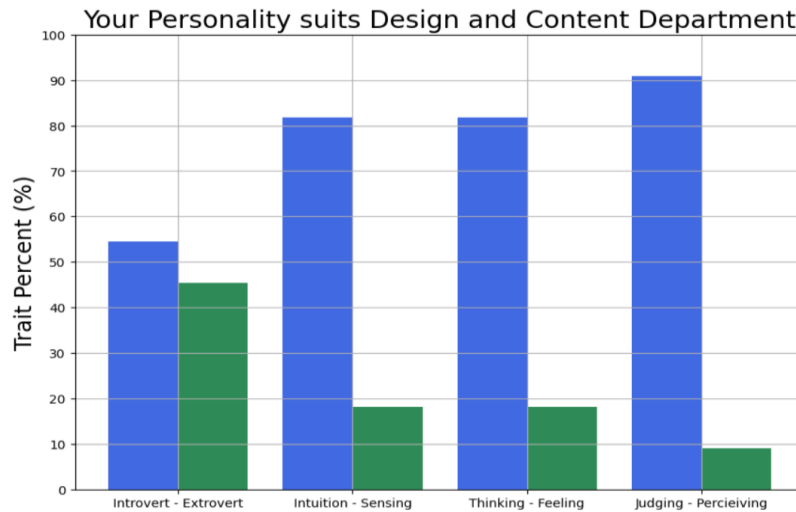
1. Harsha Bhogle:



2. Noah Kagan:



3. Ken Prewitt:



5. CONCLUSION

Basic classifiers like Naive Bayes and SVM didn't give expected results. Improved Naive Bayes Classifier provides significant results. Implementing these on a bigger dataset could produce astonishing results which might actually be much better than a human judging someone's personality. Currently, this model determines each trait with around 60% accuracy despite the limited dataset. This model could be useful for the clubs in VIT to recruit candidates who are fit and to assign the roles that are suitable for them with their social media posts.

Moreover, implementing something with a better learning curve like neural networks might help us determine a person's personality with better accuracy.

6. REFERENCE

- [1] Kollipara PB, Regalla L, Ghosh G, Kasturi N. Selecting project team members through MBTI method: an investigation with homophily and behavioural analysis. In 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP) 2019 Feb 25 (pp. 1-9). IEEE.
- [2] Kuchhal M, Jangid P, Saini M, Jindal R. Personality Prediction And Group Detection Using Social Media Posts. In 2022 IEEE 7th International conference for Convergence in Technology (I2CT) 2022 Apr 7 (pp. 1-5). IEEE.
- [3] Sahono MN, Sidiastahta FU, Shidik GF, Fanani AZ, Nuraisha S, Lutfina E. Extrovert and Introvert Classification based on Myers-Briggs Type Indicator (MBTI) using Support Vector Machine (SVM). In 2020 International Seminar on Application for Technology of Information and Communication (iSemantic) 2020 Sep 19 (pp. 572-577). IEEE.
- [4] Cui B, Qi C. Survey analysis of machine learning methods for natural language processing for MBTI Personality Type Prediction. Final Report Stanford University. 2017.
- [5] Pradhan T, Bhansali R, Chandnani D, Pangaonkar A. Analysis of personality traits using natural language processing and deep learning. In 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA) 2020 Jul 15 (pp. 457-461). IEEE.
- [6] Rathi S, Verma JP, Jain R, Nayyar A, Thakur N. Psychometric profiling of individuals using Twitter profiles: A psychological Natural Language Processing based approach. Concurrency and Computation: Practice and Experience. 2022 Aug 30;34(19):e7029.
- [7] Celli F, Lepri B. Is big five better than MBTI? A personality computing challenge using Twitter data. Computational Linguistics CLiC-it. 2018 Dec;2018:93.
- [8] Sang Y, Mou X, Yu M, Wang D, Li J, Stanton J. MBTI Personality Prediction for Fictional Characters Using Movie Scripts. arXiv preprint arXiv:2210.10994. 2022 Oct 20.
- [9] Basto C. Extending the Abstraction of Personality Types based on MBTI with Machine Learning and Natural Language Processing. arXiv preprint arXiv:2105.11798. 2021 May 25.

- [10] Pansare A, Panwar P, Kosamkar P. Personality Prediction with Natural Language Processing using Questionnaire Responses. In 2022 IEEE Pune Section International Conference (PuneCon) 2022 Dec 15 (pp. 1-6). IEEE.
- [11] Chen TY, Huang Y. Predicting MBTI Personalities based on Language Usage by Natural Language Processing Algorithms.
- [12] Plank B, Hovy D. Personality traits on twitter—or—how to get 1,500 personality tests in a week. In Proceedings of the 6th workshop on computational approaches to subjectivity, sentiment and social media analysis 2015 Sep (pp. 92-98).
- [13] Harrouk AI, Barbar AM. A psycholinguistic approach to career selection using nlp with deep neural network classifiers. In 2018 IEEE international multidisciplinary conference on engineering technology (IMCET) 2018 Nov 14 (pp. 1-6). IEEE.
- [14] Kadambi P. Exploring personality and online social engagement: an investigation of MBTI users on twitter. arXiv preprint arXiv:2109.06402. 2021 Sep 14.
- [15] Gjurković M, Šnajder J. Reddit: A gold mine for personality prediction. In Proceedings of the second workshop on computational modeling of people's opinions, personality, and emotions in social media 2018 Jun (pp. 87-97).