

# Correlation based data unification for personality trait prediction

Radha Divi<sup>1</sup>, Chandra Sekhar Potala<sup>2</sup>

<sup>1</sup>Department of Computer Science and Engineering, Malla Reddy College of Engineering and Technology, Scholar GITAM University, Hyderabad, India

<sup>2</sup>Department of CSE, GITAM, Visakhapatnam, India

## Article Info

### Article history:

Received Jun 12, 2022

Revised Sep 6, 2022

Accepted Sep 19, 2022

### Keywords:

Correlation  
Data consolidation  
Heterogeneous  
Neuroticism  
Personality traits  
Prediction

## ABSTRACT

The data points created by users with their online behavior are the primary source of data analysis in various research studies. One of such studies that even more contributes directly or indirectly to other application domains is personality trait analysis. There are well-defined models witnessed to be strong enough to characterize the individual personalities. Some prominent models are the Big Five, Interpersonal Circumplex, and DISC. To make it possible the model should always be fed with a huge amount of data. Here comes the limitation is that the user's online behavior is spread across various platforms resulting in various forms of data points. The segregated form of the data limits the model performance, which is the primary focus of the proposed work. Among these, the big five personality traits model has become so familiar because of its simplicity and dominance. The proposed work illustrated the mapping of various data forms namely measures of the source dataset (MBTI) and Interpersonal Circumplex data to Big five data.

*This is an open access article under the [CC BY-SA](#) license.*



## Corresponding Author:

Radha Divi

Department of Computer Science and Engineering, Malla Reddy College of Engineering and Technology  
Scholar GITAM University, Hyderabad, India

Email: radharavavarapu@gmail.com

## 1. INTRODUCTION

Earlier studies have witnessed how an individual personality can be drawn from his handwriting skills. This high dimensional online data availability has vanished the traditional approaches and forced us to come up with altered study models for human psychological behavior. As a result, The Big Five Personality trait model, the first-ever kind of psychological prediction model, has been introduced. Its main aim is to analyze the personality based on the responses given to the predefined psychology-based questionnaire [1]. To be more precise, the Big Five Personality Trait model [2], also known as the OCEAN model, identifies and maps the individual to one of the following assessment aspects [3].

- Openness: People with dynamic personalities, creative, attentive, and willing to take on more challenges come under this personality trait. Such people tend to be more curious and ready to explore new things.
- Conscientiousness: High levels of thought, carefulness, and discipline are the characteristics of this trait. They seem to be well organized and stick to the time in completing tasks.
- Extroversion: These people are too expressive and have good social behavior. They clearly show their excitement and enthusiasm and like to see the results quickly.
- Agreeableness: Cooperation, affection, trustworthiness, and kindness are the main characteristics of these people. They are considered to be generous and understanding.

- Neuroticism: These people are emotionally unstable and are highly prone to mood swings. High levels of this trait are irritable.

Human behavior is a complex interplay of actions, cognitions, and emotions, a deep understanding of it, will help in decision making, stress management, professional insights, setting and reaching goals, and emotion control [4]. The process of trait analysis has proportionately increased the complexity of the emerging technologies. Different sources will generate heterogeneous cross-domain and context-based data, however, mapping or consolidating such data to a uniform format well contributes to trait analysis [5].

## 2. RELATED WORK

Personality prediction is one of the prominent application areas in the field of individual counselling, personality trait detection, online marketing, recruitment agencies and many more [6]. This section presents the exiting work in the relevant field. Initially the personality prediction started along with sentiment on email data using Bayesian classifier [7]. Their work attempted to study the personality prediction on the social network data with a focus on email messages. Later it was extended to the correlation analysis over user's personalities and their respective online behavior on social media platforms.

Han *et al.* [8], explored the challenges in predicting the Big Five Personality scores using Linear Regression and Support Vector Regression. Their study is carried on Facebook posts and status to predict the personality. Another study given in [9], highlighted the impact of social media addiction on their behavioral pattern. Their questionnaire-based study among the teenage students witnessed how the social media influenced the individual empathetic concern and perspective talking.

Aung and Myint [10], has come up with a novel personality recognition model that works on personality lexicon. Their word embedding techniques could come up with a lexicon-based approach for personality recognition. Dandannavar *et al.* [11], investigated the influence on demographics and Big Five personality dimensions on social media. The word vector-based Bi-RNN model is developed for personality prediction based on word vector representation [12].

The CNN based another model is proposed Thomas *et al.* [13] and applying different activation functions to detect personality of the individual purely on the text data. Christian *et al.* [14] explored the provision for personality detection using the mobile technology, in which correlation and clustering methods identify extroversion and neuroticism using BIG5 model. Another work proposed in Ren *et al.* [15] attempted to identify correlation among personality and handwriting data in which Convolutional neural network applied to find the correlation among human handwriting and personality detection on BIG5 model.

## 3. MAPPING ILLUSTRATION

The Big Five Personality taxonomy was introduced in the 1980s with the basic idea of extracting the semantic associations among the words that actually describe the personality [16]. This has been carried out with the help of statistical-based factor analysis on personality survey data that gets mapped to one of the five dimensions namely openness, conscientiousness, extraversion, agreeableness, and neuroticism. To address this, the other sources of data are being considered to understand human behavior with respect to the context and content. As a consequence, dealing with this heterogeneous data under various contexts has become a challenging task. In this work, a mapping illustration is primarily focused on unifying various kinds of data gathered across different online platforms to the Big Five context model. Though accurate personality trait estimation is highly difficult from heterogeneous data, the advent of natural language processing (NLP) techniques gave a comfortable walkthrough to reach the goal [17]. This thought is extended in this work and experimented with mapping on various datasets as demonstrated below. Mapping illustration is given as follows. Firstly, an in-depth detail of the Big five model is presented. Secondly, the mapping process of each individual data with proper correlation of suitability is presented.

### 3.1. Big Five Personality Model

To begin with, a Big Five Personality Trait Model is chosen as a target model for data consolidation. In this, a sample questionnaire [8] is framed and the assessment procedure follows with a survey on a five-point scale of 1-disagree, 3-neutral and 5-strongly agree. Assuming no falsification from the participant will result in mapping an individual to one of its five dimensions which are presented along with their correlation analysis in the Table 1.

From the obtained values given in Table 1, it is clearly evident that all the five dimensions are uncorrelated with each other and thus can be considered as significant classes of personalities of human behavior. As shown in Figure 1, the datasets considered for fundamental mapping are MBTI dataset [18], Interpersonal Circumflex data [19], [20]. This way of consolidation would help the deep learning models in

increasing the data size and also to carry the assessment on context and content-based data. The more detailed description of individual data unification is given as follows.

Table 1. Correlation analysis on Big Five Dataset

	EXT	NEU	AGR	CON	OPN
EXT	1	-0.16	0.12	0.13	0.079
NEU		1	-0.089	-0.148	-0.047
AGR			1	0.134	0.018
CON				1	-0.027
OPN					1

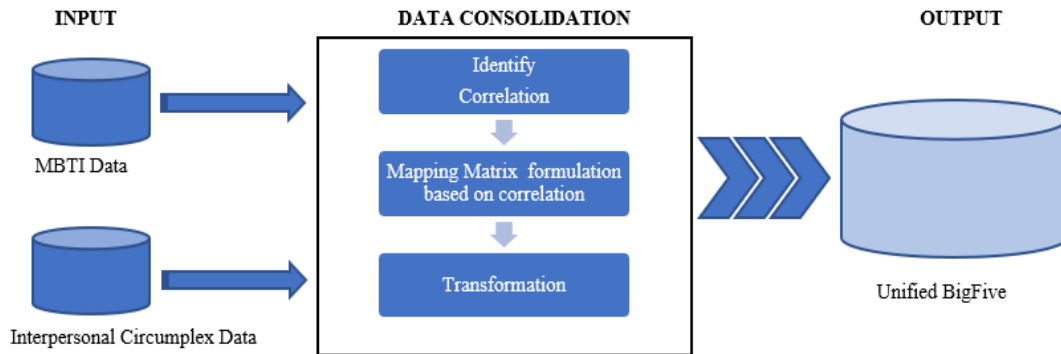


Figure 1. System flow

### 3.2. Myers briggs type indicator or MBTI

MBTI in short, is a personality type detection system that categorizes the individual across four axes, for example someone who prefers to be extrovert, sensing, thinking and perceiving will be labeled as ESTP. The consistent four axis are: i) Introversion (I) – Extroversion (E), ii) Intuition (N) – Sensing (S), iii) Thinking (T) – Feeling (F), iv) Judging (J) – Perceiving (P). The personality type of the individual is represented as four letter MBTI code, and the 16 possible combinations of these four-letter codes [21] are given in Table 2.

Table 2. MBTI - 16 Personality types with four-letter code

Sno	four-letter code	Sno	four-letter code
1	ISTJ	8	INFJ
2	ISTP	10	INFP
3	ESTP	11	ENFP
4	ESTJ	12	ENFJ
5	ISFJ	13	INTJ
6	ISFP	14	INTP
7	ESFP	15	ENTP
8	ESFJ	16	ENTJ

To ensure the uncorrelated or bipolar nature of four axes of MBTI, the correlation analysis is carried out on the experimental data. The linear correlation is extracted from the data using the Pearson correlation coefficient [22] whose values range from -1 to +1. To say that two sets are not correlated the respective correlation values should lie in -0.3 to 0.3, the values given in Table 3 clearly depicts that the pair of sets have no correlation and an individual personality can be categorized across these one of four axes.

Table 3. Pearson Correlation Analysis on MBTI

	IE	NS	TF	JP
IE	1	-0.046	-0.070	0.160
NS		1	-0.081	0.015
TF			1	-0.00447
JP				1

**3.3. Mapping procedure of MBTI to Big FIVE**

As mentioned earlier, the personality trait of an individual is represented as a four-letter code, and the proposed work attempted to reinterpret this code representation in correlation with the Big Five model. The experimentation is aimed to find the correlated attributes among MBTI and Big Five, and so map the MBTI codes to Big Five dimensions. The partial correlation table in the literature is given in table4 from the source [23].

It is required to have an overlap between the measures of the source dataset (MBTI) and the target dataset (Big Five) to proceed further with the mapping process. The correlation values given in Table 4, clearly provides a feasible solution to obtain either a negative or positive correlation among the two trait models. The Pearson correlation analysis between two is summarized in Table 5.

Table 4. Correlation of MBTI and Big five

MBTI / Big Five	E	I	S	N	T	F	J	P
Neuroticism	-.30	.31	.15	-.14	-.13	.12	.07	-.07
Extraversion	.71	-.72	-.28	.27	.00	-.00	-.13	.16
Openness	.28	-.32	-.66	.64	-.17	.13	-.25	.26
Agreeableness	-.02	.02	.01	-.00	-.41	.28	.05	-.06
Conscientiousness	.13	-.13	.10	-.13	.22	-.27	.46	-.46

Table 5. Mapping Compatibility with Pearson Correlation Coefficient on MBTI and Big Five

Big Five	MBTI	Correlation Score	Type of Correlation
Extroversion (EXT)	Extroversion(E)	0.7	Positive
	Introversion(I)	-0.7	Negative
Openness (OPN)	Intuition(N)	0.6	Positive
	Sensing(S)	-0.7	Negative
Agreeableness (AGR)	Feeling (F)	0.3	Positive
	Thinking (T)	-0.4	Negative
Conscientiousness (CON)	Judging(J)	0.4	Positive
	Perceiving (P)	-0.5	Negative
Neuroticism (NEU)	Introversion(I)	0.3	Positive
	Extroversion(E)	-0.3	Negative

**3.4. MBTI to Big five fusion analysis**

From the observations recorded in Table 5, it is apparently visible that two trait models are compatible with each other for the fusion or unification process. Each sample in MBTI is mapped to Big Five traits on the bases of existing positive or negative correlation. Basing upon this relative analysis, further the fusion process with the possible options of these two trait models is illustrated in Table 6.

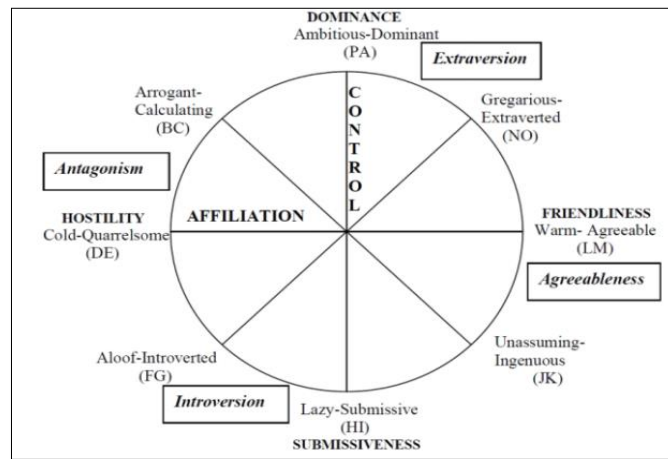
Table 6. MBTI 16 four-letter code classes to Big five dimensions

	EXT	OPN	AGR	NEU	CON
ISTJ	N	N	N	Y	Y
ISFJ	N	N	Y	Y	Y
INFJ	N	Y	Y	Y	Y
INTJ	N	Y	N	Y	Y
ISTP	N	N	N	Y	N
ISFP	N	N	Y	Y	N
INFP	N	Y	Y	Y	N
INTP	N	Y	N	Y	N
ESTP	Y	N	N	N	N
ESFP	Y	N	Y	N	N
ENFP	Y	Y	Y	N	N
ENTP	Y	Y	N	N	N
ESTJ	Y	N	N	N	Y
ESFJ	Y	N	Y	N	Y
ENFJ	Y	Y	Y	N	Y
ENTJ	Y	Y	N	N	Y

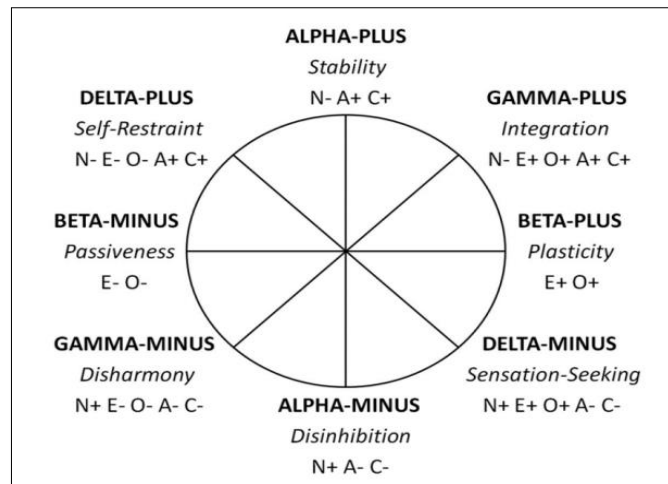
**3.5. Interpersonal circumflex data model**

In Figure 2 gives the details about another personality trait model namely, Interpersonal Circumflex or circle- a well-known conceptual model for assessing individual interpersonal traits. This circumplex is represented by two orthogonal axes [24]. The vertical axis signifies Status, Dominance, Power, Ambitiousness

where as horizontal axis meant for Agreeableness, Compassion, Solidarity, Friendliness, Warmth, Love. This interpersonal circumplex can be divided into two or four broad segments, or sixteen narrow segments [25], [26] shown in Figure 2(a). But most of the studies considered the circle into eight partitions/octants which implies a blend of the two axial dimensions. Each octet is spread across 45 degrees and their correlation angle subtend is presented in Figure 2(a), and the mapping of corresponding Big five traits are given in Figure 2(b). The correlation among the dimensions ranges will be figured as -1 for linear angles, 0 for orthogonal angles +1 for 00 and for the sample dimension “Dominance” with other measures at different angles is given in Table 7.



(a)



(b)

Figure 2. Details about another personality trait model (a) Interpersonal Circumplex to Big five Traits and (b) Correlation among dimensions Range

Table 7. Correlation summary of Interpersonal Circumplex

Angle subtend with dominance	Correlation
Gregarious Extraverted	0.5
Friendliness; Agreeableness	0
Unassuming-Ingenuous	-0.5
Submissiveness	-1.0
Aloof-Introverted	-0.5
Hostility/Coldness	0
Arrogant-Calculating	0.5

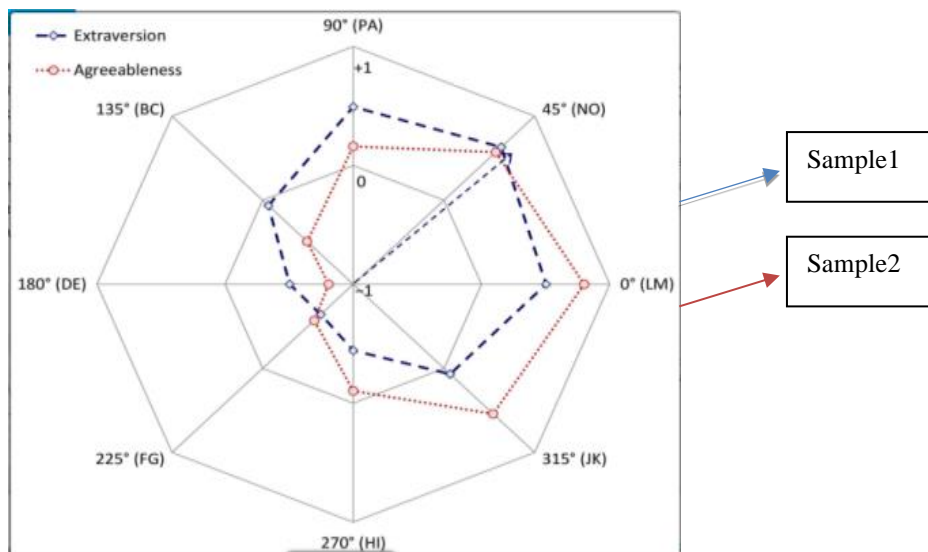
**3.6. Mapping from IC to Big Five**

From the understanding of blended concepts given in Figures 2(a) and (b), the relative mapping of IC to Big five model with angle subtend is given in Table 8. Using this mapping Table 8, any data sample on the interpersonal circumplex can be mapped to Big five.

**Table 8. Correlation of Interpersonal Circumplex angles to Big Five**

IC Angle \ Big Five	E	O	A	C	N
0 <sup>0</sup>	Positive	Positive			
45 <sup>0</sup>	Positive	Positive	Positive	Positive	Negative
90 <sup>0</sup>			Positive	Positive	Negative
135 <sup>0</sup>	Negative	Negative	Positive	Positive	Negative
180 <sup>0</sup>	Negative	Negative			
225 <sup>0</sup>	Negative	Negative	Negative	Negative	Positive
270 <sup>0</sup>			Negative	Negative	Positive
315 <sup>0</sup>	Positive	Positive	Negative	Negative	Positive
360 <sup>0</sup> /0 <sup>0</sup>	Positive	Positive			

Figure 3 gives a view on how a user profile can be shown on Interpersonal circumplex. Each of the intersection point over IC octants represents user behavioral towards that dimension. If the distance from circle center to the octet intersection point is high, that is towards circumplex then it implies a positive correlation and if the distance is less, it implies a negative correlation with that dimension. In order to map the IC to Big five, for any user profile, the octet with the first largest distance towards circumplex is considered and proceeds further for the second largest circumplex and so on until the complete mapping of all Big five-dimension set is obtained. Mapping or Consolidation of two samples given in figure 3 are presented in table 9. The existing personality trait prediction model has given less attention to multiple forms of data, because of which the behavioral analysis is restricted to a single vertical aligned to BIG Five data model [D12].



**Figure 3. Octet Representation of the interpersonal circumplex source**

**Table 9. Mapping of IC data sample to Big Five**

Big Five \ Largest Circumplex Angle	E	O	A	C	N	
Sample 1	45 <sup>0</sup>	Positive	Positive	Positive	Positive	Negative
	90 <sup>0</sup>			Positive	Positive	Negative
Final Mapping	-	Positive	Positive	Positive	Positive	Negative
Sample 2	0 <sup>0</sup>	Positive	Positive			
	45 <sup>0</sup>	Positive	Positive	Positive	Positive	Negative
Final Mapping	-	Positive	Positive	Positive	Positive	Negative

#### 4. CONCLUSION

The research focused on understanding various forms of psychology trait models and attempted to obtain a unifying approach. The primary advantage of this unification is to perform the data consolidation with an adequate mapping and thus to increase the sources of data. This way of rising the data volume from heterogeneous forms facilitates any model to have proper learning with optimal fitness avoiding the biased prediction due to falsification by the individuals while taking surveys. Taking this as a major focus, the proposed work concentrated on data unification or fusion or consolidation from heterogeneous psychology data models to the familiar Big five data model. For illustration, two forms namely MBTI and Interpersonal Circumplex have been chosen and proposed the suitable mapping procedures. As a future work, a context-based NLP learning model can be developed to experiment over this unified data and test for accuracy levels with the existing techniques.




#### REFERENCES

- [1] E. Ezpeleta, I. V. de Mendizabal, J. M. Gómez Hidalgo, and U. Zurutuza, "Novel email spam detection method using sentiment analysis and personality recognition," *Logic Journal of the IGPL*, vol. 28, no. 1, pp. 83–94, 2020, doi: 10.1093/jigpal/jzz073.
- [2] M. Gavrilesco and N. Vizireanu, "Predicting the Big Five personality traits from handwriting," *EURASIP Journal on Image and Video Processing*, vol. 2018, no. 1, p. 57, Dec. 2018, doi: 10.1186/s13640-018-0297-3.
- [3] T. R. Reddy, V. V. Bulusu, and P. V. Reddy, "A survey on author profiling techniques", *International Journal of Applied Engineering Research*, vol. 11, no. 5, pp. 3092-3102, 2016.
- [4] P. Howlader, K. K. Pal, A. Cuzzocrea, and S. D. M. Kumar, "Predicting facebook-users' personality based on status and linguistic features via flexible regression analysis techniques," in *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, Apr. 2018, pp. 339–345, doi: 10.1145/3167132.3167166.
- [5] L. Liu, D. Preoțiuc-Pietro, Z. R. Samani, M. E. Moghaddam, and L. Ungar, "Analyzing personality through social media profile picture choice," in *Proceedings of the 10th International Conference on Web and Social Media, ICWSM 2016*, 2016, pp. 211–220.
- [6] M. A. Rahman, A. Al Faisal, T. Khanam, M. Amjad, and M. S. Siddik, "Personality detection from text using convolutional neural network," in *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, May 2019, pp. 1–6, doi: 10.1109/ICASERT.2019.8934548.
- [7] M. Dalvi-Esfahani, A. Niknafs, Z. Alaedini, H. Barati Ahmadabadi, D. J. Kuss, and T. Ramayah, "Social media addiction and empathy: moderating impact of personality traits among high school students," *Telematics and Informatics*, vol. 57, p. 101516, Mar. 2021, doi: 10.1016/j.tele.2020.101516.
- [8] S. Han, H. Huang, and Y. Tang, "Knowledge of words: An interpretable approach for personality recognition from social media," *Knowledge-Based Systems*, vol. 194, p. 105550, Apr. 2020, doi: 10.1016/j.knosys.2020.105550.
- [9] K. Kircaburun, S. Alhabash, Ş. B. Tosuntaş, and M. D. Griffiths, "Uses and gratifications of problematic social media use among university students: a simultaneous examination of the big five of personality traits, social media platforms, and social media use motives," *International Journal of Mental Health and Addiction*, vol. 18, no. 3, pp. 525–547, Jun. 2020, doi: 10.1007/s11469-018-9940-6.
- [10] Z. M. M. Aung and P. H. Myint, "Personality prediction based on content of facebook users: a literature review," in *2019 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, Jul. 2019, pp. 34–38, doi: 10.1109/SNPD.2019.8935692.
- [11] P. S. Dandannavar, S. R. Mangalwede, and P. M. Kulkarni, "Social media text - a source for personality prediction," in *Proceedings of the International Conference on Computational Techniques, Electronics and Mechanical Systems, CTEMS 2018*, Dec. 2018, pp. 62–65, doi: 10.1109/CTEMS.2018.8769304.
- [12] J. Lee and N. Bastos, "Finding characteristics of users in sensory information: from activities to personality traits," *Sensors*, vol. 20, no. 5, p. 1383, Mar. 2020, doi: 10.3390/s20051383.
- [13] S. Thomas, M. Goel, and D. Agrawal, "A framework for analyzing financial behavior using machine learning classification of personality through handwriting analysis," *Journal of Behavioral and Experimental Finance*, vol. 26, 2020, doi: 10.1016/j.jbef.2020.100315.
- [14] H. Christian, D. Suhartono, A. Chowanda, and K. Z. Zamli, "Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging," *Journal of Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00459-1.
- [15] Z. Ren, Q. Shen, X. Diao, and H. Xu, "A sentiment-aware deep learning approach for personality detection from text," *Information Processing and Management*, vol. 58, no. 3, 2021, doi: 10.1016/j.ipm.2021.102532.
- [16] T. R. Reddy, V. V. Bulusu, and P. V. Reddy, "Profile specific document weighted approach using a new term weighting measure for author profiling", *International Journal of Intelligent Engineering and Systems*, vol. 9, no. 4, pp. 136-146, Nov 2016, doi: 10.22266/ijies2016.1231.15.
- [17] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: state of the art, current trends and challenges," *Multimedia Tools and Applications*, Aug. 2022, doi: 10.1007/s11042-022-13428-4.
- [18] M. B. Gurtman, "Exploring personality with the interpersonal circumplex," *Social and Personality Psychology Compass*, vol. 3, no. 4, pp. 601–619, Jul. 2009, doi: 10.1111/j.1751-9004.2009.00172.x.
- [19] V. Zeigler-Hill and T. K. Shackelford, *Encyclopedia of Personality and Individual Differences*. Cham: Springer International Publishing, 2020.
- [20] W. Strus and J. Ciecuch, "The circumplex of personality metatraits and the HEXACO model: Toward refinement and integration," *Journal of Personality*, vol. 89, no. 4, pp. 803–818, Aug. 2021, doi: 10.1111/jopy.12616.
- [21] I. Pervaz, I. Ameer, A. Sittar, and R. M. A. Nawab, "Identification of Author Personality Traits using Stylistic Features - Notebook for PAN", *Conference and Labs of the Evaluation forum CLEF 2015*, France, Sep. 2015.
- [22] T. R. Reddy, P. V. Reddy, M. Tejaswinic, N. Shravyad, and C. Upendrae, "A combined feature selection techniques and distance measures based approach for author profiling", *Design engineering*, vol. 7, pp. 7021-7040, July 2021.




- [23] C. G. DeYoung, Y. J. Weisberg, L. C. Quilty, and J. B. Peterson, "Unifying the Aspects of the Big Five, the Interpersonal Circumplex, and Trait Affiliation," *Journal of Personality*, vol. 81, no. 5, pp. 465–475, Oct. 2013, doi: 10.1111/jopy.12020.
- [24] E. K. Traupman, T. W. Smith, B. N. Uchino, C. A. Berg, K. K. Trobst, and P. T. Costa, "Interpersonal circumplex octant, control, and affiliation scales for the NEO-PI-R," *Personality and Individual Differences*, vol. 47, no. 5, pp. 457–463, Oct. 2009, doi: 10.1016/j.paid.2009.04.018.
- [25] K. D. Locke, "Development and validation of a circumplex measure of the interpersonal culture in work teams and organizations," *Frontiers in Psychology*, vol. 10, no. APR, May 2019, doi: 10.3389/fpsyg.2019.00850.
- [26] A. Furnham, J. Moutafi, and J. Crump, "The relationship between the revised neo-personality inventory and the myers-briggs type indicator," *Social Behavior and Personality: an international journal*, vol. 31, no. 6, pp. 577–584, Jan. 2003, doi: 10.2224/sbp.2003.31.6.577.

## BIOGRAPHIES OF AUTHORS



**Radha Divi**    working as an associate professor in CSE department, MallaReddy College of Engg and Technology. She was pursuing ph.d from GITAM university, vizag with the specialization of deep learning. Recently published a springer conference in deep learning. She can be contacted at email: radharavavarapu@gmail.com.



**Chandra Sekhar Potala**    working as an Associate professor in CSE department, GITAM university. He published so many papers in some of the reputed journals and lifetime member of CSI and ISPS. He got Best Faculty award from GITAM University. He can be contacted at email: Chandrasekhar.pothala@gitam.edu.