International Journal For Research In Advanced Computer Science And Engineering

ISSN: 2208-2107

Volume 09 Issue 04 December 2023

Available on-line at:https://ijracse.org/index.php/cse/issue/view/479

DOI:https://doi.org/10.53555/cse.v9i4.2266 *Received: 07-Oct-2023 Accepted: 10-Nov-2023*

PERSONALITY CLASSIFICATION OF TEXT THROUGH MACHINE LEARNING AND DEEP LEARNING: A REVIEW (2023)

Himaya Perera[1*], Lakshan Costa[2]

Corresponding Author:

Abstract

Personality classification from text is a very popular domain of research among the domain of natural language processing. Personality of an individual has been found to be a very important characteristic when analyzing an individual for a particular purpose. Especially in fields such as e-recruitment, personality is a determining factor of if an individual has a placement at a particular workplace. The author aims to explore various personality classifications such as 'TheBig Five' and the "Myer Briggs Type Indicator' and various approaches in which text classification when it comes to detecting personality. Both machine learning and deep learning approaches are examined, and their inner workings, benefits and limitations are detailed as well. We expect that this article will provide a thorough insight to personality classification of text by a numerous number of approaches.

Index Terms—*Machine Learning, Deep learning, Personality Classification*

ISSN: 2208-2107

I. INTRODUCTION

Personality is an individual's most defining characteristics that cause them to perceive,feel and behave in certain ways [1].

Each person has a unique constellation that defines the way in which they function as an individual. The word personality originates from the latin word 'persona' [1] and the history of personality has been studied for years and years by many psychologists. Your personality includes traits, your character and your temperament. A personality trait is a characteristic pattern that defines how one thinks, feels or acts. Personality traits will be on a varying scale with some traits being more dominant and other traits more submissive [2]. One's personality is influenced by many factors including heredity, experiences, parenting styles etc. An individual's personality will fluctuate the most when they are children. When an individual is an adult it is often clearer what a person's defined personality is and there is less room for influence and change compared to when one is a child [2].

The importance of one's personality when recruiting for a role has been brought to light by many researchers. A candidate's personality is a key factor that contributes to their job performance when hired to work in a specific company. Compatibility of a candidate to a company is often a direct indicator of how happy and motivated they will be while working and it is a factor that can be used to predict the amount of output this candidate is able to provide to the company [3]. Research shows that past work experiences is only able to predict job performance by 16% accuracy where else, by assessing a candidate's personality one is able to predict their job performance by around 78% accuracy [4].

Personality detections are generally based on a couple of popular personality models such as the Big Five Factor personality model [5] and the Myers-Briggs Type Indicator (MBTI) [6]. Further research is necessary to find better solutions for personality detection, this review aims to provide a thorough overview of the personality models and personality detection approaches currently used in today's world.

II. PERSONALITY CLASSIFICATIONS

Psychologists have attempted to nail down the exact number of existing traits within a personality. In history the amount of traits to have been existing within a personality has been estimated to be 40000 by Gordon Allport, 16 by Raymond Cattell, and even 3 by Hans Eysenck [7]. However, many psychologists believed that 4000 is too many and 3 is too few, hence the following personality classifications were born.

A. The Big Five

Researchers such as Tupes and Christal reanalysed the aforementioned data and found that the big five traits account for the data very well [5]. The Big Five are namely,

1) extraversion - Often described as sociability or extrovertness. Individuals who are on the high end of extraversion are extroverts and are very sociable people. Individuals who are on the lower end of extraversion are introverts and they find difficulty in socializing and tend to need a lot of time in solitude to recharge.
2) agreeableness - Often described as kindness, ability to understand and trust. People who are on the higher end of the scale of agreeableness are co-operative and while those on the lower end of the spectrum are more competitive.
3) openness - Can be described as creativeness and intrigue. Individuals who are on the extreme end of this trait are considered to be adventurous and creative and individuals who are on the lower end of the spectrum of this trait are believed to be more traditional and find it difficult to be creative or abstract.
4) conscientiousness - Is described as thoughtfulness, great impulse control, and goal orientedness. Often individuals who are very conscientious are super organized and well aware of details. Individuals who are on the lower end of conscientiousness are more scatterbrained when it comes to details and are significantly less organized causing them to miss deadlines or be tardy.
5) neuroticism - Often describes sadness and emotional Instability. Individuals who are high in neuroticism are moodier and experience anxiety, sadness etc. Individuals who are on the lower end of the scale tend to be more resilient and emotionally stable [7].

B. The Myers-Briggs Indicator

The Myers-Briggs Type Indicator (MBTI) personality was developed by Isabel Briggs Myers and her mother Katharine Briggs and is based on the theory of psychological types described by C.G Jung [6]. The MBTI personality consists of 16 personality types with the variance of the four dichotomies described below:

- Extraversion or Introversion (E or I) - Does the individual prefer the outer world or the internal word.
- Sensing or Intuition(S or N) - Does the individual focus on basic information or does the individual interpret and then add meaning as well.
- Thinking or Feeling (T or F) - When one is making decisions do they look at logic or do they think about the people and circumstances involved.
- Judging or perceiving (J or P) - When dealing with the external world does the individual prefer to have things decided or are they open to new information and angels.

ISSN: 2208-2107

When describing one's MBTI personality type each of the above dichotomies will be represented via one letter [6]. For example if an individual is classified as an ESTJ personality type: they are considered to be an extraverted individual that is sensitive, thinks as opposed to feel and is judgemental.
Once all of the above details are considered there will be an outcome of 16 MBTI personality types.

III. MACHINE LEARNING APROACHES FOR PERSONALITY DETECTION

There are many techniques in machine learning that can be applied for personality detection. Logistic Regression, Native byers etc. Some of the most popular methods are described below:

A. Linear Regression

Linear regression is commonly used for predictive analysis, and it mainly checks whether a group of factors can predict a dependent outcome, which of those factors are significant in the prediction process and how they impact the final outcome. [8] aimed to predict users personality traits based on the content filtered out from one's Facebook profile. The research runs linear regression on each personality factor and produces weights for each feature. Then the research uses M5 'Rules and Gaussian Processes to predict the score of a personality trait.

B. Logistic Regression

In NLP logistic regression is the foundation supervised ML algorithm which is used for classification. It can be used to classify observations into two or more classes. The main difference between naive byers and this algorithm is that it's a discriminative classifier while naive byers is a generative classifier. Generative models aim to have an understanding of what each class looks like whereas discriminative models only try to learn how to distinguish classes from each other. Logistic regression generally works better in an environment where larger datasets are present since it is able to find many correlated features [9].

C. Naive Bayes Classifier

Naive Bayes is based on the bayes theorem and In this classifier it represents a piece of text as a 'bag of words' which is an unordered set of words. The frequency of these words are considered regardless of their position [9]. Then these words are cross checked with features that are present in the classification categories. If most features present in the analyzed text are present in category b then this text will be categorized into category b. In summary Naive Bayes selects the category with the highest probability as the predicted category for the input data. [10] proposes a solution that uses Naive Bayes Classifier to classify text into personality types.

D. Random Forest Classification

Random forest is based on several decision trees which act as the pillar of this algorithm. First K data points are selected from the training data and then decision trees are constructed [11]. Once multiple decision trees are constructed the most prominent features are selected from the subset of features. And the decision trees are used for the classification task, for example this situation can be described like if each tree votes for the most prominent class and finally the class with the most aggregate votes will be the final classification. Random forest can handle many input variables and can detect the variables that are important in the classification. It is also run well with large databases and the trees/ forests generated can be saved for use in the future [12].
Some limitations of random forest are as follows. Sometimes two trees may have correlations that lead to an increase in the error rate [11].

E. K- Nearest Neighbour Algorithm (KNN)

This algorithm focuses entirely on keeping similar objects close to each other and works with feature vectors and class labels in a data set. This algorithm stores all cases and classifies new inputs by measuring similarity. In KNN the input text is represented in spatial vectors and texts with the most similarity is selected and finally the text is classified using K neighbors [11].
Limitations of KNN are that accuracy will depend on the quality of the data and if there is a large amount of data the classification process will be slow. It also requires high memory and because of this reason it may be quite expensive to train [11].

E. Support Vector Machines (SVM)

SVM is a supervised algorithm that figures out the best decision boundary among vectors. SVM divides the plane of vectors into different classes, it takes the training data and marks it into various categories and then based and then predicts the new input into those categories. [13] SVM requires preprocessing steps. Removing missing data, removing numbers and special characters and lower casing, tokenization, lemmatization and word vectorization [14].

IV. DEEP LEARNING APROACHES FOR PERSONALITY DETECTION

A. *Recurrent Neural Networks (RNN)*

RNNs are the basis of approaches such as Long Short Term Memory (LSTM). For classification tasks the input text is sent through the RNN word by word while generating a new layer at each step. FInally the token for the last text is used as a compressed representation of the entire input. This token is passed into a feedforward network that chooses the

probable class. We can also use a pooling function to gather the hidden states for each word and compile their mean to pass into the feedforward network instead of just the last token [9].

B. Convolutional Neural Networks (CNN)

CNN is popular for image classification tasks but has also been proved to be effective in NLP tasks like text classification. CNN treats the text like a 2D image, with the dimensions being the sequence of words and the features of the words and then apply convolutional filters to get the extracted features. This output is then connected to neural network for the classification process. Text classification from CNN includes embedding layer, convolutional layer, max pooling layer, and a fully connected layer. Some limitations of CNN include having fixed length inputs, limited understanding of semantics and only recognizing patterns associated with a particular class. CNNs have limited interpretability which makes it difficult to understand how the CNN arrived at its classification decision. CNNs also require large amounts of training data and have trouble handling temporal data [15].

C. LSTM

RNN can be difficult to work with in instances where distant information is critical for application functioning. To address this shortcoming LSTM is used. LSTM removes information that is no longer necessary and adds information that would later be useful. This is achieved by the inclusion of a context layer into the LSTM architecture as well as using 'gates', special neural units to control information flow. (implemented through extra weights) Some of the gates include forget gates, output gates, add gates etc [9].

D. Pre-trained Language Model (PLM)

PLMs are massive neural networks that are used for various natural language processing tasks. The work under the pre-train and finetune approach. These models are generally first pre-trained by using large text data and then are fine tuned to perform a specific language task.

Fluent speakers of a language are estimated to 100000 depending on the resources they were able to use to learn languages. This means the average child is expected to learn at least 10 words per day to achieve the above mentioned level of literacy by the age of 20. The bulk of this vocabulary growth is achieved as a by-product of reading. The importance of the above phenomenon is the ability to use this knowledge gained through years of learning to perform specific tasks long after it was initially learnt [9]. When relating to PLMs the above concept can be likened to the process of pretraining.

PLMs provide good consistency when compared to other models as well as accuracy [16]. They are very effective especially when a project faces low training time as well as less available datasets for domain specific tasks.Even with the above stated limitations PLMs are able to excel among other models in aspects such as consistency and accuracy which are highly important when conducting language related tasks. In neural networks when the parameters are too large they might not be fully trained by just their training data. Via the pre-training method these models have a better initial state [17].

Pretrained models are able to extract generalizations from large amounts of texts however to make practical use of the above generalizations interfaces need to be created from these models to downstream these applications. In fine tuning a set of application specific parameters are added on top of the existing model. Here labeled data is used to train these specific parameters. A variety of PLMs are described below.

1) BERT-

The Bidirectional Encoder Representations from Transformers (BERT) model is the start of a new era in natural language processing. BERT is regarded as the leading PLM in efficiency, universality and usability and it outperforms most of the other PLMs. BERT revolutionizes the relationship between specific NLP tasks and pre-trained word vectors. BERT is based on transformers as opposed to LSTM which is a one way language model(LM). This model pretrains bidirectionally on unlabelled text by conditioning on both left and right contexts in all of its layers [17].

Transformers handle distant information by mapping input vectors to sequences of output vectors. They are made up of transformer blocks which are each a network consisting of many layers that is a combination of linear, feedforward and self attention layers. Causal transformers that are (left-to-right) that are the foundation to some LMs that can be applied for domains such as translation, summarization etc. However LMs based on causal transformers tend to fall short in domains such as classification and labeling problems since they are based on incremental processing. Bidirectional encoders solve this limitation by letting self-attention mechanisms read over the entire input and not just the input on the left. This is the approach taken in BERT. One of the greatest strengths of BERT is that its upper and lower layers are directly connected to each other compared to RNN and LSTM, hence its more efficient and is able to collect longer distance dependencies. BERT consists of a vocabulary of 30000 tokens by using the Word-Piece algorithm, has hidden layers which are of the size 768, and has 12 layers of transformer blocks which has 12 multihead attention layers each [17]. This results in BERT having over 100M parameters and it is based on subword tokens rather than just words. Input sentences should be tokenized and then future processing of the input will take place on the token and not the words [9].

BERT follows two main pre-training tasks. Masked LM and Next Sentence Prediction. This is a novelty in PLMs. Masked LM uses a deep bidirectional model and this is different to other standard language models since they cannot be trained bidirectionally. Next Sentence Prediction is catered towards solving problems such as question answering etc. a next sentence prediction task is added to the model for it to understand the relationship within two sentences. Pre-training on this task can achieve around 97-98% accuracy [17].

The fine-tuning process involves adjusting the weights of the pre-trained BERT model using gradient descent optimization, to minimize the difference between the predicted personality traits and the true personality traits of the text samples in the training set.

2) RoBERTa

RoBERTa is a modification of BERT. RoBERTa is trained with significantly more data which leads to improved accuracy, and it is trained on longer sequences (up to 8K sequences). Next sentence prediction In BERT is trained to check whether there is a relation between consecutive sentences which removes the possibility of increasing downstream task performance. In BERT masking is done only once at the initial point of preprocessing, however RoBERTa duplicates training data ten times and each sequence is differently masked [18].

RoBERTa like BERT accepts sentences that have been encoded or transformed into tokens. These tokens along with the attention mask (describes token padding) and token type ids (binary value that indicates if two sentences are a pair) will determine the classification class. Once RoBERTa is pre-trained it can be finetuned to perform specific personality tasks such as personality classification. This can be done by adding several Keras layers to the outer layers and the use of various other layers such as flatten layer to reduce dimensions, dense layers to receive input, dropout layer to reduce overfitting etc. [18].

3) DistilBERT

DistilBERT attempts to optimize BERT by reducing its size and increasing its speed while retaining as much performance as possible. It uses similar architecture to BERT however with less encoder blocks and the removal of toke type embeddings and pooling functions. DistilBERT is not trained in next sentence prediction like BERT is and is only pretrained using masked language modelling. This model is also trained using three loss functions [19].

4) ALBERT

ALBERT like DistilBERT aims to reduce the size of the model but without a tradeoff in performance. This difference is structural. While DistilBERT trains from BERT in a teacher student manner, ALBERT is trained from scratch [20]. ALBERT can retain good performance with a smaller model by using parameter reduction techniques such as factorized embedding parameterization, cross layer parameter sharing, and removing dropout layers [20].

5) Flan-T5

Flan is a combination of a model and a network and is an improved version of the T5 language model developed by Google. This model works well with multiple tasks but namely translation and summarization. Flan T5 is better at all tasks than T5 and works on a wide range of NLP tasks [21]. The model is highly regarded for its speed and efficiency and is highly useful when it comes to real time applications. In training Flan-T5 is fed a large amount of data and is trained to predict the missing words in a piece of text. This process is done until the model has fully learnt to generate similar text to the input text. Text classification is one of the potential use cases of this model. Some limitations of this model include data biases, being resource intensive and sometimes-receiving unreliable output when met with new inputs. Specifically, its unreliability with new inputs makes it difficult to use in real time scenarios where accuracy is key [21].

6) GPT-3

GPT-3 is the successor to GPT-2 and was introduced by Open AI in 2020. With over 175 billion parameters, GPT-3 is believed to be the largest model so far [22]. However, a model of this size will also require massive amounts of data to train it. This model relies on transformer-based architecture that includes modified initialization, pre-normalization, reverse tokenization, and alternating sparse and dense patterns. With this model many tasks can be done without fine tuning because of its task agnostic nature. However, to have better accuracy with a specific task the few shots setting should be used with this model where like machine learning, inputs and corresponding outputs are provided to the model. Moreover, unlike machine learning models there will be no update to the weights of the model and will just provide a solution based on the ‘shots’ given. Typically, the above method will require between 10-100 shots for one specific task [22].

Limitations of GPT-3 include not performing well on text synthesis tasks. Furthermore, since GPT-3 unlike models such as BERT is not bidirectional, its more suitable for tasks that do not depend on fine tuning and are context level tasks. GPT-3 can also not be retrained or downloaded since it has “closed-API” access [22].

7) XLNet

XLNet is a BERT like model but is generalized with AR pretraining method. XLNet has permutation language modelling uses permutations of the occurrences for a specific word, it trains each word in a sequence and this takes much longer that

the process of BERT. This can be regarded as a limitation of this model. XLNet like RoBERTa conducts longer training by using larger batches of data and step sizes along with some changes in the masking procedure [23].

V. CONCLUSION

This paper is an attempt to explore the various ways in which personality classification can be done through text. The importance of detecting personality is highlighted especially in field such as e-recruitment etc. The two personality classification models 'The Big Five' and 'MBTI' are described and compared in this use case. The selection between these two classifications will depend on the requirements of how one's personality should be detected. This paper also elaborates on many techniques both using machine learning and deep learning for personality detection from text. It covers initial approaches used in text classification such as SVM and Naive Bayes and the newer approaches such as PLMs. The innerworkings of these different approaches are described as well as their strengths and weaknesses when assessing the task at hand.

REFERENCES

[1]. K. Dumper, "10.1 What is Personality? – Introductory Psychology," *opentext.wsu.edu*. https://opentext.wsu.edu/psych105/chapter/10-2-what-is-personality/
[2]. H. Gillette, "Are You Born with Personality or Does It Develop Later On?," *Psych Central*, Jan. 25, 2022. https://psychcentral.com/health/personality-development
[3]. D. H. M. Pelt, D. van der Linden, C. S. Dunkel, and M. Ph. Born, "The general factor of personality and job performance: Revisiting previous meta-analyses," *International Journal of Selection and Assessment*, vol. 25, no. 4, pp. 333–346, Nov. 2017, doi: https://doi.org/10.1111/ijsa.12188.
[4]. F. Schmidt, "The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 100 Years of Research Findings | Request PDF," *ResearchGate*, Oct. 2016. https://www.researchgate.net/publication/309203898_The_Validity_and_Utility_of_Selection_Methods_in_Personnel_Psychology_Practical_and_Theoretical_Implications_of_100_Years_of_Research_Findings
[5]. M. Barrick and M. Mount, "THE BIG FIVE PERSONALITY DIMENSIONS AND JOB PERFORMANCE: A META-ANALYSIS," *Personnel Psychology*, vol. 44, no. 1, pp. 1–26, Mar. 1991, doi: https://doi.org/10.1111/j.1744-6570.1991.tb00688.x.
[6]. Y. Cohen, H. Ornoy, and B. Keren, "MBTI Personality Types of Project Managers and Their Success: A Field Survey," *Project Management Journal*, vol. 44, no. 3, pp. 78–87, Jun. 2013, doi: https://doi.org/10.1002/pmj.21338.
[7]. K. Cherry, "The Big Five Personality Traits," *Verywell Mind*, Oct. 19, 2022. https://www.verywellmind.com/the-big-five-personality-dimensions-2795422
[8]. J. Golbeck, C. Robles, and K. Turner, "Predicting personality with social media," *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems - CHI EA '11*, 2011, doi: https://doi.org/10.1145/1979742.1979614.
[9]. D. Jurafsky and J. Martin, "Speech and Language Processing," *Stanford.edu*, 2018. https://web.stanford.edu/~jurafsky/slp3/
[10]. S. Katiyar, S. Kumar, and H. Walia, "Personality Prediction from Stack Overflow by using Naïve Bayes Theorem in Data Mining," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 3, pp. 1555–1559, Jan. 2020, doi: https://doi.org/10.35940/ijitee.c8601.019320.
[11]. K. Shah, H. Patel, D. Sanghvi, and M. Shah, "A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification," *Augmented Human Research*, vol. 5, no. 1, Mar. 2020, doi: https://doi.org/10.1007/s41133-020-00032-0.
[12]. M. Z. Islam, J. Liu, J. Li, L. Liu, and W. Kang, "A Semantics Aware Random Forest for Text Classification," *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, Nov. 2019, doi: https://doi.org/10.1145/3357384.3357891.
[13]. B. Y. Pratama and R. Sarno, "Personality classification based on Twitter text using Naive Bayes, KNN and SVM," *2015 International Conference on Data and Software Engineering (ICoDSE)*, Nov. 2015, doi: https://doi.org/10.1109/icodse.2015.7436992.
[14]. Y. Wahba, N. Madhavji, and J. Steinbacher, "A Comparison of SVM against Pre-trained Language Models (PLMs) for Text Classification Tasks," *arXiv:2211.02563 [cs]*, Nov. 2022, Accessed: Mar. 05, 2023. [Online]. Available: https://arxiv.org/abs/2211.02563
[15]. M. Amin and N. Nadeem, "Convolutional Neural Network: Text Classification Model for Open Domain Question Answering System." Accessed: Mar. 05, 2023. [Online]. Available: https://arxiv.org/pdf/1809.02479.pdf
[16]. Y. Elazar *et al.*, "Measuring and Improving Consistency in Pretrained Language Models," *Transactions of the Association for Computational Linguistics*, vol. 9, pp. 1012–1031, 2021, doi: https://doi.org/10.1162/tacl_a_00410.
[17]. J. Duan, H. Zhao, Q. Zhou, M. Qiu, and M. Liu, "A Study of Pre-trained Language Models in Natural Language Processing," *2020 IEEE International Conference on Smart Cloud (SmartCloud)*, Nov. 2020, doi: https://doi.org/10.1109/smartcloud49737.2020.00030.
[18]. R. Khusuma, W. Maharani, and P. H. Gani, "Personality Detection On Twitter User With RoBERTa," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 7, no. 1, pp. 542–553, Feb. 2023, doi: https://doi.org/10.30865/mib.v7i1.5598.
[19]. V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," *arXiv.org*, 2019. https://arxiv.org/abs/1910.01108

[20]. R. Soricut, Z. Lan, Research Scientists, and Google Research, "ALBERT: A Lite BERT for Self-Supervised Learning of Language Representations," *Google AI Blog*, Dec. 20, 2019. https://ai.googleblog.com/2019/12/albert-lite-bert-for-self-supervised.html

[21]. J. Jacob, "What is FLAN-T5? Is FLAN-T5 a better alternative to GPT-3? | exemplary.ai," *exemplary.ai*, Feb. 02, 2023. https://exemplary.ai/blog/flan-t5 (accessed Mar. 05, 2023).

[22]. T. Brown *et al.*, "Language Models are Few-Shot Learners," 2020. Available: https://arxiv.org/pdf/2005.14165.pdf

[23]. P. Rajapaksha, R. Farahbakhsh, and N. Crespi, "BERT, XLNet or RoBERTa: The Best Transfer Learning Model to Detect Clickbaits," *IEEE Access*, vol. 9, pp. 154704–154716, 2021, doi: https://doi.org/10.1109/access.2021.3128742.

[24]. Z. Ren, Q. Shen, X. Diao, and H. Xu, "A sentiment-aware deep learning approach for personality detection from text," *Information Processing & Management*, vol. 58, no. 3, p. 102532, May 2021, doi: https://doi.org/10.1016/j.ipm.2021.102532.

[25]. X. Sun, B. Liu, Q. Meng, J. Cao, J. Luo, and H. Yin, "Group-level personality detection based on text generated networks," *World Wide Web*, Sep. 2019, doi: https://doi.org/10.1007/s11280-019-00729-2.

[26]. X. Sun, B. Liu, J. Cao, J. Luo, and X. Shen, "Who Am I? Personality Detection Based on Deep Learning for Texts," *2018 IEEE International Conference on Communications (ICC)*, May 2018, doi: https://doi.org/10.1109/icc.2018.8422105.

[27]. Md. A. Rahman, A. Al Faisal, T. Khanam, M. Amjad, and M. S. Siddik, "Personality Detection from Text using Convolutional Neural Network," *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, May 2019, doi: https://doi.org/10.1109/icasert.2019.8934548.

[28]. C. Martin, "The Big Five OCEAN Personality Types: Introduction and Discussions," *blog.flexmr.net*. https://blog.flexmr.net/ocean-personality-types

[29]. N. Majumder, S. Poria, A. Gelbukh, and E. Cambria, "Deep Learning-Based Document Modeling for Personality Detection from Text," *IEEE Intelligent Systems*, vol. 32, no. 2, pp. 74–79, Mar. 2017, doi: https://doi.org/10.1109/mis.2017.23.

[30]. A. Kazameini, S. Fatehi, Y. Mehta, S. Eetemadi, and E. Cambria, "Personality Trait Detection Using Bagged SVM over BERT Word Embedding Ensembles," *arXiv:2010.01309 [cs]*, Oct. 2020, Available: https://arxiv.org/abs/2010.01309