# A Deep Learning Approach To Personality Identification

**Abstract.** Identifying personality traits is a crucial aspect of mental healthcare, which may be effectively achieved through integrating deep learning techniques within the field of psychology. Particularly, "personality" distinguishes an individual from others by describing feelings, thinking, and behaviors. Determining the correct personality proposes a good understanding of the individual and supports the psychologist in diagnosing and treating patients. In this paper, we developed a deep learning approach to identify personality from the MBTI (Myers-Briggs Type Indicator) Kaggle dataset. The model included three main components: convolutional layers to extract information, the LSTM layer that processes information in long-term dependency, and fully connected layers to classify the personality. As a result, our model outperformed the existing benchmarks of several studies in the 16-personalities classification task. Our models also led in performance in 2 over 4 binary personality dimension classification tasks. With these investigations, we recognized that our models could be applied to classify personality with better accuracy than several existing models.

**Keywords:** Deep Learning, Personality Identification, Psychology

## 1 Introduction

Personality has a crucial influence on processing information and making decisions. In psychology, it is defined as "a complex, dynamic integration shaped by many forces including hereditary and constitutional tendencies, physical maturation, early training; identification with significant individuals and groups, culturally conditioned values and roles; and critical experiences and relationships" [1]. Although the complexity of the dimensions of personality and disagreement in personality theories, it is an agreement that personality is an organization of some permanent patterns and distinct characteristics, so each person's behaviors are individual and consistent. Therefore, with strong theoretical backing, there are several models conducted to identify personality traits, including Myers–Briggs Personality Model (1920)[2], Eysenck's Three-Factor Theory (1963)[3], Keirsey Temperament Model (1984)[4], Five-Factor Model of Personality (1985)[5], HEXACO Six-Factor Model of Personality (2000)[6].

In the current research, the identification of personality has always been a crucial task among a vast of fields such as psychology [7], neural linguistic programming [8], behavioral finance [9, 10], and the business research domain [11, 12]. Especially with the development of AI and social media, understanding and identifying the individual personality becomes increasingly essential and important. The Myers-Briggs Type Indicator (MBTI) is the most commonly used to determine personality types among personality prevailing models. It has been reported that the Myers-Briggs Type Indicator model has been used by at least 115 countries, in 29 languages, and employed in 88 Fortune 100 companies during recent years[13]. Possible applications of personality identification could be in the recruitment process to select the candidates [14], or even Facebook has its own personality prediction system based on the interactions between users and its functionalities. MBTI has been applied in education to explore the different learning styles and teaching methodologies to develop an effective curriculum [15-17]. Thus, it is reasonable to conduct a study on the MBTI model, considering its popularity and widespread employment for personality identification.

Machine learning and deep learning have become one of the most noticeable emerging fields in recent years due to their successful applications in other areas. Deep CNNs, and RNNs models such as 1D Resnet, Bert, and so on have shown a powerful ability to extract information from text. On the other hand, machine learning techniques have long been recognized as traditional, reliable, and frequent tools that have been deeply studied during the last decades. Although both approaches are commonly used in personality prediction models, choosing the effective method depends on database characteristics, research questions, and research fields. For example, in social media research, it was observed that deep learning is more accurate than machine learning in personality prediction [18], while in psychology, machine learning methods were preferent methods [19].

One of the most studied MBTI datasets is the Myers–Briggs personality type dataset from Kaggle [20], containing 8675 rows of data. Considerable research has been conducted on this dataset. Most of them are based on supervised and conditional Machine Learning models such as Naïve Bayes [25,27], Logistic Regression [27], and XGBoost [26]. Some of them used Deep Learning [24,25,27]. Despite the efforts being made, since the data is highly imbalanced, most of them have been reporting limited accuracies on this dataset. In this study, we developed a personality prediction model based on this dataset by utilizing Natural Language Processing techniques with the core components based on the 1D convolutional layers and the Bidirectional LSTM layer. We, therefore, formulated the personality identification task as a text-based classification task and applied natural language processing to the raw data. The contributions of this paper are stated below:

1. We developed a novel Deep Learning based model for personal identification. We recognized that this is a novel model, and despite being simple in architecture, its performance is sufficiently good.
2. We further compared the developed model's performance with several existing studies. We recognized that the model outperformed existing studies in the 16 personality recognition tasks and outperformed machine learning approaches in other studies.
3. We employed the usage of focal loss to tackle the challenges of imbalance presented in the dataset. By ablation study, we further proved that the imbalance problem could be eased by using our approach.

The remaining of this paper is organized as follows:

In the second section, we reviewed and discussed related works to the current task. In the third section, we presented the methodology and experiment design. In the next section, we report the results and hold discussions. Finally, we conclude the paper with findings and investigations.

## 2 Related Works

Personality identification has long been studied. The possibly earliest study on personality identification was conducted using machine learning models [21]. They used personal information presented on Twitter users to predict the MBTI personality traits. In another study, Komisin and Guinn [22] employed the Naïve Bayes and Support Vector Machine (SVM) while learning users' word choice as a significant feature. The supported database is constructed from samples of writing collected from 40 graduate students along with their MBTI-identified personalities. Recently, there has been interest attended to the MBTI Kaggle dataset [20]. This Kaggle dataset was collected from the Personality Cafe forum in 2017 [23], containing posts written by 8675 people whom the MBTI type identified from this online social forum platform. Simple statistics on this dataset revealed the average of words by each post is 1226 with a standard deviation of 311 words. It is also interesting to note that this dataset provides a classification of human personalities in 16 ways via four dimensions, namely, (1) introversion/extroversion (I/E), (2) sensing/intuition (S/N), (3) thinking/feeling (T/F) and (4) judging/perceiving (P/J). **This also implies 5 possible tasks that can be studied on this dataset: a 16-personalities classification task (formulated as a multiclass classification) and 4 binary classification tasks for each dimension**. We selected this dataset for our study since it is publicly available, easy to access and due to its prior utilization in existing research. This allowed us to establish a comparison with benchmarks later.

Researching personality type prediction from textual data, especially when conducted on this Kaggle MBTI dataset, is challenging. However, machine learning and deep learning techniques have carried out this endeavour. Most recently, Yash Mehta and his colleagues [24] considered traditional psycholinguistic features and language model embeddings as features for prediction and employed a BERT-based and Multilayer Perceptron to predict 4 binary classification tasks (I/E, N/S, T/F, P/J). On the other hand, a study [25] employed both machine learning approaches and recurrent neural networks (RNNs) to employ all 5 possible classification tasks on this dataset. Naïve Bayes and SVM are selected with an 8-layer RNN to compare the performance. In addition, research [26] used Extreme Gradient Boosting (XGBoost), which combined a sequence of weak learners fitted onto data and predictions from them to form a weighted majority vote. They also presented a detailed study and benchmarked on 4 binary classification tasks (I/E, N/S, T/F, P/J).

We employed these 3 studies as our pool of existing benchmarks to compare with. In detail, they are representatives of the 3 most common approaches in this task: (1) the traditional machine learning approach such as Naïve Bayes [25] or SVM [25], (2) the model ensemble approach like XGboost [26], and (3) the deep learning approaches such as RNN [25] or BERT [24].

In the natural language processing (NLP) field, LSTM (Long Short-Term Memory), which is a popular variant of Recurrent Neural Network (RNN), is usually used for text classification tasks. Its key strength lies in its ability to capture long-term dependencies among words in a sentence, making it well-suited for tasks requiring contextual understanding. LSTM is also less susceptible to the vanishing gradient problem associated with traditional RNNs by selectively retaining and discarding information over time. A more advanced variant of LSTM is the Bidirectional LSTM which can leverage both past and following contexts to learn better representations. Bidirectional LSTM is particularly useful in NLP tasks where the meaning of an entity depends not only on the preceding but also on the following context. Taking inspiration from these insights, we designed our architecture utilizing a combination of Bidirectional LSTM and Convolutional 1D layers. It is worth mentioning that our model shares some similarities to the method outlined in [25], with the sole deviation being the exclusion of Dropout layers from our design. Furthermore, a notable contrast between our approach and the one proposed in [25] is the non-utilization of Focal Loss and ELU activation in their model.

# 3 Methodology and Experiment Design

## 3.1 Dataset Overview

The Kaggle MBTI dataset is publicly available and contains 8675 rows of data. The data collection process begins as users foremostly complete a questionnaire or survey, allowing professionals to identify their MBTI type. Each entry in the dataset is given as a tuple of the corresponding user's MBTI type and a section from their latest 50 posts separated by 3 pipes ("|||") characters. Four different dimensions were available inside each MBTI label. For instance, if the MBTI type label is ITSJ, we can conclude the first personality dimension of this person is Introvert (I), the second, third, and fourth dimensions are Thinking (T), Sensing (S), and Judging (J). We visualized the plot of occurrences of 16 MBTI labels in Figure 1 while reporting the distributions across 4 personality dimensions in Figure 2.

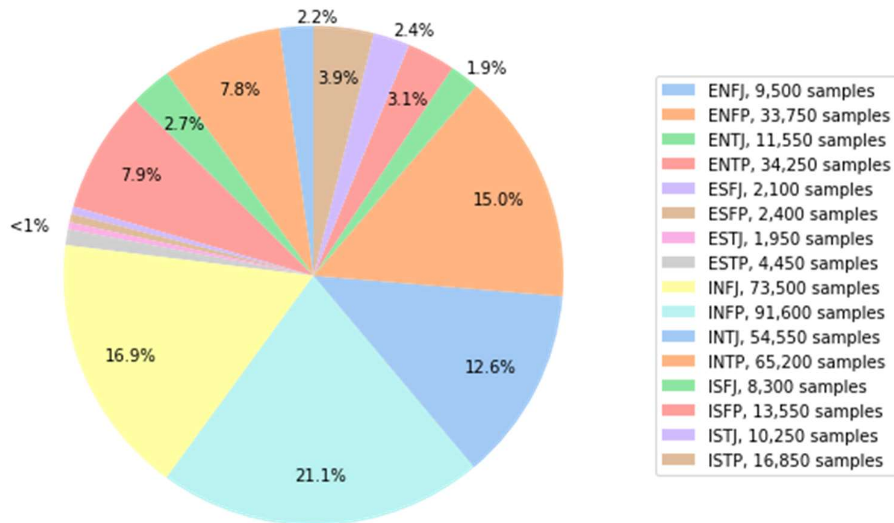


**Fig. 1.** Distribution of 16 MBTI labels.

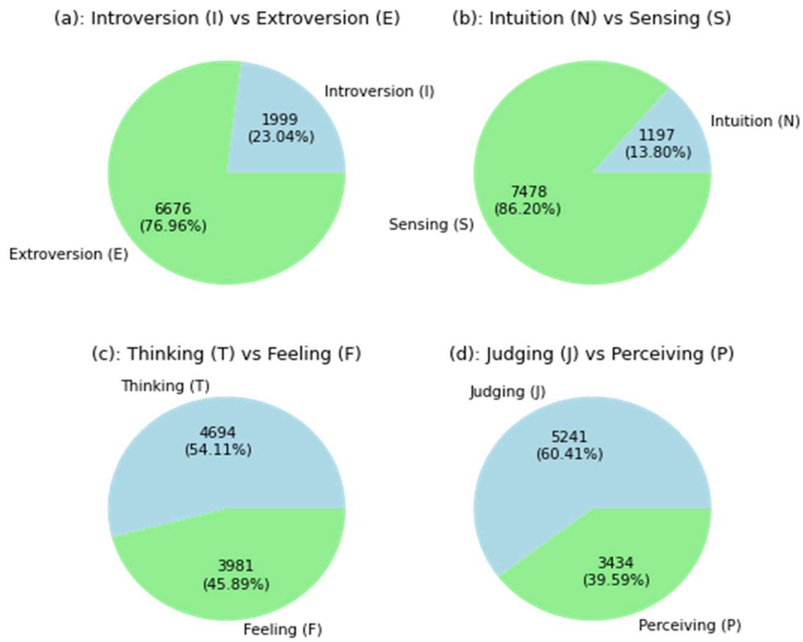


**Fig. 2.** Distribution of the dataset across four personality dimensions, namely (a) Introversion-Extraversion, (b) Sensation-Intuition, (c) Thinking-Feeling, and (d) Judging-Perceiving. Each dimension is represented by a pie chart that illustrates the proportion of the dataset falling into the two categories of each dimension.

### 3.2 Data Pre-processing

It can be observed that this dataset is highly imbalanced in label distribution. Besides, the post entries are raw and unprocessed. Therefore, sufficient processing steps should be taken to tackle these issues. The pipeline (Figure 3) for text processing consists of 4 main steps: standardization, lemmatization, tokenization, and word embeddings. We first standardize the content of the posts by removing hyperlinks, punctuations, and stop words. At the end of this step, we make the text become low case to avoid the capital case sensitivity introduced to the model. Lemmatization refers to the process of converting a word to its root form. For instance, words like "runs", "ran", and "running" are converted to "run". The motivation behind this step is to avoid grammar affection introduced to the model later. Processed text up to this step is further separated into smaller chunks called tokens. The final processing step involves word embeddings, where we encoded each word into a vector of real values such that the words that are closer in the distance in the vector space are expected to be more likely similar in definition or meaning. For the word embeddings, we employed the pre-trained word vectors called wiki-news-1M. These pre-trained vectors contain 1-millionword vectors trained on Wikipedia 2017. After pre-processing, we visualized the most frequent words based on occurrences as a word cloud image in Figure 4.

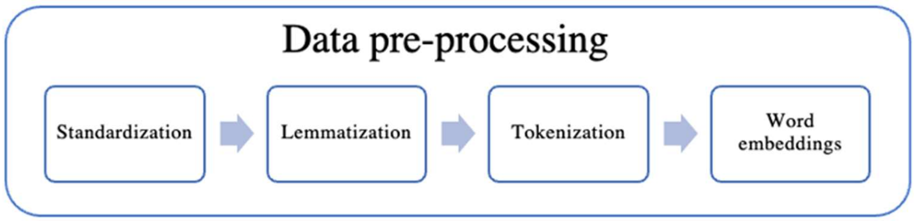


**Fig. 3.** Data pre-processing pipeline.

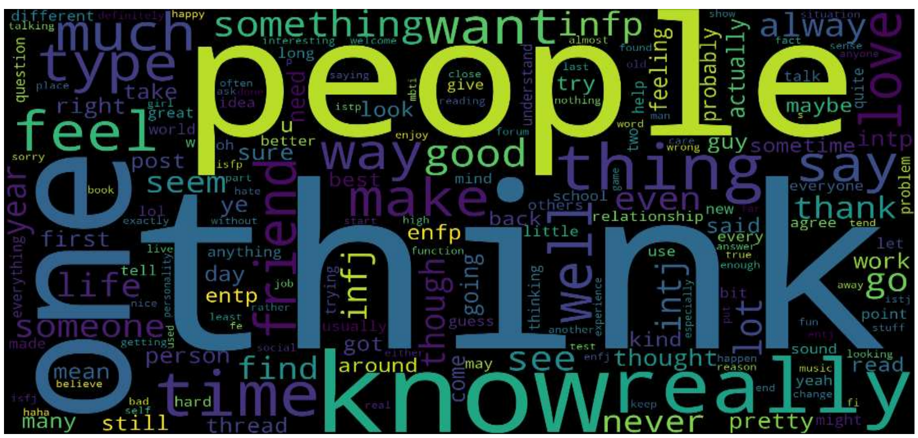


**Fig. 4.** Word cloud of most popular words in the dataset.

### 3.3 Model Development

In this section, we constructed a novel deep learning network utilizing 1D convolutional layers as a lower-level feature extraction block while using bi-directional LSTM as a means to model the dependencies inside text sequences and extract useful higher levels of features. A simple dense layer is employed to perform the classification task. We visualized the architecture in detail in Figure 5. Upon completion of the pre-processing stage, the raw textual input undergoes a transformation process whereby it is converted into a word embedding vector that functions as the input for our model. The model architecture comprises three convolutional blocks, each comprising a 1D convolutional layer endowed with 3x3, 4x4, and 5x5 kernel sizes, respectively, alongside 128 channels. These convolutional layers are subsequently followed by a max pooling layer that selectively retains the most prominent features. The resulting convolved data is then propagated to a Long Short-Term Memory (LSTM) layer, which is capable of processing sequential data while effectively capturing long-term dependencies. Finally, a set of linear layers is utilized to classify the input data into the corresponding output classes. This design enables the model to effectively learn and leverage the relevant features in the input text, thereby enhancing its predictive performance.

Defining the loss function and the learning rate is vital when training a deep neural network. From the dataset, we observed that there existed an imbalance in sample numbers between different classes in the dataset. Thus it is essential to formulate appropriate training strategies to tackle this issue. In this study, we employed Focal loss as the objective function. The formula of the focal loss is shown in Equation 1.

$$FL(p_t) = -\alpha_t * (1 - p_t)^{\gamma} * \log{(p_t)} \quad (1)$$

Where $p_t$ denotes the predicted probability of the ground-truth class, $\alpha$ and $\gamma$ are hyperparameters of the loss function.

Every model's rate of updating the weights depends on the learning rate. A learning rate too low can decelerate the convergence speed of the model, while a learning rate too high can overshoot the minimum loss values, thereby causing convergence failure. Therefore, we adopted a learning rate scheduler technique. Based on the validation loss values, the learning rate is adjusted so that the learning rate is halved unless a new minimum occurs after certain epochs.

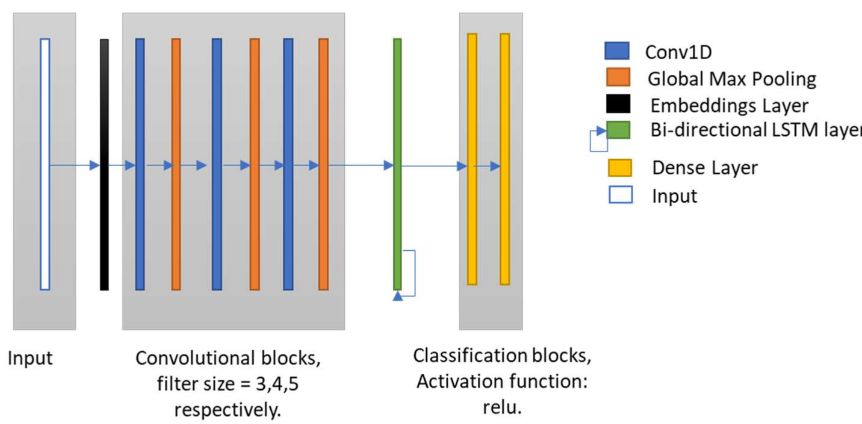


**Fig. 5.** Model Architecture.

### 3.4 Experiment Design

In this section, we described the contrast experiments conducted in this paper. In the aforementioned section, we compare our approaches to existing studies conducted by

Sakdipat Ontoum [25], Amirhosseini MH [26], and Yash Meta [24]. We employed the same training-splitting strategies mentioned in both papers [25,26] and repeated the experiment 3 times to allow consistency in evaluation. We also compared the model's performance to related studies such as Yash Meta [24] and the mentioned benchmarks.

In this study, we utilized accuracy as the primary performance metric. Our rationale for selecting accuracy as the primary performance metric for our model was informed by the prevailing use of this metric in several prior studies on personality identification using machine learning and deep learning approaches [24,25,26,27,28]. In addition, two of the three benchmarking papers [24,26] employed accuracy as the sole performance metric, which further solidified our decision to prioritize this measure. While paper [25] reported precision, recall, and F1 scores, it is noteworthy that these figures were limited to only four binary classification tasks and were unavailable to the 16 MBTI personalities classification task. Moreover, given that the remaining benchmarking papers did not furnish these metrics, we maintain that accuracy serves as the most appropriate and consistent performance metric for evaluating our model in this study. A set of hyper-parameters that are used in the training phase are recorded in Table 1 to support the reproducibility.

**Table 1.** Model Configuration.

| Hyper-parameters | Values |
|---|---|
| Number of epochs | 20 |
| Batch size | 256 |
| Conv1D Channels | 128 |
| Conv1D kernel size | 3/4/5 |
| Activation function | ELU |
| Max Pooling kernel size | 2 |
| LSTM hidden size | 128 |
| The dense layer's hidden dim | 128 |
| LSTM batch first | True |

## 4 Results and Discussions

We reported the accuracy of our model on the 16-classes classification task in Table 2, together with the benchmarks available in the existing studies. The table includes our model's average performance, estimated as the average accuracy of repeating the experiment 3 times with different seeds. In addition, the best versions of our models for each task are reported as in the line "best version". Table 3 provides a comparative analysis of our proposed models and existing benchmark approaches on four binary classification tasks.

**Table 2.** Comparison between the performance of our models with the performance of machine learning and recurrent neural network models in the study [25].

| Model | Overall Accuracy |
|---|---|
| Naïve Bayes [25] | 0.4103 |
| SVM [25] | 0.4197 |
| RNN [25] | 0.4975 |
| Our model (average performance) | **0.5334** |
| Our model (best version) | **0.6168** |

**Table 3.** The comparison between our models and other existing benchmarks on 4 binary classification tasks.

| Model | Accuracy on each task | | | |
|---|---|---|---|---|
| | I/E | N/S | F/T | P/J |
| Naïve Bayes [25] | 0.7828 | 0.8695 | 0.8063 | 0.7478 |
| SVM [25] | 0.8215 | 0.8732 | 0.8049 | 0.7270 |
| RNN [25] | 0.8359 | 0.9322 | 0.8000 | 0.7740 |
| XGBoost [26] | 0.7901 | 0.8596 | 0.7419 | 0.6542 |
| Bert + MLP [24] | 0.788 | 0.863 | 0.761 | 0.672 |
| Our model (average performance) | 0.8007 | 0.8906 | 0.8217 | 0.6690 |
| Our model (best version) | **0.8321** | 0.9077 | **0.8360** | 0.7337 |

After observing the reported results, we recognized that for the 16-classes classification task, our model outperformed machine learning approaches and deep recurrent neural networks [25]. We recognized that although our approaches were shallower compared to his network, our model still performed better. We argued that this is thanks to the application of focal loss. In an ablation study, we re-run the experiments in Table 1 with the focal loss replaced by the conventional cross-entropy loss. The average performance is reduced to 0.4229, much lower than the current performance. Observing from Table 2, our models are the best performer in 2 tasks. For 2 other tasks, our best models still achieved the second-highest ranking.

It is interesting, yet in our expectations, that the developed models outperformed traditional machine learning approaches. Surprisingly, the ensemble approach XGBoost is also surpassed by our model. One noticeable thing is that despite not using a well-known pre-trained network, our approach also outperformed the BERT+MLP model mentioned in[24].

From Table 2, the slight fluctuation between the average performance and the best model versions indicated that our approaches are consistent. Due to the limited resources, we cannot perform 10-fold cross-validation or repeat the experiment more than 3 times; however, given this consistency, we shall expect that repeating the experiment more times should also yield similar performance with slight fluctuations.

We recognized that there are several limitations in this study. Firstly, we could not repeat the experiment more times or with K-fold cross-validation (K=5,10…) due to insufficient resources and time. The second limitation is that we can still not tackle the challenges of flexible input inside the network. Specifically, we implicitly set a fixed maximum length for each text sequence and padding sequences with a length less than the fixed number. In practice, this can result in losing the nature of languages. It would be more convenient to find a method to make the network accept flexible-length text sequences to secure the nature of languages as much as possible.

## 6 Conclusion

In conclusion, in this paper, we have presented an approach for personality identification using deep learning. The proposed model is built upon convolutional layers and a bi-directional LSTM layer. Through experiments, we showed that our model outperformed several existing benchmarks in different studies. We also recognized the limitations of this study and would consider future experiments to improve them.

## Conflicts of Interests

The authors have no conflict of interest to declare.

## Appendix

**Table 4.** An ablation study of our model with different setups on the MBTI Kaggle dataset on the 16 personalities classification task.

| Model | Overall Accuracy |
|---|---|
| Our model without using Lemmatization | 0.4227 |
| Our model without using Word Embeddings | 0.2570 |
| Our model without using Focal Loss | 0.5182 |
| Our model (average performance) | 0.5334 |
| Our model (best version) | 0.6168 |

**Table 5.** The GPU and CPU configurations to train the models.

| Component | Value |
|---|---|
| VRAM | 12GB |
| System memory | 32GB |
| Number of CPU cores | 16 |
| Number of CPU threads | 32 |

## References

1. *APA dictionary of psychology, 2nd ed*. APA dictionary of psychology, 2nd ed., ed. G.R. VandenBos. 2015, Washington, DC, US: American Psychological Association. xv, 1204-xv, 1204.
2. Briggs, K.C., *Myers-Briggs type indicator. Form G*. 1987: Palo Alto, Calif. : Consulting Psychologists Press, [1987] ©1987.
3. Center, D., The Three Factor Theory of Personality. 2006.
4. Keirsey, D., Please understand me II: Temperament, character, intelligence. 1998: Prometheus Nemesis Book Company.
5. McCrae, R.R., *The Five-Factor Model of Personality: Consensus and Controversy*, in *The Cambridge Handbook of Personality Psychology*, G. Matthews and P.J. Corr, Editors. 2020, Cambridge University Press: Cambridge. p. 129-141.
6. Lee, K. and M.C. Ashton, *HEXACO Model of Personality Structure, The*, in *Encyclopedia of Personality and Individual Differences*, V. Zeigler-Hill and T.K. Shackelford, Editors. 2020, Springer International Publishing: Cham. p. 1932-1936.
7. Borghans, L., et al., *Identification problems in personality psychology.* Personality and Individual Differences, 2011. **51**(3): p. 315-320.
8. Cordell, A., Neuro-Linguistic Programming (NLP), in The Negotiation Handbook. 2018. p. 24-26.
9. Rabbani, A.G., Z. Yao, and C. Wang, *Does personality predict financial risk tolerance of pre-retiree baby boomers?* Journal of Behavioral and Experimental Finance, 2019. **23**: p. 124-132.
10. Oehler, A., et al., Investors' Personality Influences Investment Decisions: Experimental Evidence on Extraversion and Neuroticism. Journal of Behavioral Finance, 2018. **19**(1): p. 30-48.
11. Pan, C.H. and M. Statman, Beyond risk tolerance: regret, overconfidence, personality and other investor characteristics. Leavey School of Business, 2010.
12. Mowen, J.C., The 3M model of motivation and personality: Theory and empirical applications to consumer behavior. 2000: Springer Science & Business Media.
13. Company., T.M.-B.C.A.m.a.t.M.a.-T.M.-B. [cited 2022; Available from: https://www.themyersbriggs.com/en-US/Resources/Creating-Clarity-Addressing-Misconceptions-of-MBTI.
14. Abidin, N.H.Z., et al., *Improving Intelligent Personality Prediction using Myers-Briggs Type Indicator and Random Forest Classifier.* International Journal of Advanced Computer Science and Applications, 2020. **11**(11).
15. Capretz, L., *Implications of MBTI in Software Engineering Education.* SIGCSE Bulletin, 2002. **34**: p. 134-137.
16. Puji, R. and A. Ahmad, Learning Style of MBTI Personality Types in History Learning at Higher Education. Scientific Journal Of PPI-UKM, 2016. **3**: p. 289-295.
17. Sample, J., *A Review of the Myers-Briggs Type Indicator in Public Affairs Education.* Journal of Public Affairs Education, 2017. **23**(4): p. 979-992.
18. R, V., Comparative Analysis for Personality Prediction by Digital Footprints in Social Media. Journal of Information Technology and Digital World, 2021. **3**: p. 77-91.
19. Stachl, C., et al., Personality Research and Assessment in the Era of Machine Learning. 2019.
20. Mitchelle, J., Myers-Briggs Personality Type Dataset. Includes a Large Number of People's MBTI Type and Content Written by Them.

21. Golbeck, J., et al. Predicting personality from twitter. In 2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing. 2011. IEEE.
22. Komisin, M.C. and C. Guinn, *Identifying personality types using document classification methods*. 2011, University of North Carolina Wilmington Wilmington.
23. Akbar, A.T., et al., A proposed method for handling an imbalance data in classification of blood type based on Myers-Briggs type indicator. Jurnal Teknologi dan Sistem Komputer, 2020. **8**(4): p. 276-283.
24. Mehta, Y., et al., Bottom-Up and Top-Down: Predicting Personality with Psycholinguistic and Language Model Features. 2020. 1184-1189.
25. Ontoum, S. and J.H. Chan, Personality Type Based on Myers-Briggs Type Indicator with Text Posting Style by using Traditional and Deep Learning. arXiv preprint arXiv:2201.08717, 2022.
26. Amirhosseini, M.H. and H. Kazemian, *Machine learning approach to personality type prediction based on the myers–briggs type indicator®*. Multimodal Technologies and Interaction, 2020. **4**(1): p. 9.
27. Brandon Cui. and Calvin Qi, *Survey Analysis of Machine Learning Methods for Natural Language Processing for MBTI Personality Type Prediction.* Available at: https://cs229.stanford.edu/proj2017/final-reports/5242471.pdf , 2017.
28. Raje, M. S., and Singh, A. (2017), *Personality detection by analysis of Twitter profiles*. Advances in Intelligent Systems and Computing, 667–675. https://doi.org/10.1007/978-3-319-60618-7_65