Deciphering personality traits from a person's digital behavior using HAN Document Classifier

Rajat Subhra Ghosh
Department of Computer Science
California State University, Northridge
California, USA
email – rajatsubhra.ghosh@gmail.com

Mayur Rahangdale
Department of Computer Science
California State University, Northridge
California, USA
email – rahangdalemayur@gmail.com

***Abstract*—Our focus is to expand the potential of machine learning approaches to personality assessment by using social media and other digital records with established personality measures like MBTI. In this project we have utilized the deep learning methods of document classifiers using Hierarchical Attention Network. Our goal is to deploy a personality assessment model in an environment where personality-profiling data availability is abundant. We investigated how a deep-learning approach based on Hierarchical Attention Networks can improve prediction performance for multiple information extraction tasks like from unstructured psychometric documents to structured resumes/portfolios, when compared to conventional methods that do not adequately apprehend syntactic and semantic contexts from free-text documents.**

Keywords—hierarchical attention networks, han, personality assessment, machine learning, mbti, hierarchical classification, hierarchical attention, document classification

I. INTRODUCTION

There is always a need for a powerful assessment tool in the psychological field using the applications of machine learning. Therefore, our main goal is to interpret the human persona from their social media posting and to determine their personality. This model is generic enough so that it can be implemented by various companies like Facebook, Google, who deal with large numbers of employees. This will help them immensely in figuring out the general/psychological profile of their employees and thereby strategize different courses of action for their individual growth from management point of view. Also, it can be applied in the medical field and can be utilized by the psychiatrists to understand their patient behavior/mental health more accurately.

We have used machine learning to build a document classifier that accepts textual inputs from various social media platforms (e.g. social media posts) and returns a particular personality type based on the Myers–Briggs Type Indicator (MBTI) [1]. For developing the document classifier, we have used Hierarchical Attention Network (i.e., HAN) [2]. There are other personality type evaluation models like Big 5, DISC. Though there are critiques on MBTI model [3], we have used the MBTI [1] as it is one of the most popular and widely used across the industries for assessing the personality traits. We have selected HAN, because the previous studies in the similar areas of problem, shows that HAN can be successfully applied for document classification [2]. The text documents have a hierarchical structure as the document breaks down into sentences and ultimately into words. This hierarchy can be captured by HAN in an intuitive way without loss of potential insights [2]. Information is summarized on the word level by the word encoder layer as relevant contexts and passed on to the sentence encoder layer which essentially repeats the same for sentences. Furthermore, the HAN architecture incorporates hierarchical attention mechanism layers at both word level and sentence level which computes importance weights and attends the words and sentences based on how important they are for the overall meaning of the document [2]. Considering the fact that extracting personality features from text depends a lot on the usage of essential keywords in the MBTI test, this characteristic of HAN is essential for a better understanding of the semantic structure of the document and eventually in the classification of our dataset. Moreover, this experiment involves a very large labeled text-document dataset. when working with larger datasets [4], it is observed that HAN implementation shows greater training and validation accuracy than RNN or CNN

A. Background Research

There is a significant growing interest on how the machine learning field can be applied for personality detection using social media content. So far, most studies have focused on different machine learning algorithms using CNN with Big Five personality traits [6], [7], RNN with MBTI [8], Extreme Gradient Boosting with MBTI [9], SVM, and Naive Bayes with MBTI [10], Grey Prediction model, The Multiple regression model and the multi-tasking model with The gray system theory [11], deep learning model with

Big Five personality traits [12], Naive Bayes, SVM, Deep learning with MBTI [13], MLP, LSTM, CNN, GRU with Big Five personality traits [14], XGBoost with MBTI [15], Random Forest, Decision Trees with MBTI [16], Naïve Bayes classifier, a SVM, and a Multilayer perceptron neural network with Big Five personality traits [17], SVM, Naive Bayes, Decision Trees with Big Five personality traits [18], Network representation learning with Big Five personality traits [19], SVM with Big Five personality traits [20], Graph Convolution network with Big Five personality traits [8]. In previous research, we found that none of them have used a large volume of dataset. We have used a dataset of 9381 data samples. Also, we found that HAN performs efficiently when the sample dataset is large and MBTI is very popular and well understood by the professional people. Hence, we have selected HAN and MBTI as our primary model for behavioral prediction.

B. Introduction to HAN

Yang et al. (2016) [2] show how hierarchical attention networks (HAN) may be effectively used on various levels and demonstrate that the attention mechanism is applicable to the document classification problem, not just sequence generation.

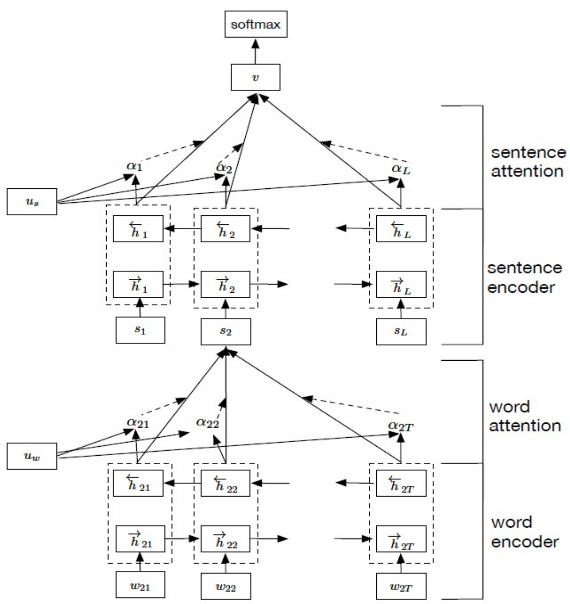

Fig 1. Representation of HAN as a combination of multiple layers (i.e attention layer and encoder layer present at word as well as sentence level) [17]

HAN has two encoder networks - i.e., word and sentence encoders. The word encoder processes each word and aligns them a sentence of interest. Then, the sentence encoder aligns each sentence with the final output. HAN enables hierarchical interpretation of results as below. The user can understand (1) which sentence is crucial in classifying the document and (2) which part of the sentence, i.e., which words, are salient in that sentence.

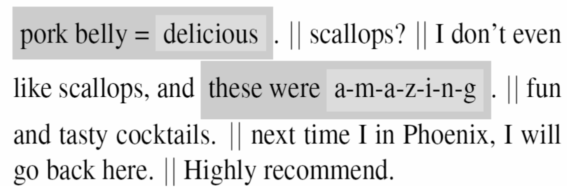

Fig 2. Representation depicts that all parts of a document are not equally important to gain the essential meaning from it. The highlighted sentences in darker grey deliver stronger meaning compared to the others. [2]

C. Introduction to MBTI

MBTI (Myers–Briggs Type Indicator) is one of the most popular personality type evaluation models, which is derived from the Latin word persona, which means describing the behavior or character of an individual… the preferences of an individual are categorized into four dimensions, and different combinations of the personality type key in these categories represent 16 different personality types based on the Myers–Briggs Type Indicator®. Figure 1 shows these 16 personality types that result from the interactions among the preferences of an individual [5].

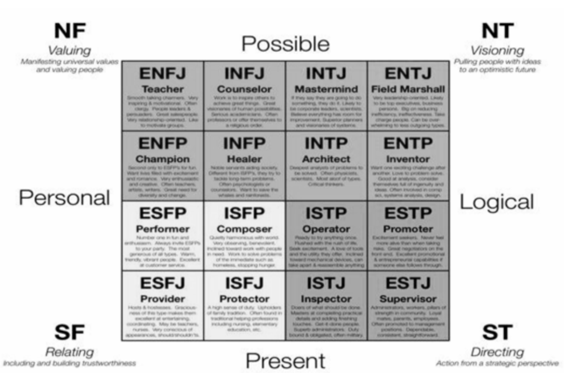

Fig 3. Representation of personality types in MBTI as combinations of four dichotomous attitudes or functioning styles: Extraversion or Introversion, Judgment or Perception, Thinking or Feeling, and Sensing or Intuition. [5]

D. Programming Language, Libraries and Platform

We have used Python as a primary programming language. Also, we are using python libraries like pandas and NumPy for data handling; nltk for text handling and text processing; spacy and fastai for deep learning; pytorch framework for model implementation; seaborn and matplotlib for data visualization; sklearn for data splitting.

II. Project Problem and Hypothesis

The project problem is to determine the human persona by their social activities. The principal idea of the HAN document classifier is to classify social media texts to infer the correct personality classification.

We are doing the hypothesis that HAN will be a better fitting neural-network architecture than previously used methods for document classification when it comes to largely labelled social media data. We believe that HAN will show better accuracy than the previously introduced methodologies discussed in the background research section. The justification for our hypothesis is discussed below.

1) The text documents have a hierarchical structure as the document breaks down into sentences and ultimately into words. This hierarchy can be captured by HAN in an intuitive way without loss of potential insights [2]. Information is summarized on the word level by the word encoder layer as relevant contexts and passed on to the sentence encoder layer which essentially repeats the same for sentences.

2) Furthermore, the HAN architecture incorporates hierarchical attention mechanism layers at both word level and sentence level which computes importance weights and attends the words and sentences based on how important they are for the overall meaning of the document [2]. Considering the fact that extracting personality features from text depends a lot on the usage of essential keywords in the MBTI test, this characteristic of HAN is essential for a better understanding of the semantic structure of the document and eventually in the classification of our dataset.

3) Moreover, this experiment involves a very large labeled text-document dataset. when working with larger datasets, it is observed that HAN implementation shows greater training and validation accuracy than RNN or CNN [4].

Hence, the given hypothesis was done based on the above discussions.

III. DATASET

The dataset used in this project was taken from the Kaggle repository named '(MBTI) Myers-Briggs Personality Type Dataset' [21], which has a usability score of 8.8. It has 2 columns named 'type' and 'posts'. Table I depicts how the sample data look like. The 'type' column contains the MBTI category values whereas the 'posts' column contains the collection of social media posts of individuals in text format. There is a total of 8675 instances in the dataset. The 'type' column has 16 unique values and can be considered as categorical data, whereas the posts column is textual data.

The dataset contains two columns one is type and another is post. Type is categorical data as per MBTI representation and post is of type textual.

	type	posts
0	INFJ	'http://www.youtube.com/watch?v=qsXHcwe3krw|||...
1	ENTP	'I'm finding the lack of me in these posts ver...
2	INTP	'Good one _____ https://www.youtube.com/wat...
3	INTJ	'Dear INTP, I enjoyed our conversation the o...
4	ENTJ	'You're fired.|||That's another silly misconce...

Fig 4. Representation of sample dataset from the training data

TABLE I. STRUCTURE OF MBTI DATASET

'posts'	Descriptive feature	Textual data
'type'	Target Feature	Categorical data (16 unique values)

Observation of textual data in the 'posts' column revealed that most of the posts contained links and Internet slangs. A thorough text-preprocessing to remove unwanted texts was required.

IV. METHODOLOGY & IMPLEMENTATION

This section describes the methodology adopted to prepare the document classifier model using python, HAN machine learning algorithm for the given MBTI dataset.

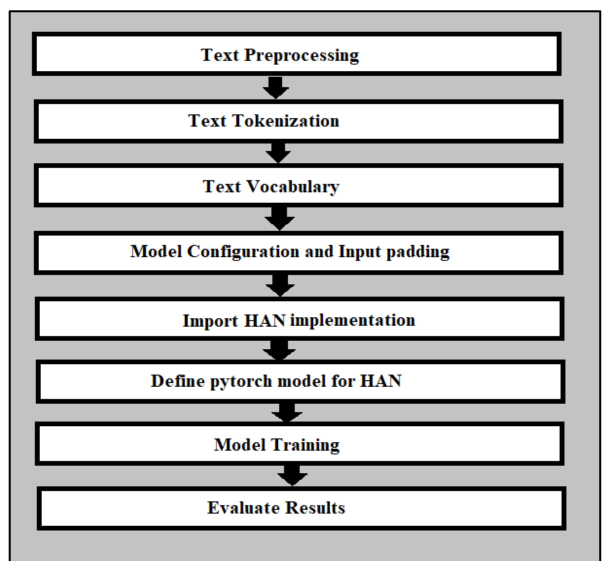

Fig 5. Depicts the system pipeline by describing all the levels of methodology in a sequential manner.

A. Oversampling

It clearly shows in the Fig 6, Before Oversampling that the data is imbalanced throughout the different classes. Oversampling is done for minority classes present in the dataset. After oversampling, the total number of instances increases from 8675 to 9381.

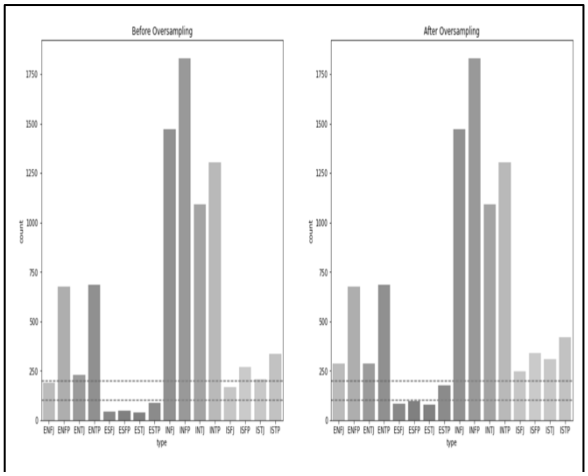

Fig 6. Represents the distribution of data before and after oversampling where horizontal dotted lines indicate the increase in size of minority classes.

The above plot of type vs the number of instances explains the distribution of data along different classes. It clearly shows that the data is imbalanced throughout the different classes. Oversampling is done for minority classes present in the dataset. After oversampling, the total number of instances increases from 8675 to 9381.

B. Data Preprocessing

We created a text preprocessing function using the NLTK library. The raw text is fed into the function. URLs are removed from the text data by defining Regex for URLs. In the next steps, all the punctuations are removed except the end of line punctuations. The text obtained is stripped of white spaces and split into sentences.

It was observed that a single instance had an average 40 number of sentences (i.e. posts). All the sentences with length less than 5 were removed. The max limit was set to 23 i.e. out of every instance, only 23 sentences were considered.

C. Text Tokenization and Text Vocabulary

Spacy library is used to create a tokenizer pipeline and clean text is fed to obtain the sentences. Sentence lengths are also computed which will be required further for padding and model configuration. Vocabulary is created using FASTAI's vocabulary class. Sentence tokens from the previous step are input to this function where the max features are set as 60000 with a minimum occurrence frequency of 5. Numericalizing of the vocabulary is done further to convert it into a searchable indexed dictionary.

D. Model Configuration and Input Padding

To find optimal model configuration, the overall structure of text data was studied more thoroughly after the preprocessing steps. The max sentence length and the max number of sentences in a document are calculated, and based on these the Model configuration is defined. These parameters are also used for padding the input tensors. The dataset is padded to a three-dimensional array: the first dimension represents the total number of documents, the second represents each sentence in a document and the last represents each word in a sentence. The shape of this array is (total_instances, maxlen_doc, maxlen_sent).

The model configuration used is shown in the table below.

TABLE II. MODEL CONFIGURATION

Configuration	Value	Description
max_vocab	60000	maximum number of unique words that should be included in the tokenized word index/ vocabulary
maxlen_sent	18	maximum number of words in each sentence
maxlen_doc	79	maximum number of sentences in one instance/document
embed_dim	100	dimensions of the word embedding

E. Importing HAN Implementation

Pytorch allows users to write custom models and custom layers for their model. Using this feature, the implementation of HAN was written. This implementation was taken from the medium article authored by Javier Rodriguez Zaurin [22], where he implemented HAN document classifier using Pytorch to predict Amazon review scores.

F. Model Definition

The first layer is the Embedding layer. The embedding layer takes input tokens as input and return token embeddings, if the tokens are found in the embedding lookup dictionary. The embedding weights get updated as the model trains.

Our Attention network model consists of two parts: Bidirectional GRU encoder layer and Attention with context layer. The attention network model exists at word level as well as at sentence level. The word attention network at word level and sentence attention network at sentence level combine to form the hierarchical attention network model. While bidirectional GRU learns the meaning behind those sequences of words and returns word vector representation corresponding to each word, the attention network gets weights corresponding to each word vector using its own shallow neural network. Then it aggregates the representation of those words to form a sentence vector representation i.e. it calculates the weighted sum of every vector. These sentence vector representations are then passed through sentence attention network model (i.e. sentence encoder and sentence attention) resulting into document vector representations. Here, similar procedure is present at word as well as sentence level to make sure that the final vector representation embodies the gist of the complete document. The last layer is a fully connected layer with SoftMax activation function for prediction.

The three dropouts used are Embedding dropouts, weight dropout and locked dropout. The implementation of dropouts is taken directly from the medium article authored by Javier Rodriguez Zaurin [22].

```
HierAttnNet(
  (wordattnnet): WordAttnNet(
    (word_embed): Embedding(70082, 50, padding_idx=1)
    (rnn): GRU(50, 32, batch_first=True, bidirectional=True)
    (word_attn): AttentionWithContext(
      (attn): Linear(in_features=64, out_features=64, bias=True)
      (contx): Linear(in_features=64, out_features=1, bias=False)
    )
  )
  (sentattnnet): SentAttnNet(
    (rnn): GRU(64, 32, batch_first=True, bidirectional=True)
    (sent_attn): AttentionWithContext(
      (attn): Linear(in_features=64, out_features=64, bias=True)
      (contx): Linear(in_features=64, out_features=1, bias=False)
    )
  )
  (fc): Linear(in_features=64, out_features=16, bias=True)
)
```

Fig 7. Represents the definition of model and its layers along with the configuration.

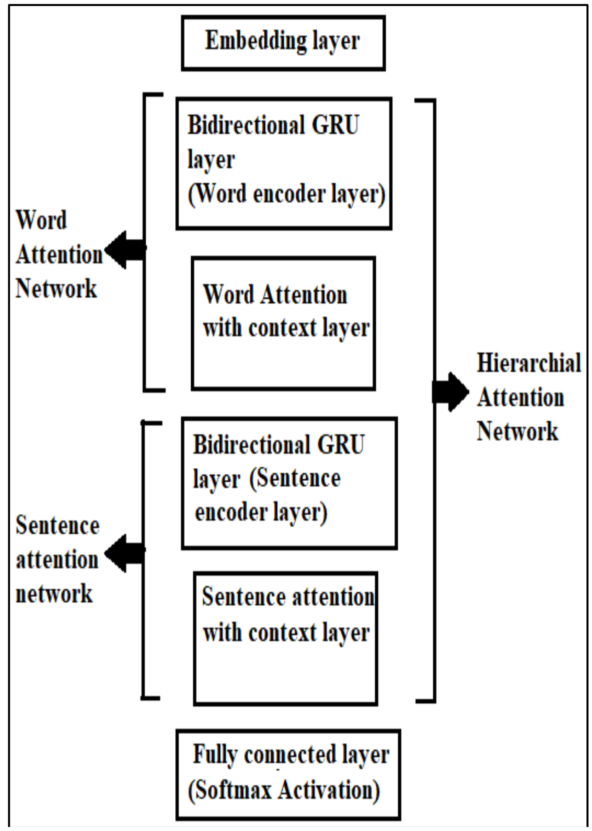

Fig 8. Block diagram of model definition where word attention network and sentence attention network combine to form hierarchical attention network.

G. *Model Training*

The model defined in the previous step is trained on the defined training dataset and validated on the defined validation dataset with train size of 0.8. While compiling the model, the loss function used is 'categorical cross entropy' and the optimizer used is 'Adam optimizer'. Early stopping is added to model training and the model is trained over 23 epochs.

V. RESULTS

As shown in Fig 9, the graph of validation accuracy tends to a constant value after epoch 9. Early stopping is triggered on the 9th epoch with training accuracy of 0.763 and validation accuracy of 0.558.

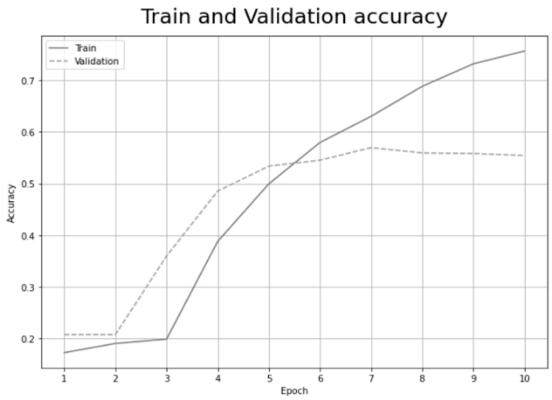

Fig 9. Graphical representation of model accuracy where train accuracy is a solid line curve increasing with epochs and its validation accuracy is a dotted line curve which tends to a constant value after 9th epoch.

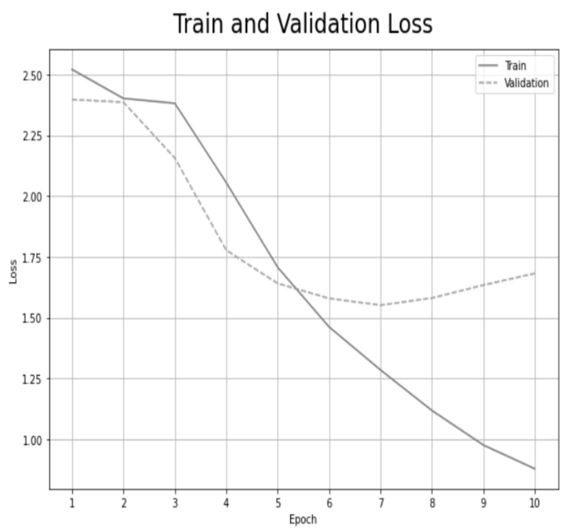

Fig 10. Graphical representation of model loss.

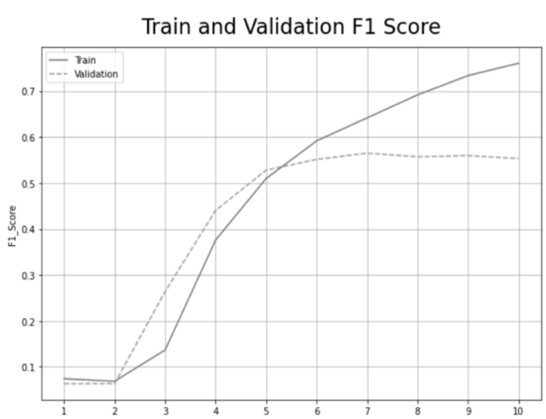

Fig 11. Graphical representation of F1 score showing similar behavior as accuracy curve

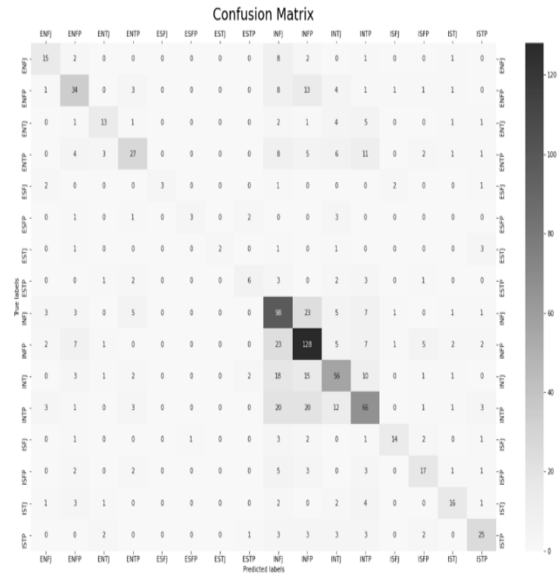

Fig 12. Graphical representation of 16X16 confusion matrix for predicted and true labels

The confusion matrix indicates a total precision of 0.63 and a total recall of 0.47. The accuracy computed by the confusion matrix is 0.557 which is equal to validation accuracy.

VI. CONTRIBUTIONS

We are the first to implement the Hierarchical Attention Network for personality trait classification using the MBTI test. According to our hypothesis, HAN document classifiers should be a state-of-the-art model for MBTI classification. Although we haven't achieved more accuracy than previous studies, it can be achieved by introducing a data collection methodology, and training over a larger size of data (5 times larger than original dataset).

Multinational companies with a large number of employees may have employee data available, which can be accumulated and used for our model. In this way, our model can be deployed in such an environment without any hassle. The large amount of data available in such scenarios will help in finer training of the model.

VII. DISCUSSIONS AND FUTURE WORK

We can conclude our project by making the statement that our HAN document classifier shows lesser accuracy than previous studies on a smaller training data.

The main limitation faced in this project is insufficient amount of data available in public repositories like Kaggle. We believe our proposed model requires a significantly larger dataset than what is currently available.

REFERENCES

[1] N. A. Indicated, 'Review of Manual: A Guide to the Development and Use of the Myers-Briggs Type Indicator', *Contemporary Psychology: A Journal of Reviews*, vol. 31, no. 10. pp. 819–819, 1986[Online]. Availablehttp://dx.doi.org/10.1037/024196.

[2] Z. Yang et al., 'Hierarchical Attention Networks for Document Classification', in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2016, pp. 1480–1489[Online]. Availablehttps://www.aclweb.org/anthology/N16-1174.

[3] J. E. Barbuto, 'A Critique of the Myers-Briggs Type Indicator and its Operationalization of Carl Jung's Psychological Types', *Psychol. Rep.*, vol. 80, no. 2, pp. 611–625, Apr. 1997[Online]. Availablehttps://doi.org/10.2466/pr0.1997.80.2.611.

[4] '[PDF] CATEGORIZATION OF TEXT WITH RNN, CNN AND ..'[Online]. Availablehttps://www.semanticscholar.org/paper/CATEGORIZATION-OF-TEXT-WITH-RNN%2C-CNN-AND-HAN-MODELS-Dhawan/2bc95a0bdd143bf24e3c5a372eb230f441d33670.

[5] P. D. Tieger et al., *Do What You Are: Discover the Perfect Career for You Through the Secrets of Personality Type*. Little, Brown, 2014[Online]. Availablehttps://play.google.com/store/books/details?id=Xe3iAAAAQBAJ.

[6] N. Majumder et al., 'Deep Learning-Based Document Modeling for Personality Detection from Text', *IEEE Intell. Syst.*, vol. 32, no. 2, pp. 74–79, Mar. 2017[Online]. Availablehttp://dx.doi.org/10.1109/MIS.2017.23.

[7] Y. Mehta et al., 'Recent trends in deep learning based personality detection', *Artificial Intelligence Review*, vol. 53, no. 4, pp. 2313–2339, Apr. 2020[Online]. Availablehttps://doi.org/10.1007/s10462-019-09770-z.

[8] R. Hernandez and I. S. Knight, 'Predicting Myers-Bridge Type Indicator with text classification', in *Proceedings of the 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA*, 2017, pp. 4–9.

[9] M. H. Amirhosseini and H. Kazemian, 'Machine Learning Approach to Personality Type Prediction Based on the Myers–Briggs Type Indicator®', *Multimodal Technologies and Interaction*, vol. 4, no. 1, p. 9, Mar. 2020[Online]. Availablehttps://www.mdpi.com/2414-4088/4/1/9[Accessed: 24October2020].

[10] M. C. Komisin and C. Guinn, 'Identifying personality types using document classification methods', University of North Carolina Wilmington, 2011[Online]. Availablehttps://pdfs.semanticscholar.org/9054/3391f5f2f7e1e5859aadd3d5a6b2c3963363.pdf.

[11] C. Li et al., 'Personality Prediction of Social Network Users', in *2017 16th International Symposium on Distributed Computing and Applications to Business, Engineering and Science (DCABES)*, 2017, pp. 84–87[Online]. Availablehttp://dx.doi.org/10.1109/DCABES.2017.25.

[12] T. Tandera et al., 'Personality Prediction System from Facebook Users', *Procedia Comput. Sci.*, vol. 116, pp. 604–611, Jan. 2017[Online]. Availablehttp://www.sciencedirect.com/science/article/pii/S1877050917320537.

[13] B. Cui and C. Qi, 'Survey Analysis of Machine Learning Methods for Natural Language Processing for MBTI Personality Type Prediction'. cs229.stanford.edu, 2017[Online]. Availablehttp://cs229.stanford.edu/proj2017/final-reports/5242471.pdf.

[14] A. S. Khan et al., 'Personality Classification from Online Text using Machine Learning Approach', *pdfs.semanticscholar.org*[Online]. Availablehttps://pdfs.semanticscholar.org/0185/5e9d101c18d4de06a8827637c3c60a55ada8.pdf.

[15] A. Genina et al., 'An Approach for Sentiment Analysis and Personality Prediction Using Myers Briggs Type Indicator', in *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2020*, 2021, pp. 179–186[Online]. Availablehttp://dx.doi.org/10.1007/978-3-030-58669-0_16.

[16] A. C. E. S. Lima and L. N. de Castro, 'A multi-label, semi-supervised classification approach applied to personality prediction in social media', *Neural Netw.*, vol. 58, pp. 122–130, Oct. 2014[Online]. Availablehttp://dx.doi.org/10.1016/j.neunet.2014.05.020.

[17] A. Imran et al., 'An enhanced approach for quantitative prediction of personality in facebook posts', *International Journal of Education and Management Engineering (IJEME)*, vol. 8, no. 2, pp. 8–19, 2018[Online]. Availablehttp://www.mecs-press.net/ijeme/ijeme-v8-n2/IJEME-V8-N2-2.pdf.

[18] X. Sun et al., 'Group-level personality detection based on text generated networks', *World Wide Web J. Biol.*, vol. 23, no. 3, pp. 1887–1906, May 2020[Online]. Availablehttps://doi.org/10.1007/s11280-019-00729-2.

[19] G. Carducci et al., 'TwitPersonality: Computing Personality Traits from Tweets Using Word Embeddings and Supervised Learning', *Information*, vol. 9, no. 5, p. 127, May 2018[Online]. Availablehttps://www.mdpi.com/2078-2489/9/5/127[Accessed: 24October2020].

[20] Z. Wang et al., 'Encoding Text Information with Graph Convolutional Networks for Personality Recognition', *NATO Adv. Sci. Inst. Ser. E Appl. Sci.*, vol. 10, no. 12, p. 4081, Jun. 2020[Online]. Availablehttps://www.mdpi.com/2076-3417/10/12/4081[Accessed: 24October2020].

[21] J. Mitchell, '(MBTI) Myers-Briggs personality type dataset'. [Online]. Availablehttps://www.kaggle.com/datasnaek/mbti-type[Accessed: 31October2020].

[22] J. Zaurin, "Predicting Amazon review scores using Hierarchical Attention Networks with PyTorch and Apache…", Medium, 2020. [Online]. Available: https://towardsdatascience.com/predicting-amazon-reviews-scores-using-hierarchical-attention-networks-with-pytorch-and-apache-5214edb3df20. [Accessed: 31- Oct- 2020].