Comparative Study of Personality Prediction Using Machine Learning Algorithms

Gayathri Kadam. C[1], Dr. D. Preethi[2]

[1]Student, Department of computer Science, Mount Carmel College Autonomous, Bengaluru, India

[2]Assistant Professor, Department of Computer Science, Mount Carmel College Autonomous, Bengaluru, India

Abstract: *Social Media has become a platform for humans to convey their happiness, emotions, opinions and expectations, with this data which is publicly viewable and accessible; we will be able to analysis the personality of the person. Personality prediction has many real - world applications as the usage of social media is expanding gradually from this, we can obtain huge amount of textual data. In this paper, we will approach a method in which the user's personality can be accurately predicted with the data accessible publicly through social media platforms viz. Twitter, You - tube, and Facebook. We will analysis and describe the data using machine learning techniques and with the help of Machine learning algorithm namely Random Forest, Stochastic Gradient Descent, XGBoost, Logistic Regression, KNN, Support vector machine. This result may help in Cognitive process to identify the personality type.*

Keywords: Machine Learning, Myers - Briggs Type Indicator, Naïve Bayes, Support vector machine, Random Forest, social media, Statistical Analysis

1. Introduction

The idea of personality begins with Hippocrates in 370 BCE and it has been studied for at least 2000 years. Personality originated from the Latin word persona, which means a theatrical mask worn by performer in order to project various roles or even to disguise their identities and in Greek the word persona indicates to speak through. An individual's personality is the combination of traits and patterns that influence their etiquette, ideas, desire, and emotion. It makes an individual to always think, sense and act in unique ways. In recent years, these patterns strongly influence personal value, attitudes, insights and expectations.

The Myers - Briggs Personality Type Indicator shown in Figure 1. is deliberated to recognize a human's personality type, strengths and inclinations. It was proposed by Isabel Myers and her mother Katherine Briggs based on the theory of personality types. Currently, the MBTI inventory is one of the most popularly used psychological instruments in the world. No one's personality type is "best" or "better" than other. It is not a tool that look for unconventionality or idiosyncrasy. Instead, its aim is clearly to assist you know more about one self.

Organization looks for employee's personality not only with the skill sets and knowledge for position at the job but also the amount of flexibility in his character to meet the changing circumstances.

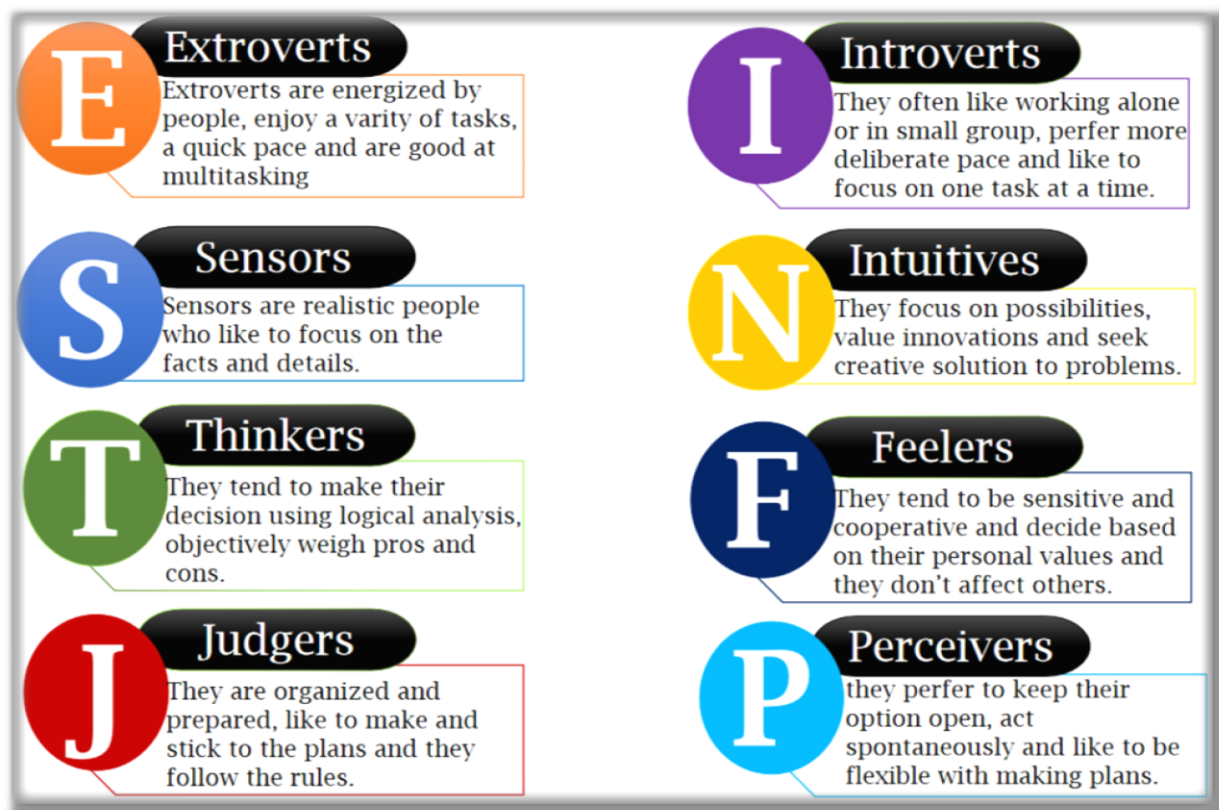**Figure 1:** MBTI Personality Types

Figure 2: MBTI personality type description

2. Literature Review

The study conducted by Thahira M, Mubeena AK, shows the accuracy of the model performance where they have used Support vector machine, Random Forest, Naïve Bayes, LSTM, Decision Tree algorithm. In this paper [1], machine learning method support vector machine classifier showed the accuracy of 58% for Openness - to - Experience and deep learning method LSTM showed the better accuracy nearly 80% accuracy. In paper [2], personality classification is having been analysed using Twitter streaming API have enforced AdaBoost, Multinomial Naïve Bayes (MNB), and LDA. It showed that MNB has highest the accuracy with 73.43, recall of 0.71and F1 - score of 0.72, precision of 0.7. They study performed by Jennifer Golbeck, Michon Edmondson, Cristina Robles, and Karen Turner [3]. Few people hadtaken a personality test through the Twitter API from whichthey collected openly available information from their profiles. For each of five personality traits they applied ZeroR and Gaussian Processes to predict scores within 11% - 18% of their actual value. The study by Bayu Yudha Pratama and Riyanarto Sarno [4]. Using Support vector machine, Multinomial Naïve bayes and KNN techniques for classification in which Multinomial Naïve bayes showed the best accuracy of 60% among other model. In paper [5], they have examined the edifice of social media platform and linguistic features of personality interactions, XGBoost classifier performed the average feature set, with accuracy of 74.2%. for the Extraversion trait that used Social Network Analysis (SNA) features and obtained the accuracy of 78.6%. In this paper [6], they have extractedthe data from social media and personality questionnaires for the dataset, they have used PCA (principal component analysis) and correlation and then implemented gray prediction model,

multiple regression model and multitasking model. That showed the gray prediction of MAE values performed good and the prediction accuracy lied between 0.8 and 0.9. The combine model Naïve Bayes and support vector machine method performed better personality prediction [7]. SVM provided higher accuracy of classification rate and minimum error rate. In addition, naïve Bayes provided a better prediction than other algorithms. A proposed system by [8] that analysed posts/ tweets from social media of an individual. LIWC, Emolex and TF/IDF is used for feature selection and the obtained feature vector is used for training data and testing data for various machine learning algorithms such as support vector machine, Neural net, Naïve Bayes. In which support vector machine with all feature vectors performed the best accuracy for all four of MBTI types. In paper [9] they have proposed a model to predict the personality type making use of 1.2 million tweets that is remarked with MBTI type. For feature selection Binary word n - gram is used. Logistic regression model showed improvement in Introversion/Extroversion and Thinking/Feeling dimensions but no changes in Sensing - Intuition and performance reduced for Perceiver - Judgers, linguistic features showed better results. In paper [10], an Experiment had conducted by using my Personality dataset and 150 users' data which was collected manually and used linguistic features such as SPLICE (structured Programming for Linguistic Cue Extraction) and LIWC (Linguistic Inquiry and Word Count) and build Multilayer perceptron, GRU and Long short - term memory and combined CNN 1D and LSTM algorithm to predict the personality. The model performance showed the accuracy of 79.49%.

3. Methodology

3.1 Process Flow

The Frame work is shown in the Figure 3. Increase in the use of social media paves way to process the textual data and predict the personality.
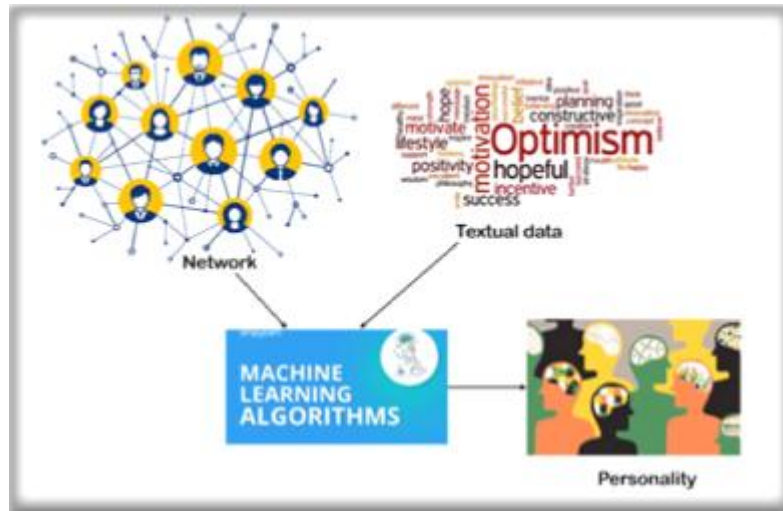

Figure 3: Process flow for personality prediction

3.2 Data Pre - Processing

For analysis, the Twitter Dataset is used which contains MBTI types and the posts of 8675. The data is pre - processed using TF - IDF vectorizerto covert the text into meaningful representation of vector which can be used for the prediction in machine learning algorithm. Figure 4. Shows the personality types in the dataset
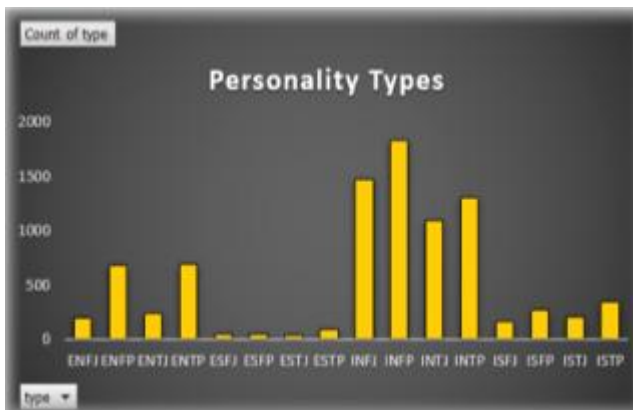

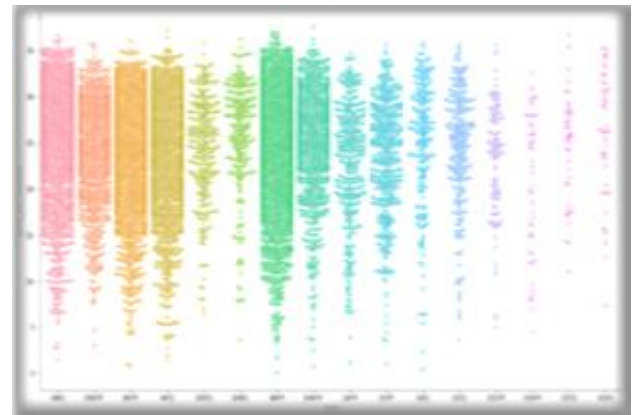**Figure 4:** Personality types

Figure 5: Number of words per post

The total number of words per post for 50 post is shown in the Figure 5. each post has been separated by pipeline [11]. The dataset is added with 4 more columns build on the four personality traits that is Thinking/Feeling (T/F), Extroversion / Introversion (E/I), Sensation / Intuition (S/I), Judgment/Perception (J/P). The retrieved data is spilt - into training dataset and testing dataset. Cross - validation used is Hold out method were data for training is 70% and for testing 30% of data is used. The testing data used to check the performance of the model classifier.
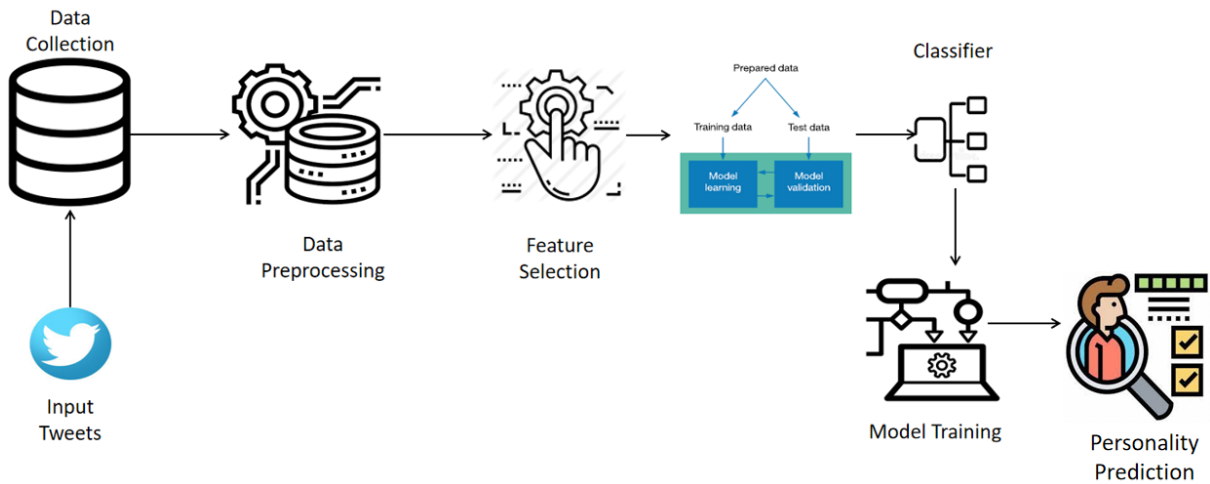
Figure 6: System Architecture

3.3 Classification Process

The personality of a person can have more than one personality classes. The task was to divide the 16 classes into four binary classes and model was created for each type indicator, The classifiers used are Stochastic Gradient Descent, Logistic Regression, Random Forest, XGBoost, K - Nearest Neighbor, Support vector machine.

4. Results

Table 1. shows the model trained for individual mbti personality type with different classifier algorithms. The

SVM model performed a better accuracy along with Logistic Regression and stochastic Gradient Descent for personality prediction.
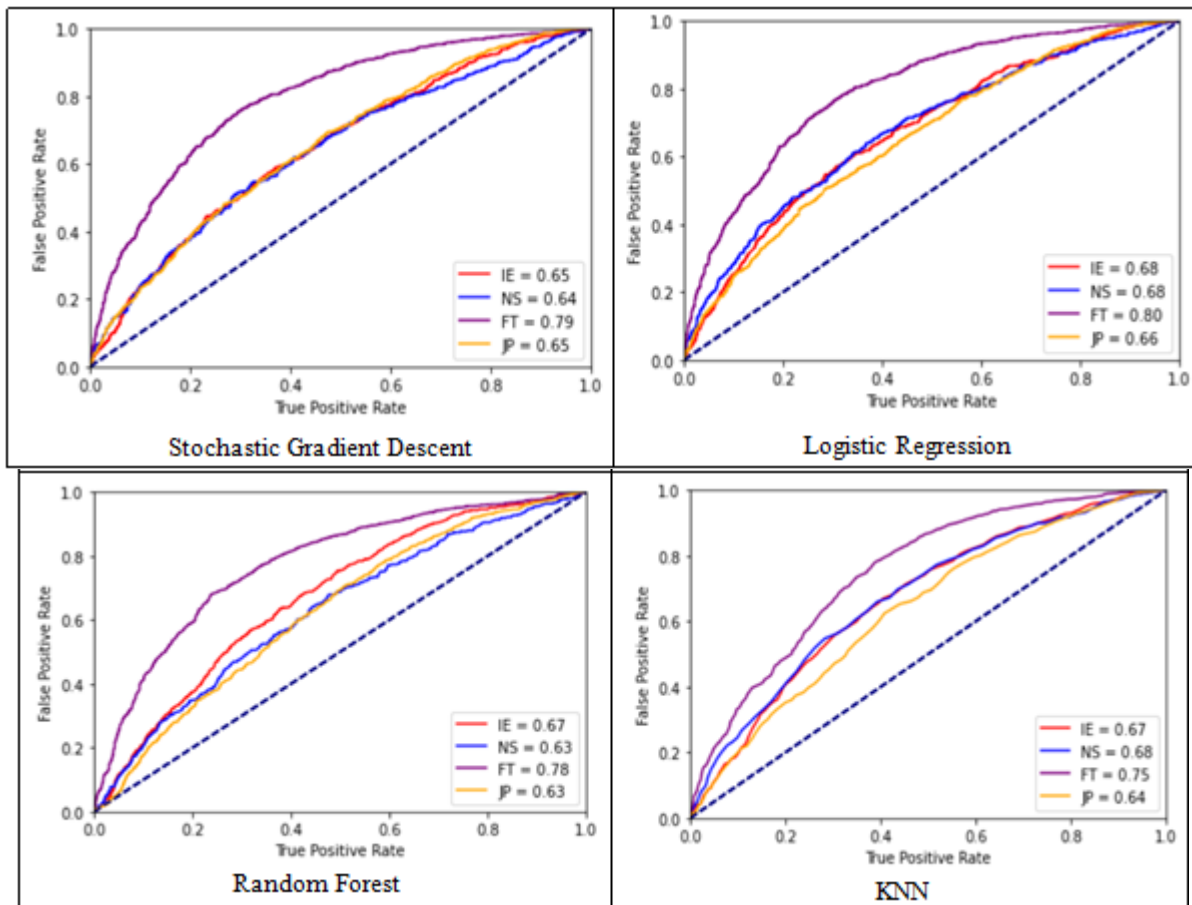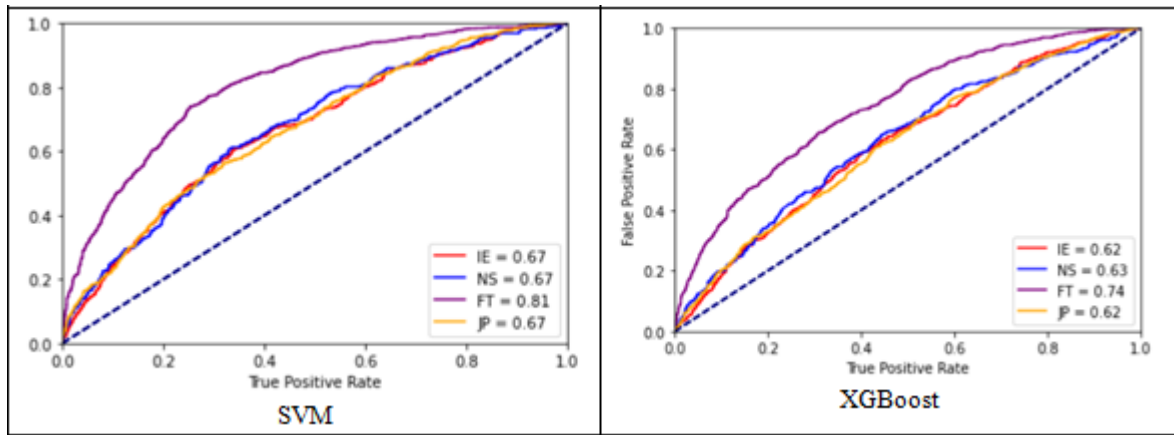
Table 1: Comparative Analysis of result

Algorithm	E/I	N/S	F/T	J/P	Accuracy
Stochastic Gradient Descent	77.40	86.03	72.09	64.55	75.01
Logistic Regression	77.84	86.06	72.44	64.51	75.21
Random Forest	76.14	86.25	70.30	59.51	73.05
KNN	77.86	86.03	60.25	60.92	71.26
XGBoost	75.53	85.94	67.15	62.15	72.69
SVM	77.56	86.21	73.03	65.50	75.57

Table 2: AUC Curve for each Model

The AUC (Area Under Curve) is used to evaluate the performance in the classification models. Table 2. represents the AUC curve for each model in which SVM model has the more area under curve and the performance of the model is good.

Figure 7: Comparative Analysis of each type

Figure 8: Model Performance

5. Conclusion and Future Enhancement

This paper shows that the MBTI personality type can be predicted by using the tweets in the Twitter. The Pearson Correlation values between all the types and each personality dimension gave the different types of features on the posts. Used six machine learning algorithms for the comparative study in which we analysed SVM can be used

for the better prediction among other algorithms which is shown in the Figure 8.

In future, we plan to expand the dataset with the live streaming of data on the social media to analyze the personality of the person. We would improve the accuracy of the machine learning algorithms. Deep learning algorithms such as RNN can be used in further analysis
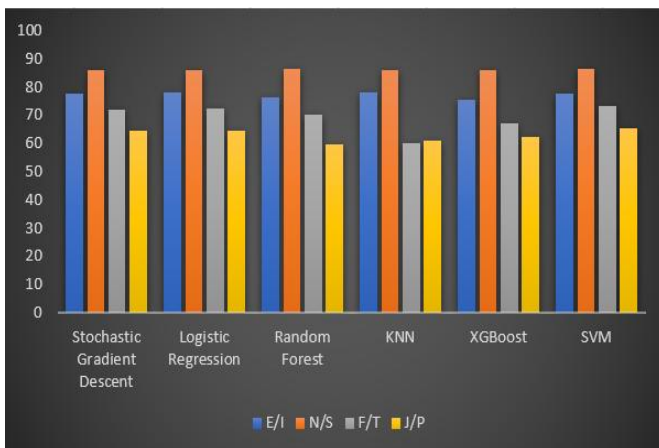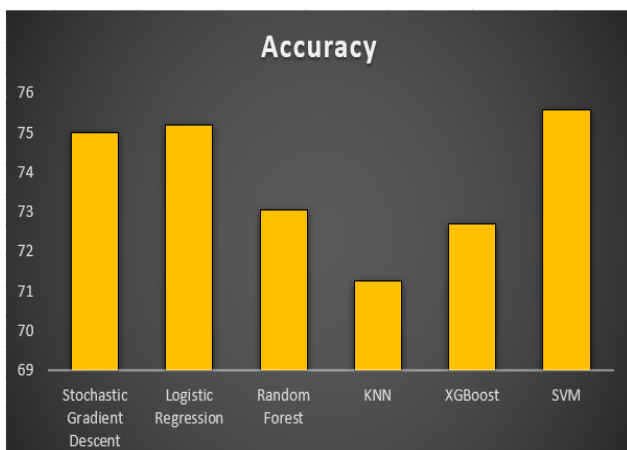
References

[1] Thahira M and Mubeena A K. "Comparative Study of Personality Prediction From Social Media by using Machine Learning and Deep Learning Method". *IJERT, ICCIDT – Conference Proceedings Volume 9, Issue 7,* pages 102 - 105, 2021.

[2] Aditi V. Kunte and Suja Panicker, "Using textual data for personality prediction: A Machine learning Approach", *IEEE, 2019 4th International Conference on Information Systems and Computer Networks (ISCON) GLA University*, *Mathura, UP, India*, pages 529 - 533 Nov 21 - 22, 2019.

[3] Jennifer Golbeck, Cristina Robles, Michon Edmondson, and Karen Turner, "Predicting Personality from Twitter", *2011, IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing,* pages 149 - 156, 2011.

[4] Bayu Yudha Pratama and Riyanarto Sarno, "Personality Classification Based on Twitter Text Using Naïve bayes, KNN and SVM", *IEEE, 2015 International Conference on Data and Software Engineering,* pages 170 - 174, Sept 27, 2018.

[5] Michael m. Tadesse, Hongfei lin, Boxu, and Liang yang, "Personality Predictions Based on User Behaviour on the Facebook Social Media Platform", *IEEEAccess*, 2018. i9

[6] Chaowei Li, Jiale Wan, Bo Wang, "Personality Prediction of Social Network Users", *IEEE, 2017 16th International Symposium on Distributed Computing and Applications to Business, Engineering and Science,* pages 84 - 87.

[7] Mr. R. Valanarasu, "Comparative Analysis for Personality Prediction by Digital Footprints in Social Media", *Journal of Information Technology and Digital World,* Volume - 03 No.02, pages: 77 - 91, 2021.

[8] S. Bharadwaj, S. Sridhar, R. Choudhary and R. Srinath, "Persona Traits Identification based on Myers - Briggs Type Indicator (MBTI) - A Text Classification Approach, " *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI),* Bangalore, pp.1076 - 1082, 2018.

[9] S. Chaudhary, R. Sing, S. T. Hasan and I. Kaur, "A comparative Study of Different Classifiers for Myers - Brigg Personality Prediction Model, " *IRJET*, vol.05, pp.1410 - 1413, 2018.

[10] Tommy Tandera, Hendro, Derwin Suhartono, Rini Wongso and Yen Lina Prasetio, "Personality Prediction System from Facebook Users", *2nd International Conference Computer Science and computational Intelligence 2017, ICCSCI 2017*, pages 604 - 611, 13 - 14 October 2017, Bali, Indonesia.

[11] Mitchelle. J, "Myers - Briggs Personality Type Dataset, Includes a large Number of people's MBTI Type", https: //www.kaggle. com/datasnaek/mbti - type, 4 September 2018