

A Machine Learning Approach for Personality Type Identification using MBTI Framework

Hira Shafi
Faculty of Engineering
Science and Technology,
Hamdard University,
Pakistan
hirashafi@live.com

Ahmed Sikender
Faculty of Engineering
Science and Technology,
Hamdard University,
Pakistan
m.sikander@hamdard.edu.pk

Ismail Mohamed Jamal
Faculty of Computer Science,
Jamhuriya University of
Science and Technology,
Somalia
ismail@just.edu.so

Jawwad Ahmad
Electrical Engineering
Department, Usman
Institute of Technology,
Pakistan
jawwad@uit.edu

Mohamed Abdelkader Aboamer
Department of Medical Equipment Technology,
Majmaah University,
Saudi Arabia
m.aboamer@mu.edu.sa

Abstract—In recent times, the prediction of personality traits with automated and programmed systems has caught human attention. Specifically, the use of multimodal data to predict personality types is the most considerable talk in artificial intelligence. There are a variety of techniques and methods available for personality type identification. The most popular and highly used personality type identifier is the Myers Briggs Type Indicator (MBTI) type indicator among all methods. In this paper, an exhaustive comparative analysis of all machine learning classical algorithms implementing the MBTI framework will be presented by giving a numerical and graphical representation of performance measures. To experience this study, a supervised machine learning approach is used to perform and analyze different classifiers using the phenomena of MBTI. The models are learned from a dataset to make predictions. The results show that the Ensemble Bagged Trees algorithm gives an overall good training accuracy of 98.4% and test accuracy of 70.75% at a moderate prediction speed of 11 K - Obs/sec by taking a training time of 14 sec. Other than that Coarse Tree algorithm in training time is 0.94009/sec and prediction speed 390 (K - Obs/sec), Fine KNN and Weighted KNN algorithm in training accuracy of 99.20% and Ensemble Boosted Trees algorithm in testing accuracy of 75.51% shows the efficient outcome respectively.

Keywords— Machine learning, dataset, classifiers, accuracy, algorithm, prediction, supervised learning, training

I. INTRODUCTION

Understanding human behavior and intellect is commonly controlled and supervised by the psychological aspects of how people think and perceive anything or any action. For instance, the situation and condition may vary but lead to an inference of personality indicator. There are a variety of personality type indicators that assist in identifying human personality by using human and psychological factors.

In this current era of technology, Machine Learning and Deep Learning have emerged as valuable and beneficial techniques to predict accurate findings and perform detailed and complex tasks with efficiency, and accuracy.

Machine learning trains the neural network for the classification and identification of distinct features. For such purposes, Machine learning classifiers are used. These classifiers are models used for training and testing data provided for the analysis. The analysis can be carried with all machine learning classifiers implemented on the personality identifier frameworks.

The most commonly used method is MBTI that contains Myers-Briggs 16 personalities which are derived from Carl Jung's theory of psychological types and commonly as personality types in which four key factors to categorize human personality are I- Introversion and E- Extraversion, S- Sensing and I- Intuition, T- Thinking and F- Feeling, and J- Judging and P- Perceiving. From these four basic dimensions, the MBTI 16 personality types are derived.

II. LITERATURE REVIEW

Human personality plays an important role in any society. Personality type prediction is nowadays a useful technique that most of the organizations, cooperate firms, military and educational sectors used to understand people. For personality type identification mainly in psychological terms, there is not a good number of datasets found.

Most of the data is based on security concerns and limited to access while publically available data which is dedicated to research and academics is available in textual and categorical forms.

For instance, the categorical datasets are further divided into personality type indicator frameworks most popularly Myers Briggs Type Indicator namely MBTI and Big 5 factors.

The interaction of digital and computerized mediums for humans is a new advanced era when it comes to artificial intelligence.

In most cases, it is required to take analysis and ask people about the trust level they consider for machine learning models specifically when it comes to psychological use.

In the research study of [10] this aspect is discussed. It was noted that humans trust the machine learning models in both observations and stated accuracy. The stated accuracy is dependent on observed accuracy and hence the overall effect of stated accuracy may get changed.

Understanding the human personality with his lifestyle is one of the key work findings of [11] in which the machine learning models implemented on textual data that give good accuracy mapping personality, lifestyle and mental health of individuals.

The study [12] shows the mental health analysis based on human stress issues in COVID'19 pandemic. The research study of [15] [16] is dedicated to finding MBTI personality types of social media users with reference to their content and posts.

This study focused on machine learning techniques and the MBTI framework. The same technique was performed in [17] with the Turkish language.

III. RESEARCH OBJECTIVE

The main objective of the research is to compare multiple Machine Learning algorithms by using the MBTI dataset at a certain type of categorical data and analyzing the results in terms of training accuracy, testing accuracy, training time and prediction speed.

IV. METHODOLOGY

The impact of analyzing and implementing a machine learning algorithm to predict MBTI personality types on a particular dataset has already been performed by different pre-trained models. In the personality type prediction, the algorithm must have good accuracy and efficiency. To find the comparative analysis of these models an exhaustive study of outcomes of each classifier in terms of training, prediction speed, and testing accuracy is proposed as the relevant measures of performance. This research study covers the comparison among all the Machine Learning classifiers and selects the best possible algorithms. To analyze the performance of each machine learning model, the following process is followed:

A. Data Collection

The MBTI typology dataset is used that contains more than 2800 nominal values of which 26% are INFJs. The dataset used is a crowd-sourced database of Rev. Emmy Kegler's personality sets

B. Applying all machine learning classifiers

For experiment and analysis, the machine learning classical models KNN, SVM, Tree, Ensemble, and Discriminant are used to examine the performance.

C. Comparative analysis of these classifiers

The machine learning models are trained and tested on the MBTI dataset and the analysis is based on training and testing accuracy and prediction speed along with training time.

The results are comprised of the confusion matrix, testing curves, testing accuracy, and ROC curves. Each model is comprehensively analyzed in form of a Confusion Matrix with True Positive and False Negative rates.

V. SIMULATION AND RESULTS

In this research, the MBTI typology dataset is used which contains 2802 number of instances and 10 attributes. The dataset is fragmented into 70% training and 30% testing data.

The fragmentation of the data is held through random permutation so the values are gathered randomly. Initially, the data was a mix of textual and numeric values. To make it nominal, the dataset mapping is performed.

In which, each unique value is labelled with a number so that the resultant dataset would be categorical. The Machine Learning classical models are trained at MATLAB. Each model is trained and executed on the classifier learner feature.

All implementation, network learning, and visualization are done in Matlab R2018b. The network training is executed in a single CPU environment.

The results are observed based on training time, prediction speed, and training and testing accuracies.

In which some algorithms show good training time and prediction speed. Other than that, some algorithms have high training accuracy and testing accuracy.

Table 1 Results of Machine Learning Algorithms

Machine Learning Algorithms	Training Accuracy (%)	Testing Accuracy (%)	Prediction Speed (K-Obs/sec)	Training Time (sec)
Fine Tree	43.7	71.58	330	2.2743
Medium Tree	34.6	75.39	300	1.1414
Coarse Tree	26.6	67.06	390	0.94009
Linear Discriminant	27.2	66.47	79	0.71396
Quadratic Discriminant	0	0.00	0	0
Linear SVM	26.7	65.04	8.4	4.0409
Quadratic SVM	29.6	66.35	2.2	17.01
Cubic SVM	68.4	62.43	3.8	21.746
Fine Gaussian SVM	82.2	65.99	0.85	9.648
Medium Gaussian SVM	32.6	65.04	1.3	9.3583
Coarse Gaussian SVM	25.4	64.09	1.3	12.619
Fine KNN	99.2	61.95	31	7.3962
Medium KNN	36.3	63.14	8.4	7.2358
Coarse KNN	27.9	65.52	5.2	8.1314
Cosine KNN	36.8	63.50	14	8.4758
Cubic KNN	36.5	63.50	1.1	11.969
Weighted KNN	99.2	65.28	7	12.357
Ensemble Boosted Trees	35.5	75.51	13	14.644
Ensemble Bagged Trees	98.4	70.75	11	14.473
Ensemble Subspace Discriminant	26.5	65.87	7.6	14.232
Ensemble Subspace KNN	90.3	71.70	4.9	15.223
Ensemble RUSBoosted Tree	18.7	53.98	27	15.059

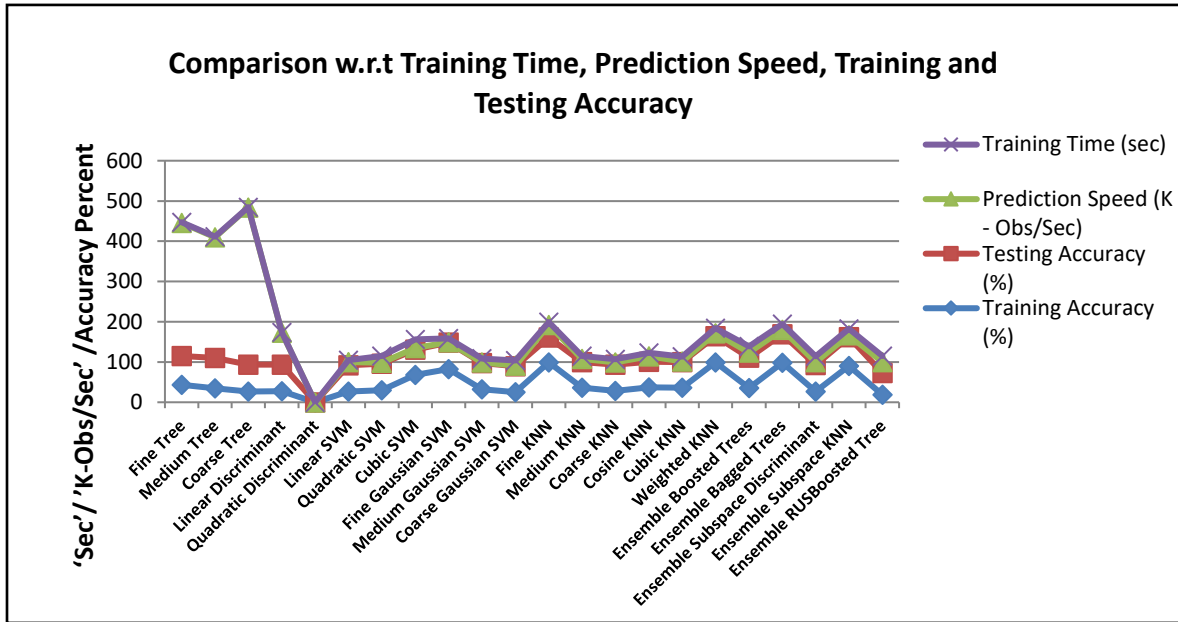


Fig. 1 Graphical representation of all Machine Learning Algorithms used in MBTI personality type identification

VI. CONCLUSION

The results depict that the Ensemble Bagged Trees algorithm gives overall good results with 98.4% training accuracy and 70.75% test accuracy at a moderate prediction speed of 11 K - Obs/sec by taking a training time of 14 sec. Hence, an Ensemble Bagged Trees algorithm is considered as a feasible option in the context of Machine Learning Classifier to facilitate computer-aided discovery for the Myers-Briggs personality type or MBTI personality type in general. In terms of training time and prediction speed, the Coarse Tree algorithm is the most efficient with the training time 0.94009/sec and prediction speed 390 (K - Obs/sec). Fine KNN algorithm and Weighted KNN algorithm are considered with the high training accuracy of 99.20% and Ensemble Boosted Trees algorithm with the testing accuracy of 75.51%.

VII. FUTURE WORK

The personality type identification is a broader area of exploration. The research is open for variety of data and techniques to come with new and effective findings. Presented study is focused on nominal data and the values are gathered from a global data inventory. However, the study could be performed on textual data using NLP technique. Further scientifically included personality type test for a real-time application on the best performance algorithms to analyze personality trait of a region or population.

REFERENCES

- [1] Yin, Ming, Jennifer Wortman Vaughan, and Hanna Wallach. "Understanding the effect of accuracy on trust in machine learning models." Proceedings of the 2019 chi conference on human factors in computing systems. 2019.
- [2] Bharadwaj, Srilakshmi, et al. "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). IEEE, 2018.
- [3] O'Brien, Terrance P., Leonhard E. Bernold, and Duane Akroyd. "Myers-Briggs type indicator and academic achievement in engineering education." International Journal of Engineering Education 14.5 (1998): 311-315.
- [4] Montequín, V. Rodríguez, et al. "Using MBTI for the success assessment of engineering teams in project-based learning." International journal of technology and design education 23.4 (2013): 1127-1146.
- [5] Cui, Brandon, and Calvin Qi. "Survey Analysis of Machine Learning Methods for Natural Language Processing for MBTI Personality Type Prediction." (2017).
- [6] Mehta, Yash, et al. "Recent trends in deep learning based personality detection." Artificial Intelligence Review (2019): 1-27.
- [7] Fernandes, Bruno, et al. "An adjective selection personality assessment method using gradient boosting machine learning." Processes 8.5 (2020): 618.
- [8] Sönmezöz, Kaan, Özgür Uğur, and Banu Diri. "MBTI Personality Prediction With Machine Learning." 2020 28th Signal Processing and Communications Applications Conference (SIU). IEEE, 2020.
- [9] Sönmezöz, Kaan, Özgür Uğur, and Banu Diri. "MBTI Personality Prediction With Machine Learning." 2020 28th Signal Processing and Communications Applications Conference (SIU). IEEE, 2020.
- [10] Goetz, Monika L., et al. "An Examination of Myers-Briggs Type Indicator Personality, Gender, and Career Interests of Ontario Veterinary College Students." Journal of veterinary medical education 47.4 (2020): 430-444.

- [11] Amirhosseini, Mohammad Hossein, and Hassan Kazemian. "Machine learning approach to personality type prediction based on the myers-briggs type indicator®." *Multimodal Technologies and Interaction* 4.1 (2020): 9.
- [12] Nadler, David W. "Decision support: using machine learning through MATLAB to analyze environmental data." *Journal of Environmental Studies and Sciences* 9.4 (2019): 419-428.
- [13] Ayodele, Taiwo Oladipupo. "Types of machine learning algorithms." *New advances in machine learning* 3 (2010): 19-48.
- [14] Hansen, Jakob Vogdrup. "Combining predictors: comparison of five meta machine learning methods." *Information Sciences* 119.1-2 (1999): 91-105.
- [15] Omary, Zanifa, and Fredrick Mtenzi. "Machine learning approach to identifying the dataset threshold for the performance estimators in supervised learning." *International Journal for Infonomics (IJ)* 3.3 (2010): 314-325.
- [16] Qi, Xinbo, et al. "Applying neural-network-based machine learning to additive manufacturing: current applications, challenges, and future perspectives." *Engineering* 5.4 (2019): 721-729.
- [17] Osisanwo, F. Y., et al. "Supervised machine learning algorithms: classification and comparison." *International Journal of Computer Trends and Technology (IJCTT)* 48.3 (2017): 128-138.
- [18] Jiang, Tammy, Jaimie L. Gradus, and Anthony J. Rosellini. "Supervised machine learning: a brief primer." *Behavior Therapy* 51.5 (2020): 675-687.
- [19] Mahesh, Batta. "Machine Learning Algorithms-A Review." *International Journal of Science and Research (IJSR)*. [Internet] 9 (2020): 381-386.
- [20] Alexander III, Leo, Evan Mulfinger, and Frederick L. Oswald. "Using big data and machine learning in personality measurement: Opportunities and challenges." *European Journal of Personality* 34.5 (2020): 632-648.
- [21] Flesia, Luca, et al. "Predicting perceived stress related to the Covid-19 outbreak through stable psychological traits and machine learning models." *Journal of clinical medicine* 9.10 (2020): 3350.
- [22] Khan, Alam Sher, et al. "Personality Classification from Online Text using Machine Learning Approach." *International Journal of Advanced Computer Science and Applications* 11.3 (2020).
- [23] Nisha, Kulsum Akter, et al. "A Comparative Analysis of Machine Learning Approaches in Personality Prediction Using MBTI." *Computational Intelligence in Pattern Recognition*. Springer, Singapore, 2022. 13-23.
- [24] Abidin, Nur Haziqah Zainal, et al. "Improving Intelligent Personality Prediction using Myers-Briggs Type Indicator and Random Forest Classifier." *INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS* (2020).
- [25] Garg, Shruti, and Ashwani Garg. "Social Sciences & Humanities Open."
- [26] Remaida, Ahmed, et al. "Personality traits analysis using Artificial Neural Networks: A Literature Survey." *2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*. IEEE, 2020.