Journal of Information Systems Engineering and Management
2025, 10(14s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Ensemble Approach for Human Personality Classification using Textual Data

[1]Niranjan Prajapati, [2]Dr. Harikrishna Jethva

[1]Research Scholar, Gujarat Technological University, Ahmedabad, Gujarat
niranjan.prajapati88@gmail.com
[2]Government Engineering College, Patan, Gujarat
hbjethva@gmail.com

ARTICLE INFO

Received: 22 Nov 2024

Revised: 14 Jan 2025

Accepted: 28 Jan 2025

ABSTRACT

Nowadays the users are active on social media platforms, such as Blogs, Twitter, Facebook, Instagram, etc. Users use these platforms to share their views on movie, current affairs, blogs, and group discussions. The views shared by users can be the charts, images, textual data, etc. The views or contents shared contain the inherent features of the users like, the persons they are following, the discussions on the topics. It will also help to provide insight of the personnel aspects of the users in the form of the job satisfaction, life expectations, the career preferences, psychological stage. Without the taken mandatory tedious tests or the feedback forms, the social media contents give the better classification of personality traits. Users share their views on the social media platforms as they are thinking about it without bothering about the tests taken by the organization. Personality classification process involves the preprocessing of the social media contents for feature extraction, mapping of the features to the personality model. Big five factor model is considered as the standards for classification of personality traits. In this paper we used the improved ensemble technique using gradient bagging for personality classification and result shows improved result as compared to recent methods of personality classification.

Keywords: Personality Classification, Radom forest, Natural Language Processing, Big five factor model, Myers-Briggs Type Indicator, Gradient Bagging

1. INTRODUCTION

The mix of a person's behavior, motivation, mental processes, and emotion is referred to as their characteristics. These exemplify a person's personality. Our personalities have a big impact on our decisions, health, aspirations, and preferences. There are numerous significant practical uses for the capacity to foresee one's personality traits. The "big five" personality traits are: conscientiousness, openness, agreeableness, extroversion, and neuroticism [7]. They are frequently referred to as "OCEAN" and occasionally "CANOE." These five personality qualities are used to distinguish between people and to guide decision-making. They cover a wide range of human activity. The model is now used by HR professionals to assess job candidates and by marketers to comprehend the target markets for their products.

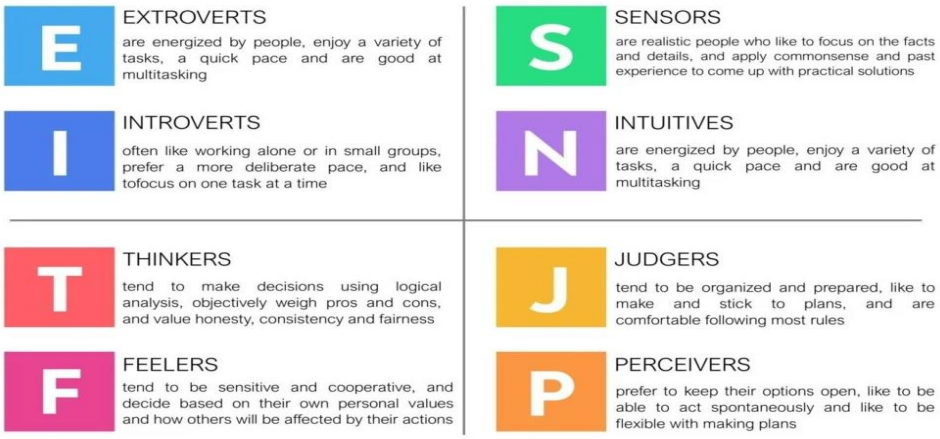

Figure 1. MBTI Personality Types Key [1]

A most famous and often used personality type indicator is the Myers-Briggs Type Indicator (MBTI). With four binary categories and a total of 16 types, it depicts how people act and engage with the environment around them. The proportion of a person's four types for each category, using the bolded identifying letter for each, makes up their MBTI personality type. For instance, someone with the personality type ENTJ would derive most of their energy from interacting with others (E), trust their gut and use intuition to interpret information in the world (N), think logically about their decisions (T), and live a carefully planned lifestyle (J) as opposed to one that is spontaneous.

The term "personality" refers to a person's self-perception, which shapes their behavior in a distinctive and dynamic way. As people learn, experience, and educate themselves, their behavior may evolve. As per Myers-Briggs Type Indicator®, an individual's inclinations are splitted into four categories, each of which represents anyone of the 16 personality types, using different combinations of the personality type key. These 16 personality types are depicted in Figure 2 as a result of interactions between a person's preferences.

Figure 2. MBTI 16 Personality types [1]

Gradient Bagging merges elements from both gradient boosting and bagging, forming a powerful ensemble method that improves predictive accuracy by lowering both bias and variance. Usually, classification and regression are employed to enhance models' performance. To understand Gradient Bagging fully, it's useful to first grasp the core concepts of bagging and gradient boosting:

1.1 Bagging (Bootstrap Aggregating):

Bagging is an ensemble learning technique where multiple models (usually of the same type, like decision trees) are trained on distinct data subsets [2]. To generate subsets sampling with replacement technique (bootstrap sampling)

is used. After training, the results of all models are combined, often through averaging in regression tasks or majority voting in classification tasks. The primary advantage of bagging is that it reduces variance, making the model less sensitive to fluctuations in the data, which helps prevent over-fitting.

1.2 Gradient Boosting:

Gradient boosting, in contrast, it train models in a sequential manner, with each new model concentrating on the residuals to try to fix the errors of the prior model (the difference between predicted and actual values). Typically, the models used in gradient boosting are weak learners (e.g., shallow decision trees), and combining these weak learners results in overall stronger model. While gradient boosting reduces bias, it can sometimes over-fit if not tuned properly, as it tends to be more sensitive to noisy data compared to bagging.

2. RELATED WORK

Two most well-liked personality models are The Big Five or MBTI, worldwide, have been the focus of the majority of studies on personality prediction. “A personality trait is a distinctive way of thinking, feeling, or acting that has a tendency to remain constant over time and in pertinent situations”, explained by Soto[7]. This justification allows us to describe The Big Five personality model as a collection of five broad trait dimensions: (1) extroversion, (2) agreeableness, (3) conscientiousness, (4) neuroticism, and (5) openness [7]. As a matter of truth, the Big Five personality model proposes five broad dimensions that are commonly utilized to define human nature and uses adjectives from everyday language. In contrast, the four aspects of the Myers-Briggs Type Indicator® are introversion/extroversion, sensing/intuition, thinking/feeling, and judging/perceiving. These dimensions are used to classify personality types into 16 groups. There is a dearth of research on personality type prediction using textual data. Notable progress has been made through the use of machine learning in this quest. MBTI personality types have been anticipated by the using both neural networks and traditional machine learning techniques. Research suggests that Greater applications of the MBTI model exist, particularly in industry and for personality type self-discovery, notwithstanding debate regarding the reliability and validity of these two models [10].

Among the first studies to apply machine learning to predict personality was by Golbeck et al [5]. As per the MBTI personality type indication and taking into account the information posted on their Twitter, they could correctly estimate a user's personality type. Komisin and Guinn [11] rely on word choice to forecast a personality type of an individual using the Support Vector Machine (SVM) and Naive Bayes approaches. Their database was created using the in-class writing samples and MBTI personality types of 40 graduate students. On their tiny dataset, they examined the effectiveness of these two strategies and found that the Naive Bayes technique outperformed SVM. Tandera et al. [6] perform on the user's Facebook information, the Big Five personality model plus some deep learning architecture is able to forecast a person's personality. Their model successfully outperformed the accuracy of earlier research that were comparable and employed standard machine learning techniques, according to comparisons they made between the results of their method and that of those studies. In order to create a classifier that could predict people's MBTI personality types based on text samples from their social media posts, Hernandez and Knight [8] used several recurrent neural networks (RNNs), such as basic RNNs and gated recurrent units (GRU), which acts as a long short-term memory (LSTM), the gating mechanism in recurrent neural networks, and bidirectional LSTM. They used the Kaggle Myers-Briggs Personality Type Dataset for their investigation.

It was discovered that to predict personality types as per the MBTI or Big Five personality type models, many methods have been used, such as support vector machine (SVM), logistic regression, random forests, K Nearest Neighbour (KNN), Naive Bayes, and Linear Discriminant analysis (LDA). According to reports, the MBTI model contains more widely used by researchers, and given the debate around the validity and reliability of these two models, the MBTI model has greater applications across a various fields [10]. Additionally, it was observed that this field didn't use several potent techniques of machine learning like gradient boosting. Because of its great degree of adaptability to the specific requirements of the application, the machine learning technique known as gradient boosting has found considerable success in a variety of real-world practical applications. According to Freund and Schapire [2], boosting is a technique based on combining imprecise, loose rules of thumb, which allude to general concepts not meant to be strictly correct or dependable in every circumstance, to create a very accurate prediction rule.As a result, Extreme Gradient Boosting, a boosted tree technique that adheres to the Gradient Boosting principle, performs better since it uses a more regularized model formalization to limit over-fitting [17]. We were certain that Extreme Gradient Boosting would be successful in this field due to the nature of personality prediction,

the requirement for classification for MBTI personality types, and the way that Extreme Gradient Boosting can handle classification tasks. As a result, this study used the Extreme Gradient Boosting approach.

3. GRADIENT BAGGING

Gradient Bagging combines the principles of bagging and gradient boosting, offering the benefits of both methods. It works in general: Step 1) Bootstrap Sampling: multiple bootstrap samples are taken from the original dataset. Step-2) Training Multiple Models: Each bootstrap sample is used for training an individual model using gradient boosting. Here, the main distinction is that models are trained concurrently on various samples, rather than sequentially as in traditional gradient boosting. Step-3) Prediction Aggregation: After training, the models' predictions are aggregated. This could mean averaging the predictions in regression tasks or using voting for classification tasks.

The parallel training of models distinguishes gradient bagging from traditional gradient boosting, where models are trained sequentially.

The steps for the loss function and use of gradient is discussed as follow:

1. Initialize the weak classifier

$$f_0(x) = argmin \sum_{i=1}^{N} L(y_i, \rho) \tag{1}$$

y_i is the true value of the sample; ρ is a constant. L denotes loss function:

$$L(y, f(x)) = ((y - f(x))^2 \tag{2}$$

The total loss of all N samples is

$$L_{all} = \sum_{i=1}^{N} L(y_i, f_m(x_i)) \tag{3}$$

where $f_m(x_i)$ is the mth predicted value.

2. Minimize the loss function

Negative gradient function is

$$-g(x_i) = -\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \tag{4}$$

We construct a negative gradient fitting function $h(x_i; \alpha)$ to fit the negative gradient $-g(x_i)$

$$\alpha_m = argmin \sum_{i=1}^{N} \left(-g(x_i) - \beta h(x_i; \alpha)\right)^2 \tag{5}$$

where α_m and $g(x_i)$ denote residual parameter and gradient, β and α are coefficients.

3. Optimize weight factor

$$\beta_m = argmin \sum_{i=1}^{N} L(y_i, f_{m-1}(x_i) + \beta h(x_i; \alpha_m)) \tag{6}$$

where β_m is weight coefficient; $f_{m-1}(x_i)$ is fitting function of the (m-1)th iteration.

4. Update forecast function

$$f_m(x) = f_{m-1}(x) + \beta_m h_m(x; \alpha_m) \tag{7}$$

$f_m(x)$ is fitting function of the current iteration; $\beta_m h_m(x; \alpha_m)$ is the (m-1)th iterative negative gradient fitting function.

Combining the best aspects of bagging and gradient boosting, gradient bagging is an advanced ensemble technique. It reduces both bias and variance, offering enhanced predictive performance while being more robust to noise compared to traditional gradient boosting methods. However, it comes with higher computational costs and complexity in terms of interpretability and hyper parameter tuning. When applied thoughtfully, it can be an extremely effective tool for addressing a variety of machine learning issues.

4. PERSONALITY CLASSIFICATION FRAMEWORK

The personality classification system proposed here consists of four main processes: (1) Categorize type indicators in four dimensions (2) Preprocessing of dataset (3) Vectorize with count and Term Frequency – Inverse Document frequency (4) Personality Classification applying Gradient Bagging.

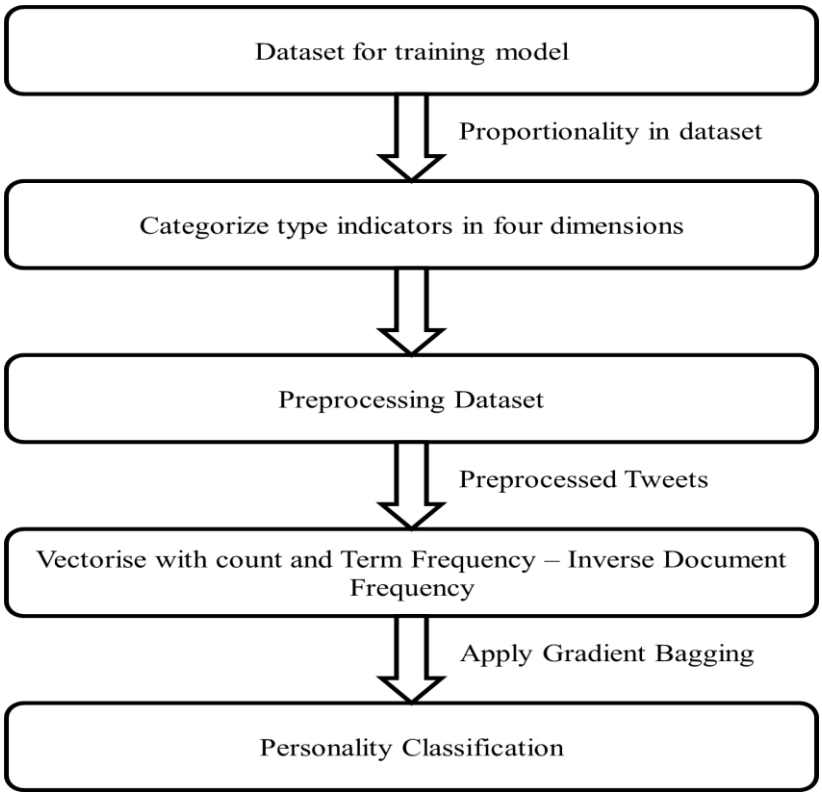

Figure 3: Personality Classification system flow diagram

Detailed flows for the system flow diagram shown in Figure 3 are as below:

Step 1: Proposed system's development tools for are setup.

Step 2: Dataset for training model is collected.

Step 3: Proportionality in Dataset, distribution of the MBTI personality types in the dataset was determined.

Step 4: Categorize the type indicators in four dimensions of Introversion (I) and Extraversion (E), iNtuition (N) and sensing (S), Thinking (T) and Feeling (F), Judging (J) and Perceiving (P)

Step 5: Preprocessing dataset to remove punctuation marks, urls, symbols, etc. It classifies the emails into appropriate labels.

Step 6: To find the words that appeared in 10% to 70% of the postings, vectorize with count and term frequency-inverse document frequency (TF-IDF).

Step 7: Apply gradient bagging to perform the personality classification.

We provide a detailed description for the above mentioned steps in the next section.

5. ALGORITHM FOR PERSONALITY CLASSIFICATION SYSTEM

5.1 Tools setup for development

The development approach made use of the distributed gradient boosting module for Python and the Natural Language Processing Toolkit (NLTK). The NLTK toolbox for NLP is a potent resource for writing Python applications to deal with data from human languages. Other Python libraries that were used include Pandas, Numpy, re, Seaborn, Matplotlib, and Sklearn.

5.2 Dataset for Training the Model

In this study, 8675 rows of the Myers-Briggs personality type dataset from Kaggle, which is available to the general public, were used. Each row in this dataset has two columns. A person's MBTI personality type is written in the first column, and fifty posts from their accounts on social media are written in the second column. Three pipe characters [16] have been used to separate each post. This information was collected from forum members who first completed a questionnaire to determine their MBTI type and then engaged in conversation with other members of the forum [8].

5.3 Dataset proportionality

This stage involved utilizing the Python data visualization tool seaborn and the Python 2D charting module matplotlib to ascertain the distribution of the MBTI personality types in the dataset. Figure4 shows each MBTI personality type's number of occurrences in the dataset.

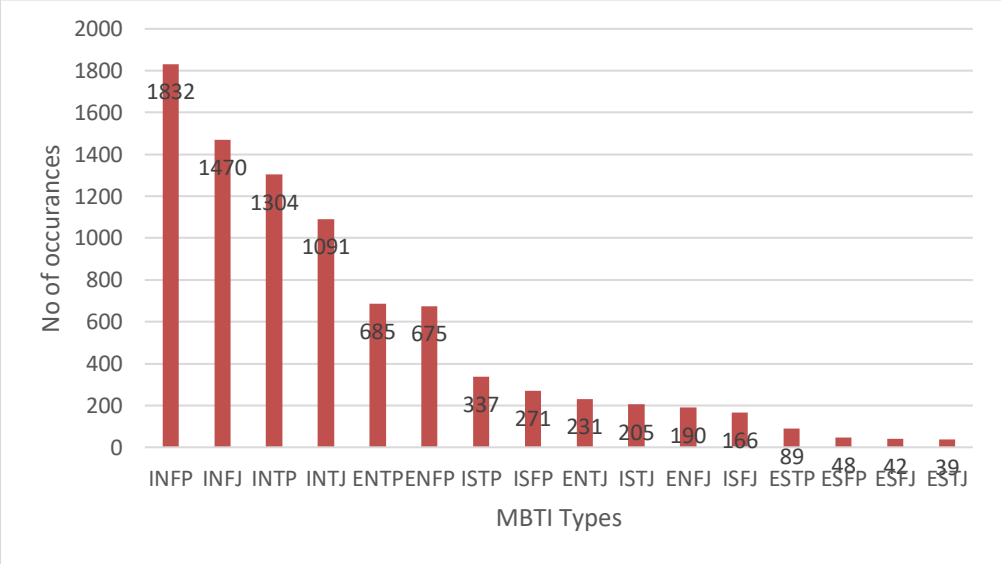

Figure 4. Each MBTI personality type's number of occurrences

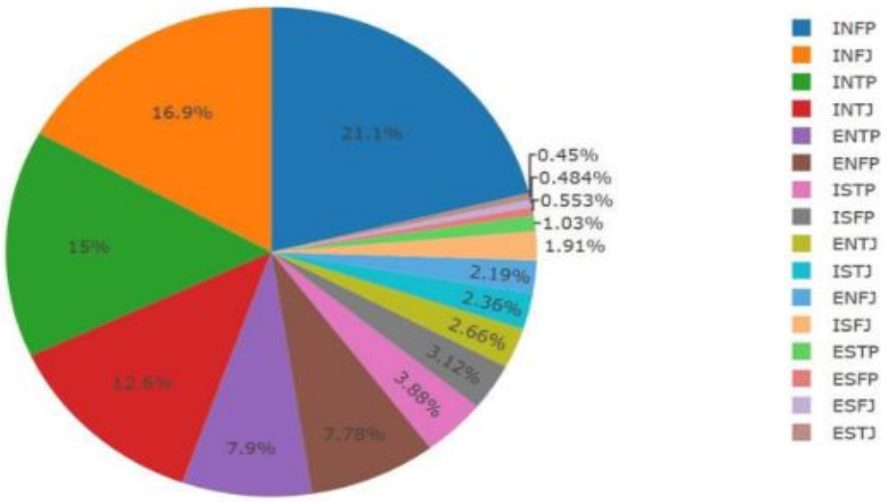

Figure 5. MBTI personality type's percentage of occurrences in dataset

Figure 4 and Figure 5 indicate an uneven representation of MBTI types in the dataset that is out of line with the proportions of those kinds in the general population shown in Table 2. It was therefore clear that the dataset would need to be cleaned up a bit to improve the accuracy of each MBTI type's proportional representation.

Table 1. Distribution of Personality type in dataset

Personality Type	Frequency in dataset
ISFJ	13.80%
ESFJ	12.30%
ISTJ	11.60%
ESTJ	8.70%
ISFP	8.80%
ESFP	8.50%
ISTP	5.40%
ESTP	4.30%
INFP	4.40%
ENFP	8.10%
INTP	3.30%
ENTP	3.20%
INTJ	2.10%
ENFJ	2.50%
INFJ	1.50%
ENTJ	1.80%

5.4 The type indicators are splitted into four categories.

The four groups of type indicators were created in order to comprehend how they were dispersed across the dataset. Introversion (I) and extraversion (E) were the first two categories, intuition (N) and sensing (S) was the second, thinking (T) and feeling (F) was the third, and judging (J) and perceiving (P) was the fourth. As a result, one letter will return for each category, resulting in a string of four letters that each stand for one of the MBTI's 16 personality types. For instance, INTJ would be the appropriate personality type if the first category got I, the second got N, the third got T, and the fourth got J. Distributions across type indicators are displayed in Figure 6 and Table 2.

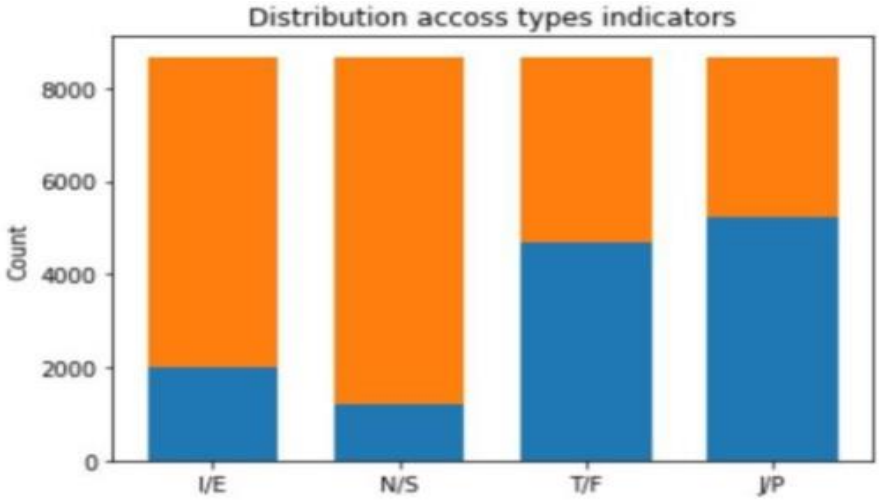

Figure 6. Personality Traits Distribution in Dataset

According to Figure 6, Extroversion is distributed significantly more evenly than Introversion for the first bar line in figure 6 of I/E. Similar to the first category, Sensing (S) has a substantially higher distribution than Intuition-N for the second bar line, N/S. Figure 6 also show that Thinking-T is distributed significantly more than Feeling-F in the third category, which is T/F. For the fourth category, Judging (J)/Perceiving (P), Judging (J) has a larger distribution than Perceiving (P).

Table 2. Distributions for type indicators

Type Indicator	Distribution
Introversion (I)	1999
Extroversion (E)	6676
Intuition (N)	1197
Sensing (S)	7478
Thinking (T)	4694
Feeling (F)	3981
Judging (J)	5241
Perceiving (P)	3434

The strength of variables and correlations can be determined using the Pearson correlation coefficient. It is possible to determine which pairs are most correlated by looking at the correlation between each of random variable (X_i) and every other value in the table, including (Xj) in a correlation matrix. It is necessary to calculate the coefficient value that can be between 0.00 and 1.00, in order to comprehend how substantial the link is among two variables. Figure 7 shows the effectiveness with which personality type indicators correlate.

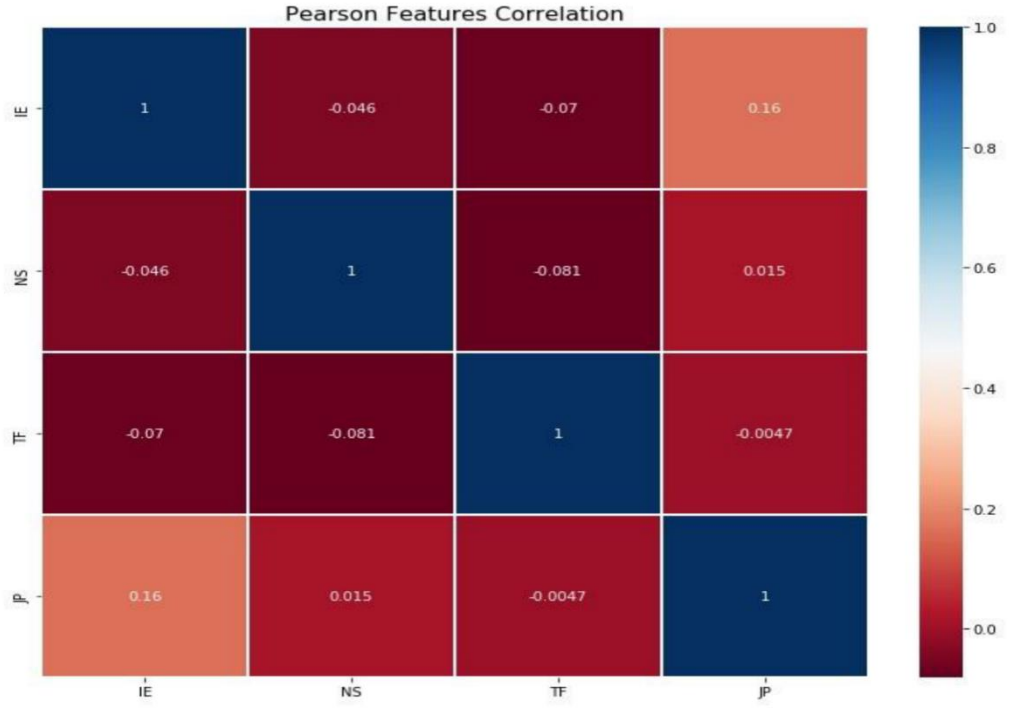

Figure 7. Correlation Coefficient (Pearson) between measures of personality types

5.5 Dataset prep-processing

As mentioned before, this dataset's data came from an online discussion board, and it became obvious after analyzing its content that some words needed to be deleted. The main cause of this was the dataset's uneven representation of MBTI types, which did not correspond to the proportions of those kinds in the general population. It was found that it was the case because the MBTI kinds were repeated far too frequently in the posts of the Internet forum where the data was gathered to debate personality types. The model's accuracy could be impacted by this as well. The MBTI types were consequently eliminated from the dataset using NLTK. Following this procedure, the MBTI personality types' distribution within the dataset was once more assessed, and it was

noted that the MBTI types' representation in the sample is consistent with their prevalence in the general population. Additionally, the dataset was cleared of all stop words and URLs. The text was then lemmatized, or the inflected versions of the words were changed into its root words, to further improve the dataset's meaning.

5.6 Count and Term Frequency-Inverse Document Frequency (TF-IDF) Vectorization

To find the words that were heavily used (ranging from 10% to 70%) of the postings, the Sklearn package was employed. Posts were first organized into a table of counts of token. In the subsequent step, the model gives a term-document matrix after learning the vocabulary dictionary. After that, the count matrix changes into a normalized TF-IDF representation that is given to the gradient bagging model. Finally, the entries include 791 words.

5.7 Classification Task

In machine learning, there are two categories of classification. The first type is to prove that classes or clusters exist in the data based on a collection of observations. The objective of the second type, where there may be several classes, is to develop a rule or set of rules that will add a new observation to one of the classes [17]. Unsupervised learning is the first type, and supervised learning is the second. [14].

The classification work was divided into four binary classification tasks after being segmented into 16 classes because each type of MBTI is made up of four binary classes. As per the personality model of MBTI, each binary classification represents a different component of personality. As a consequence, four separate binary classifiers that each focus on a different component of personality was trained. As a result, a model for each type of indication was generated separately in this step. Both the MBTI type indicators and Term Frequency-Inverse Document Frequency (TF-IDF) were carried out. In the TF-IDF representation, postings were represented by variable X, and the binarized MBTI type indicator was represented by variable Y.

5.8 Developing Gradient Bagging Model for the Dataset

Gradient Bagging, implemented using NumPy and scikit-learn, combines the principles of gradient boosting and bootstrap aggregating (bagging). It aims to increase the accuracy and stability of prediction models. Here's how it typically works:

Data Sampling:

Multiple data subsets are generated (bootstrapping) are generated using random sampling with replacement. Each subset is identical in size to the original dataset but may have instance repetition and omit others.

Base Model Training:

A base model, usually a decision tree, is trained on each bootstrapped datasets. Unlike standard bagging where models are trained independently, in Gradient Bagging, the training of later models is influenced by the errors of previous models, similar to gradient boosting.

Error Correction:

After training completion each base model, the algorithm focuses on the instances that were incorrectly predicted. It assigns higher weights to these instances, ensuring that subsequent models focus more to them. This iterative process of error correction is characteristic of gradient boosting.

Aggregation:

Final output is produces by combining response of all base models. This is typically done by averaging the predictions (for regression) or by voting (for classification).

From the sklearn library, the train_test_split() function is used to taught individual MBTI type indications, data was partitioned into training and testing datasets. 30% of the data were utilized as the test set, while 70% were used as the training set. The model was fitted using the training data, and predictions were made using the testing data. After that, the effectiveness of gradient bagging model during training on testing dataset was evaluated, and early termination was monitored. After this step, for smaller numbers, it will be necessary to build additional trees and drop the rate of learning in gradient bagging to 0.1 or less. Additionally, trees depth should be set between 2 and 8, as there is little value to having deeper trees. Additionally, 30% to 80% of the dataset of training should be designed for row sampling. Consequently, tree_depth in the newly constructed gradient bagging was specified, and gradient

bagging parameters were set up. The n_estimators is set to 200, max_depth is set to 2, nthread is set to 8, and learning_rate is set to 0.2.

The data was partitioned into training and testing data after MBTI type indicators were trained. The training data were used to fit the model, and the testing data were used to make predictions. In this stage, the efficacy of the gradient bagging model on the test dataset was assessed once again and the output is presented.

The best approach for configuration of the model for the best performance was found by using the scikit-learn library's ability to search through combinations of parameters. This is what the gradient bagging model refers to as hyper parameter tuning. Therefore, the factors to be consider for tuning are (1) the size of trees and number of trees, (2) number of trees and the rate of learning, and (3) the row and column sub-sampling rates.

6. EXPERIMENTAL RESULTS

Type indicators of MBTI were trained independently after developing the gradient bagging model, and the data was partitioned into training and testing datasets. Predictions for testing data were generated after the model had been fitted for the training set of data. After that, forecasts were assessed.

Predictions were once again assessed after the XGBoost model's tree_depth configuration.

Table 3 shows that following configuration, the model's accuracy and performance were significantly increased in the Feeling (F)-Thinking (T) category while only marginally improved in the Introversion (I)-Extroversion (E) category. However, there was a modest decline in accuracy in the intuition (I)-sensing (S) and judging (J)-perceiving (P) categories.

Table 3. Accuracy prediction comparison between before and after configuration

Binary Class	MBTI Personality Type	Accuracy before Configuration	Accuracy after Configuration	Difference
IE	Introversion (I)– Extroversion (E)	78.17%	81.01%	2.84
NS	Intuition (I)– Sensing (S)	86.06%	87.90%	1.84
FT	Feeling (F)– Thinking (T)	71.78%	77.21%	5.43
JP	Judging (J)– Perceiving (P)	65.70%	72.02%	6.32

For the comparison of accuracy of predictions following configuration to the most recent and effective existing approach. Knight and Hernandez proposed this method[8]. They used the identical dataset as their research and followed the exact same pre-processing procedure. As a result, the research's comparison of their method and the one that was presented used the identical set of data. They constructed their classifier using a variety of recurrent neural networks (RNNs), including basic LSTM, GRU, RNN, and Bidirectional LSTM. They have used two distinct evaluation techniques: a post classification and a user classification. Test set was preprocessed by them for post classification and forecasted the class for each unique post. For each of MBTI dimension, they subsequently created confusion matrix and accuracy score. On the other side, classify people, they needed to find a way to refine the class forecasts of individual posts—all written by the same author—into a forecast for the author's class. They therefore averaged the value to either 0 or 1 for the rounded class probability forecasts across all of a user's posts in their corpus.

Model of recurrent neural network with user-driven classification methodologies performed more precisely than those with post-driven classification methodologies. Gradient Bagging's classification accuracy was compared with their classifier for recurrent neural network utilizing the user classification. In reality, we compared the

performance of the models in the same way, and therefore the evaluation technique was the same. The outcomes of this comparison are shown in Table 4.

Table 4. Comparison of the precision of the recurrent neural network model with the Gradient Bagging model.

Binary Class	MBTI Personality Type	Accuracy of Recurrent Neural Network	Accuracy of Extreme Gradient Boosting	Accuracy of Gradient Bagging	Difference in accuracy of Gradient Bagging and Recurrent Neural Network	Difference in accuracy of Gradient Bagging and Extreme Gradient Boosting
IE	Introversion (I)– Extroversion (E)	67.60%	78.17%	81.01%	13.41%	2.84%
NS	Intuition (I)– Sensing (S)	62%	86.06%	87.90%	25.9%	1.84%
FT	Feeling (F)– Thinking (T)	77.80%	71.78%	85.21%	7.41%	14.03%
JP	Judging (J)– Perceiving (P)	63.70%	65.70%	72.02%	8.32%	6.32%

Table 4 demonstrates that the Gradient Bagging classifier outperforms the recurrent neural network and extreme gradient boosting in terms of accuracy when categorizing MBTI personality types in three dimensions. Gradient Bagging has much better accuracy than recurrent neural network for the categories such as: intuition (N) - sensing (S), introversion (I) - extroversion (E), judging (J) - perceiving (P), and feeling(F) - thinking(T). Gradient Bagging has marginally better accuracy than the extreme gradient boosting. Therefore, for this dataset, the Gradient Bagging classifier outperforms the recurrent neural network and extreme gradient boosting in terms of overall performance.

7. CONCLUSION

Using the personality type indicator MBTI, the experiment has created a novel method for machine learning for automated prediction of personality type and meta program recognition. The development process made use of the gradient bagging package implemented in Python, and the Natural Language Processing Toolkit (NLTK) for the implementing algorithms of machine learning within the Gradient Bagging framework. Moreover, some libraries of Python used were Seaborn, Pandas, Sklearn, Numpy, re, Matplotlib, and others. The model's performance and accuracy were assessed using the identical dataset as the latest and effective current approach. The findings shows that, compared to other approaches already in use, the methodology described here offers higher reliability and accuracy. The current methodology greatly improve the accuracy of identifying the intuition (N)-sensing (S) and extrovert (E)-introvert(I) personality characteristics, as well as slightly improving the accuracy of identifying the Judging (J)-Perceiving (P) personality category, according to the implementation in this paper. This will helpful to psychologists and NLP practitioners to classify different personality types and the cognitive processes that go along with them.

REFERENCES

[1] Simple Psychology, https://www.simplypsychology.org/the-myers-briggs-type-indicator.html

[2] Schapire, R.E., Freund, Y., “A decision-theoretic generalization of on-line learning and an application to boosting”, Jou. of Computer and System Sciences 1997, 55, pp. 119–139.
[3] Nguyen, D., Rosé, C.P., Doğruöz, A.S., Jong, F.D, “Computational sociolinguistics: A survey”, Computational Linguistics 2016, 42, pp 537–593.
[4] Gjurkovic M., Snajder J., “Reddit: A gold mine for personality prediction”, In Proc. of the 2nd Workshop on Compu. Modelling of People’s Opinions, Personality and Emotions in Social Media, New Orleans, 6 June 2018, pp. 87–97.
[5] Golbeck J., Turner K., Edmondson M., Robles C., “Predicting personality from Twitter”, In Proc. of the IEEE 3rd International Conference on Privacy, Security, Risk and Trust and IEEE 3rd International Conference on Social Computing, Boston, 9–11 October 2011.
[6] Tandera T., Prasetio Y., Wongso R., Suhartono D., “Personality prediction system from Facebook users”, In Proceedings of the Second International Conference on Comp. Sci. and Compu. Intelli., Bali, Indonesia, 13–14 October 2017.
[7] Soto, C.J.,“Big Five personality traits”, In the SAGE Encyclopedia of Lifespan Human Development; Thousand Oaks, 2018, pp. 240–241.
[8] Hernandez R., Knight I.S., “Predicting Myers-Bridge Type Indicator with text classification”, In Proceedings of the thirty first Conference on Neural Info. Processing Syst., Long Beach, 4–9 December 2017.
[9] Li C., Wan, B., Wang j., “Personality Prediction of Social Network Users”, In Proceedings of the Sixteenth International Sympo. on Distri. Compu. and App. to Business, Engineering and Science, Anyang, China, 13–16 October 2017.
[10] John E., Barbuto J.R., “A critique of the Myers-Briggs Type indicator and its operationalisation of Carl Jung’s Psychological type”, Psychological Reports, 1997, 80, 611–625.
[11] Komisin M., Guinn C., “Identifying personality types using document classification methods”, In Proce. of the twenty fifth International Florida AI Research Society Conference, Marco Island, 23–25 May 2012; pp. 232–237.
[12] Cui B., Qi C., “Survey Analysis of Machine Learning Methods for Natural Language Processing for MBTI Personality Type Prediction”.
[13] Friedman J. H., “Greedy function approximation: A gradient boosting machine”, Annals of Statistics, 2001, 1189–1232.
[14] Michie, D. Taylor C., Spiegelhalter D., “Machine Learning, Neural and Statistical Classification”, Ellis Horwood Limited: Hemel Hempstead, UK, 1994.
[15] Natekin, A.; Knoll, A.,“Gradient Boosting machines, a tutorial. Front”, Neurorobot. 2013, 7, 21.
[16] Mitchelle, J., “Myers-Briggs Personality Type Dataset”.
[17] Punnoose, R.; Ajit, P.,“Prediction of employee turnover in organizations using machine learning algorithms, A case for Extreme Gradient Boosting”, International Journal of Advanced Research in Artificial Intelligence(IJARAI) 2016, 5, 22–26.
[18] Prakruthi V, Sindhu D, Dr S Anupama Kumar, “Real Time Sentiment Analysis of Twitter Posts”, Third IEEE International Conference on Compu. Syst. and Info. Tech. for Sust. Solu., December 2018, IEEE.
[19] Vasanthakumar G U, Shashikumar D R, Suresh L, “Profiling Social Media Users, a Content-Based Data Mining Technique for Twitter Users”, First Int. Conf. on Adva. in Info. Tech. (ICAIT), 2019.
[20] Azhar Imran, Muhammad Faiyaz, Faheem Akhtar, “An Enhanced Approach for Quantitative Prediction of Personality in Facebook Posts”, Int. Jou. of Edu. and Mgt. Engineering (IJEME), Vol.8, No.2, pp.8-19, 2018.
[21] Golam Mostafa, Ikhtiar Ahmed, Masum Shah Junayed, “Investigation of Different Machine Learning Algorithms to Determine Human Sentiment Using Twitter Data”, International Journal of Info. Tech. and Comp. Sci. (IJITCS), Vol.13, No.2, pp.38-48, 2021.