Multimodal Technologies and Interaction

MDPI

Article

Machine Learning Approach to Personality Type Prediction Based on the Myers–Briggs Type Indicator®

Mohammad Hossein Amirhosseini * and Hassan Kazemian

School of Computing and Digital Media, London Metropolitan University, London N7 8DB, UK; h.kazemian@londonmet.ac.uk

* Correspondence: m.amirhosseini@londonmet.ac.uk

Received: 22 December 2019; Accepted: 12 March 2020; Published: 14 March 2020

Abstract: Neuro Linguistic Programming (NLP) is a collection of techniques for personality development. Meta programmes, which are habitual ways of inputting, sorting and filtering the information found in the world around us, are a vital factor in NLP. Differences in meta programmes result in significant differences in behaviour from one person to another. Personality types can be recognized through utilizing and analysing meta programmes. There are different methods to predict personality types based on meta programmes. The Myers–Briggs Type Indicator® (MBTI) is currently considered as one of the most popular and reliable methods. In this study, a new machine learning method has been developed for personality type prediction based on the MBTI. The performance of the new methodology presented in this study has been compared to other existing methods and the results show better accuracy and reliability. The results of this study can assist NLP practitioners and psychologists in regards to identification of personality types and associated cognitive processes.

Keywords: machine learning; personality type prediction; Myers–Briggs Type Indicator®; extreme Gradient Boosting

1. Introduction

Neuro Linguistic Programming (NLP) is a collection of techniques that can help to identify how people think, how they communicate and how they behave. In other words, NLP can be used to detect patterns in people's behaviour [1]. Meta programmes are a vital factor in NLP. Brian [2] explained that meta programmes are cognitive strategies that a person runs all the time, and they are different ways in which a person can sort information. Further, Davis [3] stated that meta programmes are habitual ways of inputting, sorting and filtering the information found in the world around us. In other words, they are our thinking styles or typical strategies and patterns. According to Ellerton [4], meta programmes can have a major influence on behaviours as well as how people communicate with others. As a result, this leads to significant differences in behaviour from one person to another.

In the early stage of NLP development, meta programmes emerged when Handler and Bandler collaborated together [5]. They discovered that people use different strategies for doing different things [6]. For instance, people use different strategies when making decisions. As a result, they presented the initial list of NLP meta programmes including 60 different patterns [4]. Many of these meta programmes have been combined together by subsequent researchers to form a much smaller and more useful set [4].

First, Cameron codified the initial list of meta programmes for therapeutic use [3]. Bailey and Stewart developed these for use in business [7], and Bailey created a profiling instrument named "LAB profile" which stands for the language and behavior profile [8]. Bailey also reduced the

number of patterns from 60 to 14 in order to make detecting and using these patterns simpler [8]. Following this, Woodsmall developed meta programmes for use in business and therapy and integrated them with the Myers–Briggs Personality Inventory [5]. The results were published in a book named "Time Line Therapy and The Basis of Personality" in 1988. He reduced the number of patterns again and made a smaller set of meta programmes, which includes only four basic and key meta programmes. These four basic meta programmes, also known as the Myers–Briggs Type Indicator® (MBTI), describe the preferences of an individual in four dimensions and these basic dimensions combine into one of 16 different personality types [9]. These four dimensions or basic meta programmes are Extroversion–Introversion (E–I), Sensation–Intuition (S–N), Thinking–Feeling (T–F), and Judgment–Perception (J–P).

Each dimension represents two types of personalities. Figure 1 shows a key of the eight personality types used in the Myers–Briggs Type Indicator®.

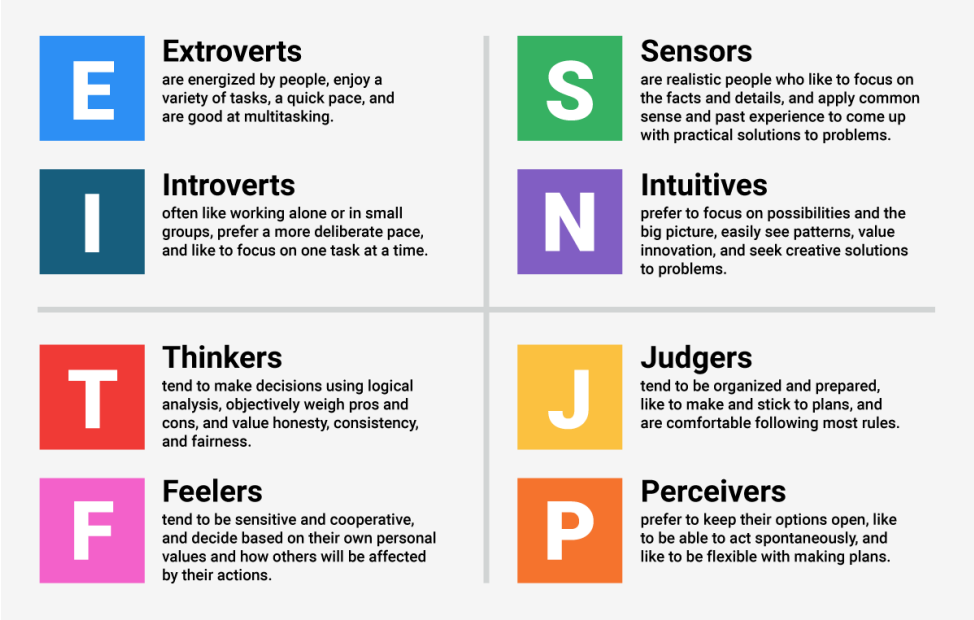

Figure 1. Personality types key [10].

1.1. Personality Types

Personality is derived from the Latin word persona, which means describing the behaviour or character of an individual [11]. It has been said that the meaning of personality is reflected in the very nature of the attitude of a person that can be distinguished from other people [12]. Personality, according to Hall and Lindzey [13], is "the dynamic organisation within the individual of those psychological systems that determine his characteristic behaviour and thought." This system determines the unique way in which an individual adapts to an environment. Personality is a description of the individual's self-image that influences their behaviour uniquely and dynamically, and this behaviour may change through the process of learning, experience, education, etc. This opinion clarifies Setiadi's view that personality is the dynamic organisation of the system that uniquely determines the individual's adjustment to the environment [14].

As discussed above, the preferences of an individual are categorised 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 2 shows these 16 personality types that result from the interactions among the preferences of an individual.

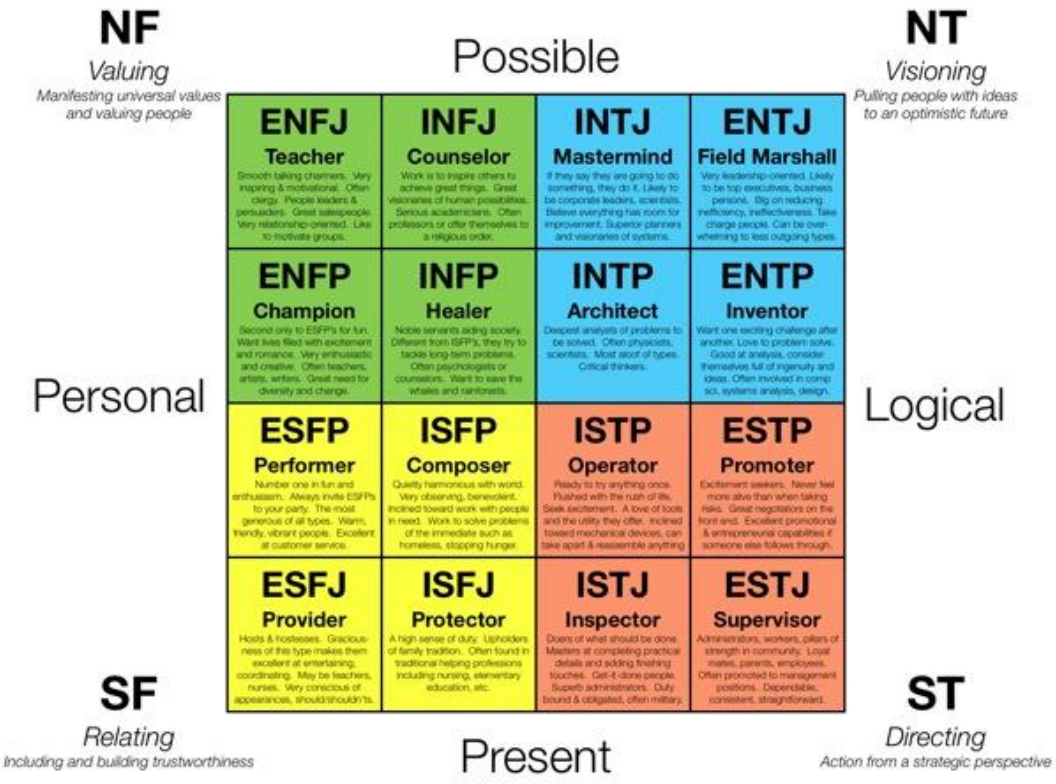

Figure 2. Personality types in the Myers–Briggs Type Indicator® [10].

As a result, the Myers–Briggs Type Indicator® (MBTI) has been used in this study in order to predict the personality type of individuals. The most popular meta programmes and personality types will also be identified, and the current organisational culture and task allocation can be modified based on this information. Each key word in Figure 2 represents a specific personality type and Figure 3 describes the cognitive functions of each MBTI personality type. The background colour of each type represents its dominant function and the colour of the text represents its auxiliary function.

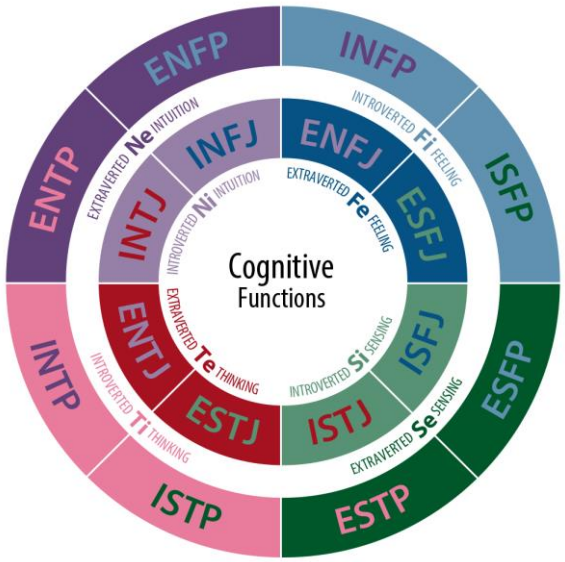

Figure 3. The cognitive functions of each personality type [15].

1.2. Background of Automating Personality Type Prediction

There is significant growing interest in automated personality prediction using social media among researchers in both the Natural Language Processing and Social Science fields [16]. So far, the application of traditional personality tests has mostly been limited to clinical psychology, counselling and human resource management. However, automated personality prediction from social media has a wider application, such as social media marketing or dating applications and websites [17].

Most studies on personality prediction have focused on the Big Five or MBTI personality models, which are the two most used personality models in the world. Soto [18] explained that "a personality trait is a characteristic pattern of thinking, feeling, or behaving that tends to be consistent over time and across relevant situations." Based on this explanation, The Big Five personality model can be defined as a set of five broad trait dimensions, namely, (1) extroversion, (2) agreeableness, (3) conscientiousness, (4) neuroticism and (5) openness [18]. In fact, the Big Five personality model uses descriptors of common language and suggests five broad dimensions commonly used to describe the human personality [19]. On the other hand, the Myers–Briggs Type Indicator® classifies personality types in 16 ways via four dimensions, namely, (1) introversion/extroversion, (2) sensing/intuition, (3) thinking/feeling and (4) judging/perceiving [20]. Research proposes that considering controversy about the reliability and validity of these two models, the MBTI model has more applications, especially in industry and for self-discovery of personality types [21].

Research on personality type prediction from textual data is scarce. However, important steps have been taken in this endeavour through machine learning. Classic machine learning techniques and neural networks have been used successfully for predicting MBTI personality types. One of the earliest studies on personality prediction using machine learning techniques was by Golbeck et al. [22]. They could accurately predict a user's personality type based on MBTI personality type indicator and by considering the information presented on their Twitter. In another study, Komisin and Guinn [23] used the Naïve Bayes and Support Vector Machine (SVM) techniques to predict an individual's personality type based on their word choice. Their database was built based on in-class writing samples that were taken from 40 graduate students along with their MBTI personality type. They compared the performance of these two techniques and discovered that the Naïve Bayes technique performs better than SVM on their small dataset. Two years later, Wan et al. [24] used a machine learning method to predict the Big Five personality type of users through their texts in Weibo, a Chinese social network, and they were able to successfully predict the personality type of the users. Li, Wan and Wang [25] used the grey prediction model, the multiple regression model and the multi-tasking model to predict the user personality type based on the Big Five model and their text samples. They compared the performance of these three models and found that the grey prediction model performs better than the two other models. In another study, Tandera et al. [26] used the Big Five personality model and some deep learning architecture to predict a person's personality based on the user's information on their Facebook. They compared the performance of their method with other previous studies that used classical machine learning methods and the results showed that their model successfully outperformed the accuracy of previous similar studies. Furthermore, in another study, Hernandez and Knight [27] used various types of recurrent neural networks (RNNs) such as simple RNN, gated recurrent unit (GRU) which is gating mechanism in recurrent neural networks, long short-term memory (LSTM) which is an artificial recurrent neural network architecture used in the field of deep learning, and Bidirectional LSTM to build a classifier capable of predicting people's MBTI personality type based on text samples from their social media posts. The Myers–Briggs Personality Type Dataset from Kaggle was used in their research. They compared the results and found that LSTM gave the best results. Recent research by Cui and Qi [28] used Baseline, Logistic Regression, Naïve Bayes, and SVM to predict an individual's MBTI personality type from one of their social media posts. They compared the results of all these methods and found that SVM performed better. They used the same database used in previous research, the Myers–Briggs Personality Type Dataset from Kaggle. Table 1 shows the studies and personality models used.

Table 1. Research on personality type prediction and personality models used.

Study	Personality Model	Method
Champa and Anandakumar (2010)	MBTI Network	Artificial Neural
Golbeck and et al. (2011)	MBTI Algorithms	Regression
Komisin and Guinn (2012)	MBTI Bayes and SVM	Naïve
Wan and et al. (2014)	Big Five Naive Bayes	Logistic Regression
Li, Wan and Wang (2017)	Big Five Learning	Multiple Regression and Multi-Task
Tandera and et al. (2017)	Big Five Architecture	Deep Learning
Hernandez and Knight (2017)	MBTI Networks	Recurrent Neural
Cui and Qi (2017)	MBTI Learning	Baseline, Naïve Bayes, SVM and Deep

In this study, it was found that classification techniques such as logistic regression, Naïve Bayes, Random forest, K Nearest Neighbour (KNN), Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM) have all been used for personality type prediction based on the MBTI or Big Five personality type models. According to the literature, the MBTI model has been more popular among researchers and, considering controversy about reliability and validity of these two models, the MBTI model has more applications in different disciplines [21]. That is why the MBTI personality model was used in this study. It was also noted that some powerful machine learning techniques such as Gradient Boosting were not implemented in this field. Gradient Boosting is a machine learning technique that has achieved considerable success in a wide range of practical applications because it is highly customisable to the particular needs of the application. Freund and Schapire [29] explained that Boosting is a method based on creating a very accurate prediction rule through combining rough inaccurate rules of thumb, which refer to principles with broad application that are not intended to be strictly accurate or reliable for every situation. Accordingly, Extreme Gradient Boosting, which is a boosted tree algorithm and follows the principle of Gradient Boosting [30], shows better performance due to the use of more regularised model formalisation in order to control over-fitting [31]. Due to the nature of personality prediction and required classification for MTI personality types and the way that Extreme Gradient Boosting can address classification tasks, we were convinced that this machine learning technique would be effective in this field. As a result, Extreme Gradient Boosting method was implemented in this study. The idea and theory behind this classification technique will be explained in the following section.

1.3. Extreme Gradient Boosting

Boosting is a method based on creating a very accurate prediction rule through combining rough inaccurate rules of thumb [29]. In this process, a sequence of weak learners is fitted onto modified data and predictions from all of them are combined through a weighted majority vote. This will help to produce the final prediction. Each step may contain some samples that were misclassified in the previous iteration. As a result, data modification is necessary at each step and includes assigning higher weights to the training samples that were misclassified. Samples that are difficult to predict during the iterations progress will receive increasing influence and this will force the weak learner to focus on the samples that are missed by its ancestor.

In Gradient Boosting, new models will be fitted during the learning process to provide a more accurate estimation of the response variable [32]. Extreme Gradient Boosting is a boosted tree algorithm that follows the principle of Gradient Boosting [30]. It is able to perform better due to using a more regularised model formalisation in order to control over-fitting [31]. In Gradient Boosting, the optimization problem is divided into two parts. In the first step, the direction of the step is

determined and then the step length is optimized. In Extreme Gradient Boosting, the step is determined directly by trying to solve Equation (1) for each x in the dataset.

$$\frac{\partial L\Big(y, f^{(m-1)}(x) + f_m(x)\Big)}{\partial f_m(x)} = 0 \tag{1}$$

Equation (2) is achieved by performing the second-order Taylor expansion of the loss function around the current estimate $f^{(m-1)}(x)$.

$$L\Big(y, f^{m-1}(x) + f_m(x)\Big)$$

$$\approx L\Big(y, f^{(m-1)}(x)\Big) + g_m(x) f_m(x) + \frac{1}{2} h_m(x) f_m(x)^2 \tag{2}$$

$g_m(x)$ in Equation (2) is the gradient and $h_m(x)$ is the second-order derivative, which is explained in Equation (3).

$$h_m(x) = \frac{\partial^2 L(Y, f(x))}{\partial f(x)^2} \tag{3}$$

Note that $f(x) = f^{(m-1)}(x)$. Equation (4) is the loss function.

$$L(f_m) \approx \sum_{i=1}^{n} \Big[g_m(x_i) f_m(x_i) + \frac{1}{2} h_m(x_i) f_m(x_i)^2 \Big] + const$$

$$\propto \sum_{j=1}^{T_m} \sum_{i \in R_{jm}} \Big[g_m(x_i) w_{jm} + \frac{1}{2} h_m(x_i) w_{jm}^2 \Big] \tag{4}$$

g_m in Equation (4) represents the sum of the gradient in region j and h_m is the sum of the second-order derivative in region j. This equation is used to determine Equation (5).

$$L(f_m) \propto \sum_{j=1}^{T_m} \Big[G_{jm} w_{jm} + \frac{1}{2} H_{jm} w_{jm}^2 \Big] \tag{5}$$

With the fixed learned structure, it is straight forward to determine the optimal weight for each region, as shown in Equation (6).

$$w_{jm} = -\frac{G_{jm}}{H_{jm}},\ j = 1, \ldots, T_m \tag{6}$$

Equation (7) will be achieved by plugging Equation (6) back to the loss function.

$$L(f_m) \propto -\frac{1}{2} \sum_{j=1}^{T_m} \frac{G_{jm}^2}{H_{jm}} \tag{7}$$

Equation (7) is the structure score for a tree. The smaller the score is, the better the structure is. In order to make each split, the proxy gain is defined as Equation (8).

$$Gain = \frac{1}{2} \left[\frac{G_{jmL}^2}{H_{jmL}} + \frac{G_{jmR}^2}{H_{jmR}} - \frac{G_{jm}^2}{H_{jm}} \right]$$

$$= \frac{1}{2} \left[\frac{G_{jmL}^2}{H_{jmL}} + \frac{G_{jmR}^2}{H_{jmR}} - \frac{\Big(G_{jmL} + G_{jmR}\Big)^2}{H_{jmL} + H_{jmR}} \right] \tag{8}$$

It can be seen in Equation (8) that all deductions do not take regularization into consideration. In order to improve generalization performance, Extreme Gradient Boosting provides variety of regularization. As a result, the loss function can be rewritten as Equation (9).

$$L(f_m) \propto \sum_{j=1}^{T_m}\left[G_{jm}w_{jm} + \frac{1}{2}H_{jm}w_{jm}^2\right] + \gamma T_m + \frac{1}{2}\lambda\sum_{j=1}^{T_m} w_{jm}^2 + \alpha\sum_{j=1}^{T_m}\left|w_{jm}\right|$$

$$= \sum_{j=1}^{T_m}\left[G_{jm}w_{jm} + \frac{1}{2}(H_{jm}+\lambda)w_{jm}^2 + \alpha\left|w_{jm}\right|\right] + \gamma T_m \tag{9}$$

γ in Equation (9) is the penalization term on the number of terminal nodes. For $L1$ and $L2$ regularization, α and λ will be used respectively. The optimal weight for each region is calculated in Equation (10).

$$w_{jm} = \begin{cases} -\frac{G_{jm}+\alpha}{H_{jm}+\lambda} & G_{jm} < -\alpha \\ -\frac{G_{jm}-\alpha}{H_{jm}+\lambda} & G_{jm} > \alpha \\ 0 & else \end{cases} \tag{10}$$

The gain of each split will be defined accordingly in Equation (11).

$$Gain = \frac{1}{2}\left[\frac{T_\alpha\left(G_{jmL}^2\right)}{H_{jmL}+\lambda} + \frac{T_\alpha\left(G_{jmR}\right)^2}{H_{jmL}+\lambda} - \frac{T_\alpha\left(G_{jm}\right)^2}{H_{jm}+\lambda}\right] - \gamma \tag{11}$$

2. Methodology for Automating Personality Type Prediction

2.1. Development Tools

The natural language processing toolkit (NLTK) and XGBoost which is an optimised distributed Gradient Boosting library in Python, were used for the development process. NLTK is a powerful natural language processing toolkit for developing Python programmes to work with human language data. Moreover, XGBoost can be used to implement machine learning algorithms under the Gradient Boosting framework. Pandas, numpy, re, seaborn, matplotlib and sklearn are other Python libraries that were used.

2.2. Dataset for Training the Model

The publicly available Myers–Briggs personality type dataset from Kaggle, containing 8675 rows of data, was used in this research. In this dataset, each row consists of two columns. The first column is for the MBTI personality type of a given person, and the second column includes fifty posts obtained from the individual's social media. Each post has been separated by three pipe characters [33]. This data has been collected from the users of an online forum, where in the first step, users take a questionnaire that recognises their MBTI type; and in the second step, communicate with other users [27].

2.3. Proportionality in Dataset

In this step, seaborn which is a Python data visualisation library and matplotlib which is a Python 2D plotting library were used for data preview and to determine the distribution of the MBTI personality types in the dataset. Figure 4 shows the number of occurrences for each MBTI personality type in the dataset.

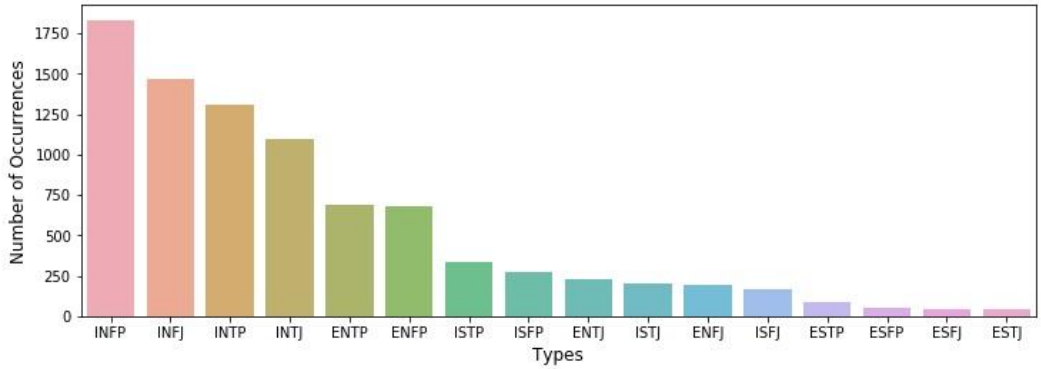

Figure 4. Number of occurrences for each MBTI personality type in the dataset.

Similarly, Figure 5 shows the percentage of occurrences for each MBTI personality type in the dataset.

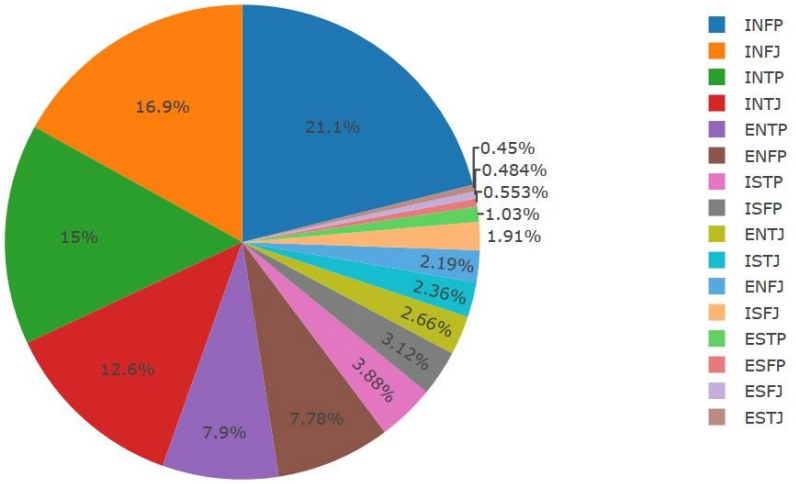

Figure 5. Percentage of occurrence for each MBTI personality type in the dataset.

Figures 4 and 5 show a non-uniform representation of MBTI types in the dataset that is not commensurate with the actual proportions of MBTI types in the general population shown in Table 2. As a result, it was clear that some cleaning in the dataset would be necessary in order to improve the accuracy of the proportional representation of each MBTI type. Pre-processing the dataset will be explained in Section 2.5.

Table 2. Personality type distribution in the general population [34].

Personality Type	Frequency in Population
ISFJ	13.8%
ESFJ	12.3%
ISTJ	11.6%
ISFP	8.8%
ESTJ	8.7%
ESFP	8.5%
ENFP	8.1%
ISTP	5.4%

Table 2. *Cont.*

Personality Type	Frequency in Population
INFP	4.4%
ESTP	4.3%
INTP	3.3%
ENTP	3.2%
ENFJ	2.5%
INTJ	2.1%
ENTJ	1.8%
INFJ	1.5%

2.4. Categorizing the Type Indicators in Four Dimensions

Four different categories were created for the type indicators in order to understand the distribution of types indicators in the dataset. The first category was for Introversion (I)/Extroversion (E), the second category was for Intuition (N)/Sensing (S), the third was for Thinking (T)/Feeling (F) and the fourth category was for Judging (J)/Perceiving (P). As a result, for each category, one letter will return and at the end there will be four letters that represent one of the 16 personality types in the MBTI. For instance, if the first category is returning I, the second category is returning N, the third category is returning T and the fourth category is returning J, the relevant personality type would be INTJ. Table 3 and Figure 6 show the distribution across type indicators.

Table 3. Distribution across 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

According to Table 3 and Figure 6, for the first category of Introversion (I)/Extroversion (E), the distribution of Extroversion (E) is much greater than Introversion (I). Similarly, for the second category which is Intuition (N)/Sensing (S), the distribution of Sensing (S) is much higher than Intuition (N). Figure 6 and Table 3 also show that for the third category which is Thinking (T)/Feeling (F), the distribution of Thinking (T) is slightly more than Feeling (F). Finally, for the fourth category which is Judging (J)/ Perceiving (P), the distribution of Judging (J) is greater than Perceiving (P).

The Pearson correlation coefficient can measure the strength between variables and relationships. Each random variable (X_i) in a correlation matrix is correlated with each of the other values in the table (X_j) and this can help to understand which pairs have the highest correlation. In order to understand how significant the relationship is between two variables, the coefficient value must be found, which can range between −1.00 and 1.00. Figure 7 shows the correlation efficient between personality type identifiers.

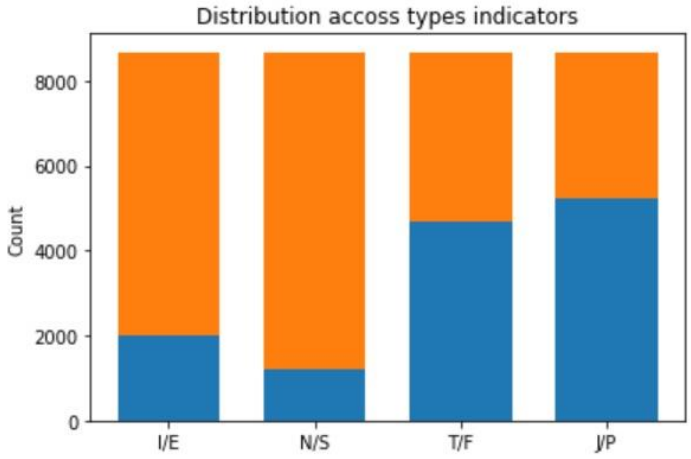

Figure 6. Distribution across type indicators.

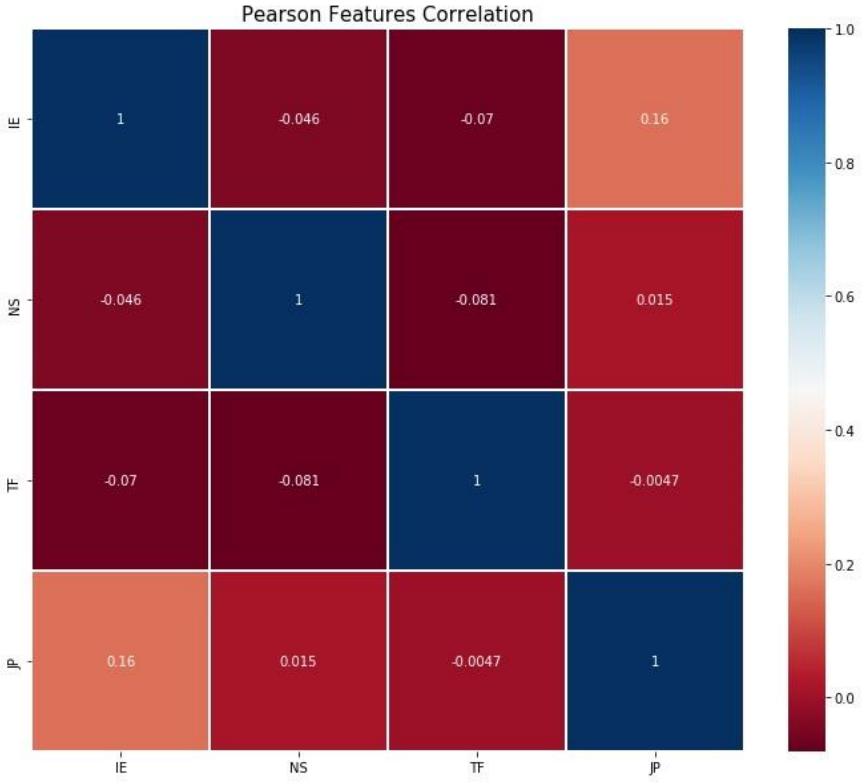

Figure 7. Pearson correlation coefficient between personality type indicators.

2.5. Pre-Processing the Dataset

As discussed earlier, data in this dataset was collected from an Internet forum and after analysing the content of the dataset, it was clear that some word removal is necessary. Non-uniform representation of MBTI types in the dataset that is not commensurate with the actual proportions of MBTI types in the general population was the most important reason for this. It was determined that this is because the data was collected from an Internet forum created for discussion about personality type and MBTI types were repeated too many times in the posts. This may also affect the accuracy of the model. As a result, NLTK was used to remove the MBTI types from the dataset. After this step, the distribution of MBTI personality types in the dataset was determined again, now finding that representation of MBTI types in the dataset is commensurate with the actual proportions of MBTI types in the general

population. In addition, all urls and stop words were removed from the dataset. Finally, in order to make the dataset more meaningful, the text was lemmatised, i.e., inflected forms of the words were transformed into their root words.

2.6. Vectorise with Count and Term Frequency–Inverse Document Frequency (TF–IDF)

Sklearn library was used to recognize the words appearing in 10% to 70% of the posts. In the first step, posts were placed into a matrix of token counts. In the next step, the model learns the vocabulary dictionary and returns a term-document matrix. The count matrix then transforms into a normalised TF–IDF representation which can be used for the Gradient Boosting model. Finally, 791 words appear in 10% to 70% of the posts.

2.7. Classification Task

In machine learning, there are two different types of classification. In the first type, based on a set of observations, the aim is to establish the existence of classes or clusters in the data. In the second type, a certain number of classes may exist, and the aim is to establish a rule or a set of rules to classify a new observation into one of the existing classes [31]. The first type is known as Unsupervised Learning and the second type as Supervised Learning [35].

The classification task was divided into 16 classes and further into four binary classification tasks, since each MBTI type is made of four binary classes. Each one of these binary classes represents an aspect of personality according to the MBTI personality model. As a result, four different binary classifiers were trained, whereby each one specializes in one of the aspects of personality. Thus, in this step, a model for each type indicator was built individually. Term Frequency–Inverse Document Frequency (TF–IDF) was performed and MBTI type indicators were binarised. Variable X was used for posts in TF–IDF representation and variable Y was used for the binarised MBTI type indicator.

2.8. Developing Gradient Boosting Model for the Dataset

Numpy, XGBoost and sklearn were used in this step to create the Gradient Boosting Model. MBTI type indicators were trained individually, and the data was then split into training and testing datasets using the train_test_split() function from sklearn library. In total, 70% of the data was used as the training set and 30% of the data was used as the test set. The model was fit onto the training data and the predictions were made for the testing data. After this step, the performance of the XGBoost model on the testing dataset during training was evaluated and early stopping was monitored. The result of this evaluation is presented in Section 3. Following this step, the learning rate in XGBoost should be set to 0.1 or lower, and the addition of more trees will be required for smaller values. Moreover, the depth of trees should be configured in the range of 2 to 8, as there is not much benefit seen with the deeper trees. Furthermore, row sampling should be configured in the range of 30% to 80% of the training dataset. Thus, tree_depth in the created XGBoost was configured and parameters for XGBoost were setup as follow:

n_estimators = 200
max_depth = 2
nthread = 8
learning_rate = 0.2

MBTI type indicators were trained individually and then the data was split into training and testing datasets. The model was fit onto the training data and the predictions were made for the testing data. In this step, the performance of the XGBoost model on the testing dataset was evaluated again and the result is presented in Section 3.

The scikit-learn library enables searching combinations of parameters and this capability was used in order to discover the optimal way to configure the model for achieving top performance. This is called Hyperparameter tuning in the XGBoost model. As a result, parameters including (1) the number

and size of trees, (2) the learning rate and number of trees, and (3) the row and column subsampling rates are the parameters to consider when tuning.

3. Results and Discussion

Evaluating the Accuracy of the XGBoost Model

As explained in Section 2.8, after creating the Gradient Boosting Model, MBTI type indicators were trained individually, and the data was split into training and testing datasets. The model was fit onto the training data and predictions were made for the testing data. Then predictions were evaluated.

After configuring the tree_depth in the created XGBoost model, predictions were evaluated again.

According to Table 4, after configuration, the performance of the model and the accuracy was slightly improved in the Introversion (I)–Extroversion (E) category and considerably improved in the Feeling (F)–Thinking (T) category. The accuracy in Intuition (I)–Sensing (S) and Judging (J)–Perceiving (P) categories, however, was slightly worse.

Table 4. Comparison of accuracy prediction before and after configuration.

Binary Class	MBTI Personality Type	Accuracy after Configuration	Accuracy before Configuration	Difference
IE 0.84	Introversion (I)–Extroversion (E)	79.01%	78.17%	
NS 0.1	Intuition (I)–Sensing (S)	85.96%	86.06%	-
FT 2.41	Feeling (F)–Thinking (T)	74.19%	71.78%	
JP 0.28	Judging (J)–Perceiving (P)	65.42%	65.70%	-

Other existing methods which used the same dataset were discussed in Section 1.3. As a result, the accuracy of prediction after configuration was compared to the latest and most successful existing method. This method was introduced by Hernandez and Knight [27] in 2017. The same dataset was used in their research and the pre-processing step was exactly the same as this research. Hence, the comparison between their method and the presented method in this research was on the same data. They used various types of recurrent neural networks (RNNs) such as simple RNN, GRU, LSTM, and Bidirectional LSTM to build their classifier. For evaluation, they used two different methods, which were (1) a post classification methodology and (2) a user classification methodology. For post classification, they pre-processed the test set and predicted the class for every individual post. They then produced an accuracy score and confusion matrix for every MBTI dimension. On the other hand, in order to classify users, they needed to find a way of tuning the class predictions of individual posts all authored by an individual into a prediction for the class of the author. As a result, they took the mean of the class probability predictions for all of the posts in a user's corpus and rounded either to 0 or 1.

The accuracy of the recurrent neural network model using user the classification methodology was better than the recurrent neural network model using the post classification methodology. Thus, the accuracy of Extreme Gradient Boosting was compared to their recurrent neural network classifier using the user classification methodology. In fact, the same strategy for evaluating the results was used in this research, as we wanted to compare the performance of the models in a same way. Table 5 shows the results of this comparison.

Table 5 shows that the Extreme Gradient Boosting classifier in three dimensions of MBTI personality types has a greater degree of accuracy than the recurrent neural network. Regarding the Intuition (I)–Sensing (S) and Introversion (I)–Extroversion (E) categories, the accuracy of the Extreme Gradient Boosting is significantly greater than the recurrent neural network; and for the Judging (J)–Perceiving (P) category, the accuracy is slightly better. However, the accuracy of the recurrent neural network for Feeling (F)–Thinking (T) is considerably better than the Extreme Gradient Boosting classifier. Thus,

the overall performance of the Extreme Gradient Boosting classifier is better than the recurrent neural network for this dataset.

Table 5. Comparison of accuracy of the Extreme Gradient Boosting model and the recurrent neural network model.

Binary Class	MBTI Personality Type	Accuracy of Extreme Gradient Boosting	Accuracy of Recurrent Neural Network	Difference
IE 10.75%	Introversion (I)–Extroversion (E)	78.17%	67.6%	
NS 24.06%	Intuition (I)–Sensing (S)	86.06%	62%	
FT 6.02%	Feeling (F)–Thinking (T)	71.78%	77.8%	
JP 2.0%	Judging (J)–Perceiving (P)	65.70%	63.70%	-

4. Conclusions

This research has developed a new machine learning method for automating the process of meta programme detection and personality type prediction based on MBTI personality type indicator. The natural language processing toolkit (NLTK) and XGBoost, which is an optimized distributed Gradient Boosting library in Python for implementing machine learning algorithms under the Gradient Boosting framework, were used for development process. Moreover, Pandas, Numpy, re, Seaborn, Matplotlib and Sklearn were other Python libraries that were used. The accuracy of the XGBoost model was evaluated and the performance was compared to the latest and most successful existing method which used the same dataset. The results show that the methodology presented in this research has better accuracy and reliability in comparison to other existing methods. Regarding the knowledge contribution in this paper, the presented methodology significantly improved the accuracy of recognising the Intuition (I)–Sensing (S) and Introversion (I)–Extroversion (E) personality categories, as well as slightly better accuracy for the Judging (J)–Perceiving (P) personality category. This can effectively assist NLP practitioners and psychologists in regards to identification of personality types and associated cognitive processes.

Author Contributions: Conceptualization, M.H.A.; methodology, M.H.A.; software, M.H.A.; validation, M.H.A.; formal analysis, M.H.A.; data curation, M.H.A.; writing—original draft preparation, M.H.A.; writing—review and editing, M.H.A. and H.K.; visualization, M.H.A.; supervision, H.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Stevenson, M. Introduction to Neuro-Linguistic Programming. Available online: http://www.freenlphomestudy.com/membersonly/iNLP/iNLPManual.pdf (accessed on 8 September 2018).
2. Brian, C. Metaprograms as a Tool for Critical Thinking in Reading and Writing. In Proceedings of the Second JALT Critical Thinking SIG Forum, Kobe Convention Center, Portopia Kobe, Tokyo, 25–28 October 2013; Available online: http://www.standinginspirit.com/wp-content/uploads/2013/10/JALT2013-Critical-Thinking-Forum-Handout-Metaprograms-Explanation-Handout.pdf (accessed on 12 September 2018).
3. Davis, K. The Meta Model Problem Solving Strategies. Available online: http://nlp-mentor.com/meta-model/ (accessed on 16 September 2018).
4. Ellerton, R. NLP Meta Programs, Part 1. Available online: http://www.renewal.ca/nlp17.htm (accessed on 24 September 2018).
5. Hall, L.M.; Bodenhomer, B.G. *Figuring Out People, Design Engineering with Meta-Programs*; Crown House Publishing Ltd.: Carmarthe, UK. Available online: http://www.nlpinfocentre.com/nlpebooks/(Ebook%20-%20Nlp)%20Michael%20Hall%20-%20Figuring%20People%20Out.pdf (accessed on 28 September 2018).

6. Hoag, J.D. The NLP Meta Model. Available online: http://www.nlpls.com/articles/NLPmetaModel.php (accessed on 27 August 2018).
7. James, T.; Woodsmall, W. *Time Line Therapy and the Basis of Personality*; Meta Publications: Toronto, ON, Canada. Available online: http://simbi.kemenag.go.id/pustaka/images/materibuku/time-line-therapy-and-the-basis-of-personality.pdf (accessed on 17 September 2018).
8. Charvet, S.R. *Words that Change Minds: Mastering the Language of Influence*, 2nd Revised ed.; Kendall/Hunt Publishing Co.: Dubuque, IA, USA, 1997.
9. Mind Academy. Basic Meta Programs. Available online: http://www.mindacademy.com/nlp/basic-metaprograms (accessed on 25 August 2018).
10. Tieger, P.D.; Barron-Tieger, B. *Do What You Are: Discover the Perfect Career for You through the Secrets of Personality Type*, 4th ed.; Sphere: London, UK, 2007.
11. Darsana, M. The influence of personality and organisational culture on employee performance through organisational citizenship behaviour. *Int. J. Manag.* **2013**, *2*, 35–42.
12. Alwi, H.; Sugono, D.; Adiwirmata, S. *Kamus Besar Bahasa Indonesia*; Balai Pustaka: Jakarta, Indonesia, 2003.
13. Hall, C.; Lindzey, G. *Theories of Personality*, 2nd ed.; Wiley: New York, NY, USA, 1970.
14. Setiadi, N.J. *Perilaku Konsumen Konsep Dan Implikasi Untuk Strategi Dan Penelitian Pemasaran*; Prenada Media: Jakarta, Indonesia, 2003.
15. Beech, J. The Cognitive Functions of each Personality Type. Available online: https://siteassets.pagecloud.com/greeleymosaic/downloads/Myers-Briggs-ID-7fbebb3b-f94d-468a-ce4f-7c488703c102.pdf (accessed on 19 September 2018).
16. Nguyen, D.; Doğruöz, A.S.; Rosé, C.P.; Jong, F.D. Computational sociolinguistics: A survey. *Comput. Linguist.* **2016**, *42*, 537–593. [CrossRef]
17. Gjurkovic, M.; Snajder, J. Reddit: A gold mine for personality prediction. In Proceedings of the Second Workshop on Computational Modelling of People's Opinions, Personality and Emotions in Social Media, New Orleans, LA, USA, 6 June 2018; pp. 87–97. Available online: https://peopleswksh.github.io/pdf/PEOPLES12.pdf (accessed on 21 September 2018).
18. Soto, C.J. Big Five personality traits. In *The SAGE Encyclopedia of Lifespan Human Development*; Borstein, M.H., Arterberry, M.E., Fingerman, K.L., Lansford, J.E., Eds.; SAGE Publications: Thousand Oaks, CA, USA, 2018; pp. 240–241.
19. Goldberg, L.R. The structure of phenotypic personality traits. *Am. Psychol.* **1993**, *48*, 26–34. [CrossRef] [PubMed]
20. Myers, I.B.; McCaulley, M. *Manual: A Guide to the Development and Use of the Myers-Briggs Type Indicator*, 15th ed.; Consulting Psychologists Press: Santa Clara, CA, USA, 1989.
21. John, E.; Barbuto, J.R. A critique of the Myers-Briggs Type indicator and its operationalisation of Carl Jung's Psychological types. *Psychol. Rep.* **1997**, *80*, 611–625.
22. Golbeck, J.; Robles, C.; Edmondson, M.; Turner, K. Predicting personality from Twitter. In Proceedings of the IEEE Third International Conference on Privacy, Security, Risk and Trust and IEEE Third International Conference on Social Computing, Boston, MA, USA, 9–11 October 2011. Available online: https://ieeexplore.ieee.org/document/6113107/ (accessed on 25 August 2018).
23. Komisin, M.; Guinn, C. Identifying personality types using document classification methods. In Proceedings of the 25th International Florida Artificial Intelligence Research Society Conference, Marco Island, FL, USA, 23–25 May 2012; pp. 232–237.
24. Wan, D.; Zhang, C.; Wu, M.; An, Z. Personality prediction based on all characters of user social media information. In Proceedings of the Chinese National Conference on Social Media Processing, Beijing, China, 1–2 November 2014; pp. 220–230.
25. Li, C.; Wan, J.; Wang, B. Personality Prediction of Social Network Users. In Proceedings of the 16th International Symposium on Distributed Computing and Applications to Business, Engineering and Science, Anyang, China, 13–16 October 2017. Available online: https://ieeexplore.ieee.org/document/8253041/?part=1 (accessed on 16 August 2018).
26. Tandera, T.; Suhartono, D.; Wongso, R.; Prasetio, Y. Personality prediction system from Facebook users. In Proceedings of the 2nd International Conference on Computer Science and Computational Intelligence, Bali, Indonesia, 13–14 October 2017.

27. Hernandez, R.; Knight, I.S. Predicting Myers-Bridge Type Indicator with text classification. In Proceedings of the 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017. Available online: https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1184/reports/6839354.pdf (accessed on 9 September 2018).
28. Cui, B.; Qi, C. Survey Analysis of Machine Learning Methods for Natural Language Processing for MBTI Personality Type Prediction. Available online: http://cs229.stanford.edu/proj2017/final-reports/5242471.pdf (accessed on 3 September 2018).
29. Freund, Y.; Schapire, R.E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [CrossRef]
30. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, 1189–1232. [CrossRef]
31. Punnoose, R.; Ajit, P. Prediction of employee turnover in organisations using machine learning algorithms, A case for Extreme Gradient Boosting. *Int. J. Adv. Res. Artif. Intell.* **2016**, *5*, 22–26. [CrossRef]
32. Natekin, A.; Knoll, A. Gradient Boosting machines, a tutorial. *Front. Neurorobot.* **2013**, *7*, 21. [CrossRef] [PubMed]
33. Mitchelle, J.; Myers-Briggs Personality Type Dataset. Includes a Large Number of People's MBTI Type and Content Written by Them. Available online: https://www.kaggle.com/datasnaek/mbti-type (accessed on 4 September 2018).
34. Myers, I.B.; McCaulley, M.H.; Quenk, N.L.; Hammer, A.L. *MBTI Manual: A Guide to the Development and Use of the Myers-Briggs Type Indicator*; CPP: Palo Alto, CA, USA, 1998.
35. Michie, D.; Spiegelhalter, D.; Taylor, C. *Machine Learning, Neural and Statistical Classification*; Ellis Horwood Limited: Hemel Hempstead, UK, 1994; Available online: https://www1.maths.leeds.ac.uk/~{}charles/statlog/whole.pdf (accessed on 27 September 2018).