Predicting MBTI Personality type with K-means Clustering and Gradient Boosting

Conference Paper · November 2020

DOI: 10.1109/INMIC50486.2020.9318078

Nosheen Sabahat
Forman Christian College

Predicting MBTI Personality type with K-means Clustering and Gradient Boosting

Zeeshan Mushtaq
Department of Computer Science
Forman Christian College University,
Lahore, Pakistan
213542608@formanite.fccollege.edu.pk

Sagar Ashraf
Department of Computer Science
Forman Christian College University,
Lahore, Pakistan
213542610@formanite.fccollege.edu.pk

Nosheen Sabahat
Department of Computer Science
Forman Christian College University
Lahore, Pakistan
nosheensabahat@fccollege.edu.pk

***Abstract*— Personality refers to a characteristic pattern of thoughts, behavior, and feelings that makes a person unique. Asking users to fill a questionnaire to get their personality insights could be inaccurate because the users are conscious and try to take a careful approach when filling the survey. However, when it comes to social media, users do not take any consideration before posting their opinions on social media. Therefore, the data obtained from social media could be precious to determine the user personality type. In this paper, we propose a way to analyze the user's data posted on social media by combining two existing machine learning algorithms, such as K-Means Clustering and Gradient Boosting, in order to predict user personality type. Moreover, this research helps to analyze the empirical relation between the user's data posted on social media and the user's personality. In this paper, we used The Myer-Briggs Type Indicator (MBTI) introduced by Swiss psychiatrist Carl Jung. MBTI is based on sixteen personality types, and they act as a valuable reference point to understand a person's unique personality. The technique of combining these two machine learning algorithms gave accurate results than the traditional naive Bayes classification and other algorithms. Results of this study can help bloggers and social media users to know what type of personality they are showing on the social media with the data they posted on the internet.**

Keywords—TF-IDF(Term Frequency-Inverse Document Frequency), k-means, Xgboost, Gradient Boosting, Stop words, Lemmatization, Scatter Plot, MBTI, k-Folds, hyper parameters

I. INTRODUCTION

The current era of the internet is witnessing a huge growth of electronic media such as social websites, and a massive amount of new data is being created every minute [1]. With the growing popularity of social media, it has become possible to disseminate this information at a rapid rate. Millions of posts are published every day on social networking sites like Facebook, Twitter, Instagram, and many others. Among these social networking sites, Facebook is the only fantastic real-time social networking tool that can be a great source of rich information for data mining [2].

In early 2018, a whistleblower revealed a piece of astonishing information regarding the presidential election campaign in 2016 [2]. The information involved an organization called Cambridge Analytica that harvested Facebook Profiles of 50 million users in the United States and performed a detailed analysis of their data to determine the success rates during the elections. Although it was a direct security breach by the organization on the user's data, we can conclude that users' data on social media (Facebook) can help determine the outcome of a particular business objective.

Data mining is a process of going through small or large datasets to identify hidden patterns and structures to solve different problems through data analysis and machine learning algorithms [3]. It is a multidisciplinary field that uses statistics, Artificial intelligence, databases, and machine learning to find the insights of the dataset. In other words, data mining is knowledge discovery, pattern or data analysis, information harvesting, and others. Insights revealed by data mining techniques can be used in different fields, such as market analysis, biogenetics [3], etc. Data mining tools and techniques help various organizations to predict future trends by analyzing the past or current data. In data mining, different association rules are created by analyzing the data and extracting useful information out of it. To get the valuable insights from a data, many data mining techniques are applied.

II. BACKGROUND AND RELATED WORK

There are many data mining techniques such as classification, clustering, sequence analysis just to name a few [4]. Data mining techniques are used in many research areas, including mathematics, biogenetics, and marketing. Different data mining techniques are being used in the business field to predict customer behavior and future trends of the business market.

The Myers-Briggs Personality Indicator is a very famous personality revealing way designed to identify a person's personality type, strength, and preferences [5]. Bharadwaj et.al [6] are of the view that users data such as essays, posts, statuses, and blogs have been analyzed to find out the user's personality by using different machine learning algorithms such as Naive Bayes, SVM and Neural net yelling an accuracy of 88%[6].

Gradient boosting is a machine learning algorithm mainly used for regression and classification problems [7]. Brownlee explains the comparison between different machine learning algorithms such as SVM, Naive Bayes, and Gradient Boosting has been done using **Google's** Bangle dataset. All of these Machine learning algorithms have shown promising results with reasonable accuracy, but Gradient boosting has shown the highest accuracy of 76.95 [8].

The previous study on personality prediction has been done by using social media Twitter. The research conducted at Cornell University analyzed the tweets of users and predicted the dark personality traits that the user might inhibit [9]. Our model predicts the MBTI personality type by combining two famous machine learning algorithms, K-Means clustering, and Gradient boosting. For example, our model would analyze the text posted by the user and predict

2020 IEEE 23rd International Multitopic Conference (INMIC) | 978-1-7281-9893-4/20/$31.00 ©2020 IEEE | DOI: 10.1109/INMIC50486.2020.9318078

978-1-7281-9893-4/20/$31.00 ©2020 IEEE

out of 16 MBTI personality types which personality type this particular user has.

Golbeck et.al [10] applied various deep learning algorithms, including Long Term Short Memory (LSTM), Convolutional Neural Network, and MLP with an accuracy rate of 68.63%, 73.4%, and 73.87. Later they combined LSTM and CNN to improve their accuracy rate for up to 74% as opposed to traditional machine learning algorithms (Naive Bayes, LDA), which showed an accuracy rate of 64% and 66.8%, respectively [10].

Zvarevashe et.al applied two machine learning algorithms such as Random Forest Recursive Feature Elimination (RF-RFE) and Gradient Boosting for gender voice recognition. The public voice dataset was analyzed. Different algorithms such as Feed Forward Neural Network (FFNN), Extreme Learning Machine (ELM), and Gradient boosting for gender voice recognition. Among these Gradient, surprisingly gave the highest Accuracy than FFNN and ELM when all these Algorithms were used with Random Forest Recursive Feature Elimination (RF-RFE) [11].

Literature reveals that personality prediction using various regression and classification models with an average accuracy of 74 - 82% has been implemented usually. The gradient boosting algorithm is a robust machine learning algorithm that has shown very considerable outcomes on many practical applications [7]. The gradient boosting algorithm is highly tunable to the particular need of the model for prediction. On the other hand, K-means clustering is one of the most popular algorithms for clustering tasks to understand the structure of a dataset.

We used the traditional TF-IDF approach, along with the K-means clustering algorithm for making clusters and Gradient Boosting to predict the MBTI personality type.

III. METHODOLOGY

This research focuses on predicting MYERS BRIGGS personality type as it is one of the most common personality types used worldwide. The foundations of MBTI type indicators were laid by Katharine Cook Briggs and Isabel Briggs Myers in 1917. The idea behind creating these type indicators was to help women select right job for them during world war II [12].

Fig. 1 shows our proposed approach.

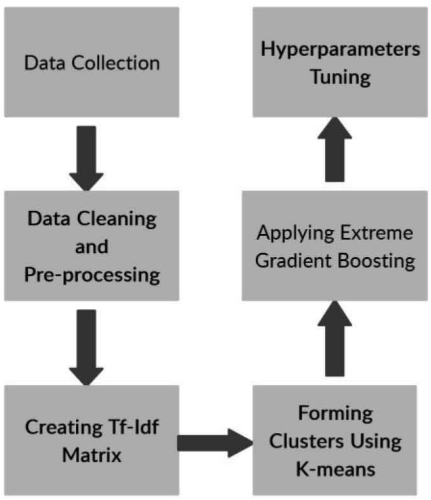

Fig. 1. Proposed Methodolgy

A. Data Collection and Visualization

The dataset we are using is an MBTI data set obtained from Kaggle. It has two columns and 8675 rows. One column indicates the personality type, i.e., ISTJ, ISTP, ISFJ, ISFP, INFJ, INFP, INTJ, INTP, ESTP, ESTJ, ESFP, ESFJ, ENFP, ENFJ, ENTP and ENTJ , while the other column contains the raw script written by that specific personality type. If you see in the figure below, the posts column contains several entries that we do not need, such as URLs. However, before we apply some preprocessing on our dataset, we visualized the data to find out patterns so that we could identify the potential setbacks we have in our data.

Fig. 2 shows our transformed data set as a bar graph where the number of posts is in the y-axis and type indicators on the x-axis. It can be observed that the data is unbalanced. For example, the INFP type indicator contains over 1750 posts, while ESTJ contains less than 250.

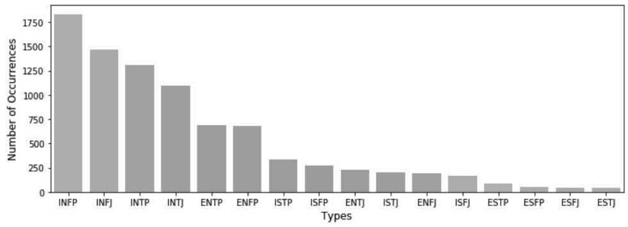

Fig. 2. Transformed Dataset

Later, the primary dataset is transformed to individual type indicators. If you have given the personality test, the results show you are a combination of 4 different type indicators, such as Introvert/Extrovert, Intuition/Sensing/, Thinking/Feeling, and Judging/Perceiving.

We created four new columns of each type indicator and assigned binary values so that it would be easy for us to count the number of individual posts each indicator has to see how the data is distributed among each indicator type. Fig. 3 shows the same.

Out[5]:

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

Fig. 3. Binarized Dataset

B. Data Cleaning and Preprocessing

Data preprocessing is a process of cleaning the data to extract useful information that is helpful for the machine learning model. We did the three necessary preprocessing steps:

- Removing URLs and MBTI Profile strings
- Converting into lowercase
- Lemmatization

Removing URLs and MBTI profile strings from the data was vital because they did not contribute to the efficiency of our algorithm. After that, every string was converted into lowercase to avoid any string differences. After this step, lemmatization was applied; it is a technique of converting every word to its base form so that it is categorized as a single term [13]. For example, 'caring' to appear in our data set would be treated as 'care.'

In natural language processing, words that do not contribute to the learning of the machine learning algorithm are called stop words. For example, words like "is," "the," "an," "in" are stop words and should be removed from the dataset. So, we looped through our dataset and removed all the stop words. The next step was to create a TF-IDF matrix from the resultant data frame.

C. Creating Term Frequency- Inverse Document Frequency ($TF - IDF$) Matrix

The term $tf - idf$ is a combination of two distinct terms, term frequency and inverse document frequency. Term frequency is the number that shows how often a particular word appears in a set of documents, and inverse document frequency is the value that tells us how common or rare a particular word is in a set of documents [14]. $Tf - idf$ approach is a traditional approach mainly used in information retrieval and search engines. The $tf - idf$ score is calculated using the eq (1) [13].

$$tf\,idf(t,d,D) = tf(t,d).idf(t,D) \qquad (1)$$

Where 'D' is the set of documents, and 't' is the term frequency of the document 'd'. Whereas term frequency is calculated by the eq. (2) given below.

$$tf(t,d) = \log\big(1 + freq(t,d)\big)\, and\ idf(t,D) = log(N/count(d \in D{:}\, t \in d)) \qquad (2)$$

Converting our data frame to a $tf - idf$ matrix would make it easier for our machine learning algorithm to grasp the word's real importance, and it would increase the efficiency of our algorithm. In the figure, you can see the number of times a word has appeared in the corpus and that word. After that, we applied principal component analysis so that the $tf - idf$ matrix can be viewed in a 2-dimensional space. It is applied when we are dealing with a big dataset.

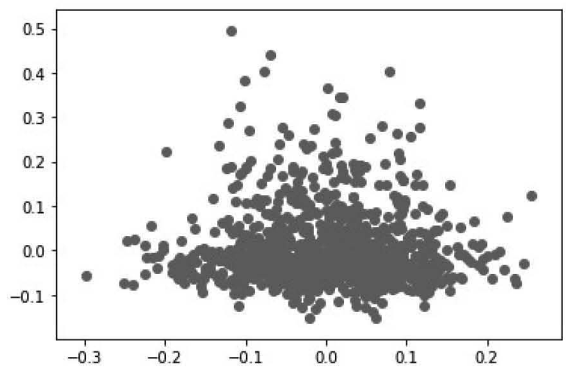

Fig. 4. PCA data points of tf-idf matrix

Principal component analysis is a process to reduce dimensionality of a large dataset so that they can be visualized on a 2-dimensional canvas while preserving as much variability as possible. The shape of our tf-idf matrix is (8675,791). This means that our tf-idf matrix has 8675 dimensions and each dimension has 791 elements which makes it impossible to visualize this matrix on a 2-dimensional space.

In the above figure, you can see that data points are converged between 0.1 and 0.2. This again shows the dataset is uneven and mostly consists of similar words and occurred multiple times in the dataset.

D. Forming Clusters Using K-Means

K-means is an unsupervised machine learning algorithm that distributes the data in the form of clusters based on a specific criterion (calculation of centroids) [15]. Clusters are a collection of various data points that are placed together due to certain similarities. Centroids are calculated based on either **using the Euclidian's distance or Jaccard's coefficient** [15]. Eq. 1 shows the formula to calculate the centroid value of any two given data points using Euclidean distance.

$$dist((x,y),(a,b) = \sqrt{(x - a)^2 + (y - b)^2} \qquad (3)$$

We used the default k-means algorithm from python to visualize the $tf - idf$ matrix to get the central intuition of the dataset we are dealing with. K means calculated by measuring the distance between the data point and the centroid value, and the resultant value would decide the cluster. Fig. 5 shows clusters of each type indicator to see how versatile the data is. Red colored datapoints represents the words related to introversion-extroversion whereas green colored datapoints represents the words that are related to Intuition and Sensing. Blue colored data points represent the words related to Feeling and Thinking and Yellow colored datapoints represents the datapoints related to Judging and Perceiving. [15]. It is observed that each cluster represents similar data points across many type indicators.

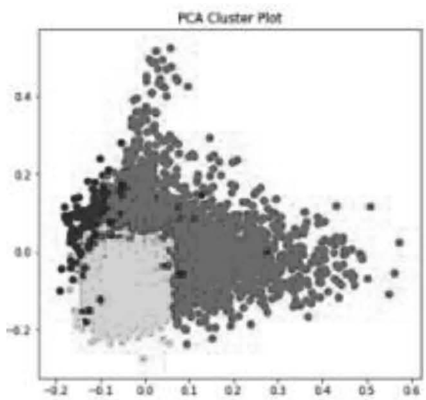

Fig. 5. PCA Cluster Plot

E. Application of Extreme Gradient- Boosting

Boosting is a process of converting weak learners into strong learners. We trained a model for a limited number of times on different parameters to increase its predictive ability [16]. The simplest variant of boosting is Adaboost. In Adaboost, different weights were assigned into the original data set and the classifier is trained on the dataset. Later, decision tree is analyzed to find the prediction rate and then the weights are modified [17].

The criteria behind this is of higher weights being assigned to those data points that are difficult to predict, and lower weights are assigned to data points that are easy to predict.

After assigning the new weights, the classifier is trained on the modified data set and the accuracy of the new prediction is calculated. Until a satisfactory prediction accuracy of model is gained, these steps are repeated. During this process ensemble trees are being made as well.

Gradient boosting is different from AdaBoost, and instead of using weights, we use the gradients in the loss function $y = ax + b + e$ where e is the error/loss function [16]. The loss function is a measure of how well the coefficients fit the underlying dataset. The nature of the loss function depends on the nature of the model and its primary purpose. Since we are constructing a classifier, our loss function calculates how good the model is of classifying people who show introverted or extroverted personalities. We have used a variation of Gradient boosting called Xgboost with some promising results.

F. Hyperparameters Tuning in XGBOOST

The concept of extreme gradient boosting was discussed by Tianqi et.al [17]. It won many Kaggle competitions and was well received by the data science community. It is an extreme version of gradient boosting. Following are some factors that make Xgboost an optimal algorithm.

- Parallelization
- Tree Pruning
- Hardware Optimization

Since we intend to build a classification model with the help of Xgboost, we convert the posts to binary type data so that our model can be easily understood [17]. Therefore, the data is split into X/Y format. X being the tf-idf matrix and Y being the binarized MBTI data. After that, X is split further into training and testing sets and we trained our first classifier on each type indicator. We used the default parameters of the Xgbclassifier to train the model. At this point, we did not tweak the parameters because we wanted to study the model's initial results.

The model returned satisfactory results. All type indicators were showing between 65-75% accuracy on average. There was room for improvement.

The next approach was to split the classifier and train separately on each time indicator while monitoring the performance through validation and loss. We modified the parameters this time. We specified ten early stopping rounds. The primary purpose of this parameter is to train the dataset until a satisfactory validation score is achieved with minimum loss value. We also modified the maximum tree depth from 6 to 10. The max depth parameter defines the number of iterations the model will perform. As we know that the boosting model builds a decision tree, so the maximum depth shows the number of trees our model will build. We increased it so that our classifier gives some more accurate results while considering the overfitting scenario. The accuracy rate was slightly increased, but it needed further improvements. We refrained from increasing the depth of the tree to avoid overfitting. We modified the hyperparameters to fit our model.

Hyperparameters are those that significantly impact the architecture of the model. Parameters like polynomial features, maximum depth, the minimum number of samples, learning rate, etc. are hyperparameters [18]. Since we are applying gradient descent, the maximum depth, estimators, threads, and learning rate are our hyperparameters. We set the max depth back to 6 and limited the number of decision trees ($\boldsymbol{n\_estimators}$) from the default value to 200. Learning rate or shrinkage ($\boldsymbol{learning\_rate}$ in XGBoost) should be set to $\mathbf{0.1}$ or lower, and smaller values will require more trees, so we tuned the learning rate to $\mathbf{0.2}$. Also, we applied the k-fold cross-validation to estimate the skill of our model. This improved our model significantly, giving an average result of $\mathbf{85-90\%}$ accurate [19].

IV. RESULTS

Table I shows the final accuracy of each classifier after hyper-parameter tuning. The feeling and Thinking classifier scored a little less as compared to other classifiers but almost all of the **classifier's** accuracy falls within 85-90%range.

TABLE I. CLASSIFIER ACCURACY

Classifier	Accuracy %
IE: Introversion (I) / Extroversion (E)	89.01%
NS: Intuition (N) – Sensing (S)	85.96%
FT: Feeling (F) - Thinking (T)	84.19%
JP: Judging (J) – Perceiving (P)	85.42%

If we take the average accuracy of all the classifiers, that would be 86.3%, which is pretty good accuracy. However, we performed some tests on some bloggers' data to predict their personality type and compare the results with their actual MBTI type.

TABLE II. ACTUAL VS. PREDICTED VALUES

Blogger Name	Actual MBTI Type	Predicted MBTI Type
Amanda Lane	INFP	INTP
Joel Laramee	ENTJ	INFJ
Nathan Stanish	ESTP	ISFP
Sophie Ryder	ENFP	INFP
Kara Nicole Bannet	INFJ	INTJ
Archana Ramanathan	ENFP	ENTJ
Kiana Lovelace	ISTP	ISTJ
Ellie Kirk	ENFJ	ESFJ
Joe Djoremy	ISTJ	ISTP

Table II shows the predicted MBTI type versus their actual MBTI type. The results obtained are promising. For instance, in the case of Kiana Lovelace, our model predicted her MBTI type as ISTJ but she claims to be an ISTP. It is almost accurate with an exception of false decision by the Judging-Perceiving classifier. Firstly, we created a dataframe from the test data before performing transformation. After data preprocessing, we transformed the data into tf-idf matrix so that it can be feed to our model for testing. We set the hyper parameters to the exact values we had for our training model and passed the tf-idf matrix to our model for prediction.

Our model delivered excellent performance, but it fails to predict the extroversion nature of the individual in various instances. That is because the data in which the model was trained contained more introverted type posts rather than the extroverted type. If the data were evenly distributed, the results would have been different. If we compare our machine learning algorithm with other algorithms like Naive Bayes or neural network-based models, our model performs slightly better if not the best. The overall accuracy of traditional Naive Bayes and support vector machine algorithms was 85% accurate, and LSTM-based models showed an accuracy of 82%.

V. CONCLUSION

In this paper, Gradient boosting techniques to predict the personality is proposed. This technique converts weak learners to strong learners, but it also shows the right intuition of the data. Extreme Gradient Boosting (Xgboost) is a perfect algorithm for datasets ranging from medium to large. If the dataset contains millions of rows, this algorithm would not be suitable because of its use of ensemble trees. The depth and iterations are required to be increased in that case and the memory could be easily choked.

This paper is intended to introduce the concept of extreme gradient boosting along with its **hyper parameters' tuning** as a means of predicting MBTI personality type from a user's data and extracting valuable insights from it. There is some room for improvement as shown by the results in Table 1. For instance, if we apply some more hyper parameter tuning techniques such as increasing the tree depth or increasing the number of iterations on a more balanced dataset can greatly improve results. It is all about fine tuning the parameters in accordance with the nature of the dataset that will result in optimal performance of the machine learning algorithm

REFERENCES

[1] B. Marr, "How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read", *Forbes*, 2020. [Online]. http://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/. [Accessed: 20- Jun- 2020].

[2] E. Graham-**Harrison and C. Cadwalladr, "Revealed: 50 million** Facebook profiles harvested for Cambridge Analytica in major **data breach," The Guardian, 17**-Mar-2018. [Online]. Available: https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election.[Accessed:19-Jun-2020].

[3] S. Bharadwaj, S. Sridhar, R. Choudhary, and R. Srinath, "Persona Traits Identification based on Myers-Briggs Type Indicator(MBTI) - A Text Classification Approach," *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Bangalore, 2018, pp. 1076-1082, doi: 10.1109/ICACCI.2018.8554828. [Accessed: 19-Jun-2020].

[4] **J. Brownlee, "A Gentle Introduction to the Gradient Boosting** Algorithm **for Machine Learning,"** *Machine Learning Mastery*, 21-Aug-2019. [Online]. Available: https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/. [Accessed: 19-Jun-2020].

[5] **"Myers Briggs Personality Types," Myers Briggs Personality** Types - Introduction and Overview. [Online]. Available: https://www.teamtechnology.co.uk/tt/t-articl/mb-simpl.htm. [Accessed: 19-Jun-2020].

[6] S. Bharadwaj, S. Sridhar, R. Choudhary, and R. Srinath, "Persona Traits Identification based on Myers-Briggs Type Indicator(MBTI) - A Text Classification Approach," 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, 2018, pp. 1076-1082, doi: 10.1109/ICACCI.2018.8554828. [Accessed: 19-Jun-2020].

[7] **J. Brownlee, "A Gentle Introduction to the Gradient Boosting** Algorithm for Machine Learn**ing," Machine Learning Mastery,** 21-Aug-2019. [Online]. Available: https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/. [Accessed: 19-Jun-2020].

[8] M. M. Jawad Soumik, S. Salvi Md Farhavi, F. Eva, T. Sinha, and M. S. Alam, "Employing Machine Learning techniques on Sentiment Analysis of Google Play Store Bangla Reviews," 2019 22nd International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 2019, pp.1-5,doi: 10.1109/ICCIT48885.2019.9038348. [Accessed: 10- Jun- 2020].

[9] Byers, Alison & Boochever, Rachel & Sumner, Chris & Byers, Alison & Boochever, Rachel & Park, Gregory. (2012). Predicting Dark Triad Personality Traits from Twitter Usage and a Linguistic Analysis of Tweets. Proceedings - 2012 11th International Conference on Machine Learning and Applications, ICMLA 2012. 2. 10.1109/ICMLA.2012.218. [Accessed: 10- Jun-2020].

[10] J. Golbeck, C. Robles, M. Edmondson, and K. Turner, "Predicting Personality from Twitter," 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, Boston, MA, 2011, pp. 149-156, doi: 10.1109/PASSAT/SocialCom.2011.33. [Accessed: 10- Jun-2020].

[11] K. Zvarevashe and O. O. Olugbara, "Gender Voice Recognition Using Random Forest Recursive Feature Elimination with Gradient Boosting Machines," 2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD), Durban, 2018, pp. 1-6, doi: 10.1109/ICABCD.2018.8465466.

[12] **"The History of Katharine Briggs, Isabel Myers, and the MBTI®," Truity,** 19-Aug-2019. [Online]. Available: https://www.truity.com/myers-briggs/story-of-mbti-briggs-myers-biography. [Accessed: 19-Jun-2020].

[13] **"Tutorials - Online Data Analysis & Interpretation: DataCamp,"** DataCamp Community. [Online]. Available: http://www.datacamp.com/community/tutorials/stemming-lemmatization-python. [Accessed: 19-Jun-2020].

[14] Scott, William. "TF-IDF for Document Ranking from scratch in python on real world dataset.", *Medium*, 2020. [Online]. Available: https://towardsdatascience.com/tf-idf-for-document-ranking-from-scratch-in-python-on-real-world-dataset-796d339a4089. [Accessed: 10- Jun- 2020].

[15] Dabbura, Imad "K-means Clustering: Algorithm, Applications, Evaluation Methods, and Drawbacks", *Medium*, 2020. [Online]. Available: https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a. [Accessed: 20- Jun- 2020].

[16] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial", *The Frontiers*, 2020. [Online]. Available: http://www.frontiersin.org/articles/10.3389/fnbot.2013.00021/full. [Accessed: 20- Jun- 2020].

[17] **Nishida, Kan. "Introduction to Extreme Gradient Boosting in Exploratory." Medium, Learn Data Science, 21 Mar. 2017,** blog.exploratory.io/introduction-to-extreme-gradient-boosting-in-exploratory-7bbec554ac7?gi=bc829c58b9bd# [Accessed: 20-Jun- 2020].

[18] **Jeremy Jordan, "Hyperparameter tuning for machine learning models.," Jeremy Jordan, 05**-Dec-2018. [Online]. Available: http://www.jeremyjordan.me/hyperparameter-tuning/. [Accessed: 19-Jun-2020].

[19] S. M, "Why and how to Cross Validate a Model?", *Medium*, 2020. [Online]. Available: https://towardsdatascience.com/why-and-how-to-cross-validate-a-model-d6424b45261f. [Accessed: 20- Jun- 2020].