Predicting Myers-Briggs Personality Types through Machine Learning

Conducted by Ted Strombeck, Mentored by Professor Pavel Bělík

URGO Research at Augsburg University, 2211 Riverside Ave, Minneapolis, MN, USA

Abstract: As time and our technological capabilities increase, we have more of a use for various machine learning practices and approaches. This can be seen as machine learning is increasingly being used in the world around us from email spam and malware filtering, picture recognition, virtual personal assistants, targeted advertisements, and even for online customer support. For personality prediction, Neuro Linguistic Programming (NLP) as of now, is the most common approach to assess personality prediction as it assumes that there is a connection between cognitive function and language tendencies that are learned through experience. Myers-Briggs Personality types are the most common type indicator to use when using various machine learning techniques as it is more applicable to consumers. However, in the world of psychology it is a frequent debate between Myers-Briggs and the Big Five type indicator. There are many different algorithms use when attempting to determine personality types. This study evaluates currently existing algorithms and improves upon the accuracy and reliability results that have been reported in recent personality prediction studies. The results from this study can help Psychologists automate personality prediction as well as assisting NLP practitioners in regard to which algorithms are worth investigating and which algorithms are not worthwhile for their own personal research.

---

1. Introduction
    a. Neuro-linguistic programming

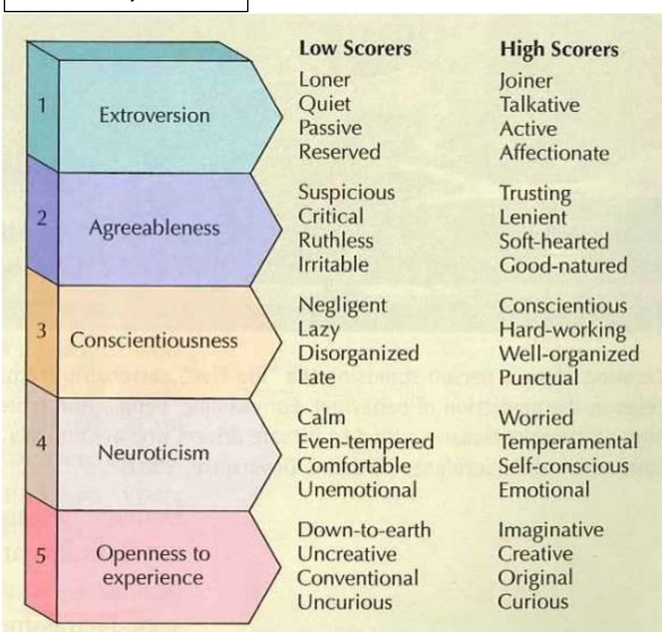
Neuro-linguistic programming was created by Richard Bandler and John Grinder in the 1970s. They claim that there is a connection between cognitive processes and language tendencies hence the “neuro-” and “linguistic” aspects. The programming refers to how humans encode this into their brains as the programming portion of the title represents the behavioral patterns learned through experience. NLP has claimed to be able to treat problems such as depression, various phobias, as well as learning disorders. However, since NLP is considered a pseudoscience, this study doesn’t base its methods around those commonly associated with NLP, but its’ results could assist NLP practitioners applying different models and practices to their work similarly to how Amirhosseini and Kazemain’s most recent article helps benefit NLP practitioners directly. This study differs from that by which it was not intended to directly assist NLP practitioners, however the results can be used by these practitioners to further their own research.

b. Myers-Briggs vs The Big 5 Personality Predictors

In the world of machine learning, the Myers-Briggs Type Indicator tends to be the most commonly used personality assessment indicator. Being used in nearly all of the most recent machine learning studies, it seems to overall reign supreme. This isn't the case in the world of psychology, as it is a frequent debate between MBTI and the Big Five personality indicators. The Big Five is generally more accepted in psychology as it is considered to be scientifically sound and is the base of modern academic psychological research. The main reason for this is because the Big Five measures conscious personality traits, where MBTI is an indicator of personality type. For example, the Big Five will tell you how much of an extrovert you are whereas the Myers-Briggs Type Indicator will tell you how likely you are to be an extroverted type, as mentioned in Park's, et al. (2014) study.

The big five factor model is classified as a taxonomy, a classification branch, aimed at covering most aspects of personality. It hopes to accomplish this with five factors; Extraversion, Neuroticism, Conscientiousness, Agreeableness and Openness. Extraversion is made up of sociability, talkativeness, assertiveness, and excitability. If someone scores high in Extraversion it means that they generally enjoy being with people, participating in social events and gatherings, and can be seen as the life of a party. On the other hand, if someone scores low in extraversion, they are quite the opposite; which can be seen in figure 1 as for them, being around people is often hard, they tend to dislike social events and gatherings, and can be often seen as boring or shy from an outsider's point of view. Neuroticism is defined as a tendency toward anxiety, depression, and other negative feelings one would associate with those listed at the start of the sentence. An individual with high neuroticism would tend to have more depressed moods and more frequently and severely suffer from these negative feelings. Conscientiousness is described as the trait of being careful when performing certain tasks. This means that an individual with high levels of conscientiousness desires to do a task well and tends to be efficient and organized. If an individual is low in conscientiousness, they are most often considered to be easy-going, unorganized, and impulsive. Agreeableness is a trait often described as cooperative, kind, and friendly. In a sense agreeableness is the ability to agree and get along with others. An individual with high levels of agreeableness is considered to be more trusting, polite, and altruistic. The fifth factor, openness is viewed as the ability to embrace new ideas and experiences as well as being generally open-minded. An individual who has high levels of openness tend to be more curious and try new things, whereas the opposite could be seen as stubborn and unwilling to try new things (Kern et al. 2013). From the descriptions of the factors of the big five taxonomy, you can see that the big five uses physical observations; they rely on actions

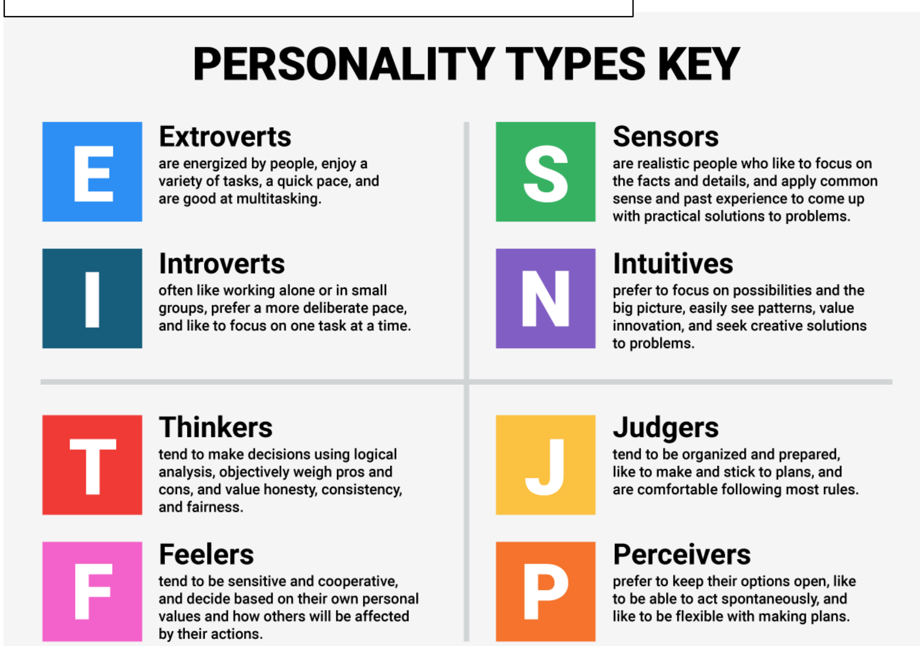
Figure 1: Big Five taxonomy



and solely on how you interact within different environments instead of attempting to theorize what occurs inside of people's heads.

This brings us to the Myers-Briggs Type Indicator or MBTI for short. Now this personality indicator is more common to be seen in research than the big five, this is because of the theoretical nature of the MBTI instead of the scientifically proven methods of the big five. The MBTI is based on a typology first introduced in the early 1900s by Carl Jung. It characterized individuals by their attitudes towards the inner and outer worlds (Introversion or Extroversion) and by an individual's cognitive assumptions (Judging or Perceiving). These are two of the four categories for the MBTI system, the other two categories being; Sensors or Intuitive and Thinkers or Feelers as shown in the four quadrants of figure 2. When these four categories are combined you get 16 combinations which refers to the 16 personality types in the MBTI system. The big five taxonomy uses high and low values of their five factors whereas the typology of the MBTI allows for probability prediction for one label or the other label. This makes it more appealing to the general public as instead of an individual saying they have low levels of extroversion; they are able to say they are introverted which has a more positive connotation than being labelled as a low extravert. This in turn makes this system more desirable to use for the general public. This is the main reason why you see the MBTI personality predictors more frequently used in research and the world outside of psychology instead of the big five personality assessment.

Figure 2. MBTI personality predictor categories

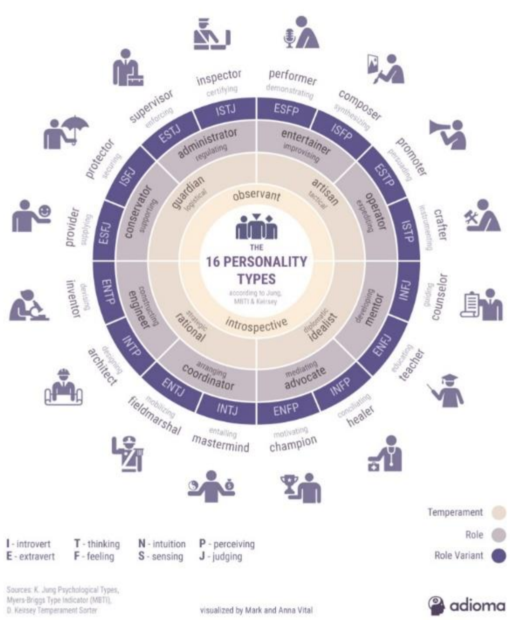


In this study we use the MBTI personality predictor. The main reason for using the MBTI system instead of the Big Five in this study is because it better characterizes people by their attitude toward the inner and outer world, inspires more research than the Big Five does, as well as being most commonly used personality predictor in businesses and education. Since it is most commonly used in various businesses and education facilities, it means that this study will help with one day creating a classification system that companies and schools will be able to use to optimize their workplace or learning environments in addition to helping NLP practitioners and psychologists (Gjurkovic Snajder (2018)).

c. Personality Types

In the MBTI system, there are 16 different personality types, being; Architect (INTJ), Logician (INTP), Commander (ENTJ), Debater (ENTP), Advocate (INFJ), Mediator (INFP), Protagonist (ENFJ), Campaigner (ENFP), Logistician (ISTJ), Defender (ISFJ), Executive (ESTJ), Consul (ESFJ), Virtuoso (ISTP), Adventurer (ISFP), Entrepreneur (ESTP) and Entertainer (ESFP). Each personality type has their own strengths and weaknesses as well as optimal career paths, parenthood, friendships, etc. (Gjurkovic Snajder (2018)). Although this system isn't as scientifically proven as the big five, for the average person, this system is able to help them understand themselves a lot better due to the sheer amount of research that has been completed on this personality prediction system. An example of the immense amount of information on MBTI personality types can be seen on the popular website 16personalities.com, where it goes into extreme detail on the personality types and topics mentioned above. This website also has a course which helps you master your personality as well as encouraging you to retake the test once every six months in order to be accurate to the individual at their current state and update as said individual changes over time. If the reader would like more in depth information about the specific differences between each personality type, we recommend that the reader would look at 16personalities.com for more specific and technical information. Figure 3 also highlights the key features and differences for the 16 different personality types.

Figure 3. MBTI 16 personality types



d. Background of Personality Prediction

As we advance our technological understanding and systems, among researchers there is an increase in interest and need in automated personality prediction using various social media platforms. Initially it was limited to only clinical psychology and counselling, however since Cambridge Analytica and the Facebook scandal in 2018, there has been an increase in machine learning applications from personality based targeted advertisements to customized personal assistants that we are all familiar with. Bleidorn, Wiebke, and Hopwood comment on the different generations of research in the machine learning application of personality assessment. The first generation introduced the topic of personality assessment autonomy through machine learning. The second generation tested different approaches on large data sets to refine and optimize the predictive accuracy of these approaches. The third and most current generation applied these different algorithms to test whether it can improve upon the already existing methods of personality assessment. This study further explained that the first generation was successful in providing initial evidence that individuals leave digital traces in their online environments that are representative of their personality traits. They also noted that second generation expanded on the first study's results using data from the "myPersonality" Facebook application. The tests offered on myPersonality were 25 different psychological tests including the big five personality test, not exclusive to the MBTI test (Bleidorn, W., & Hopwood, C. J. (2018)). However, this study mostly examined the correlation between whether or not there was a correlation and if machine learning personality assessment was better than pre-existing systems, all while using the big five personality test. Since in this study we are using the MBTI system, we can see that there is evidence of individuals having their personality traits present on social media platforms, but we are unable to use the specific validity rules that this study mentions and presents. This is due to Facebook removing their myPersonality dataset and our study having to use the Kaggle dataset which is structured and organized differently. The URL address for Kaggle is listed in the references section.

e. Recent Results of Personality Prediction and Automation

In the realm of personality prediction, a lot has been done with personality prediction and automation, however not much of it is done through textual data. Golbeck et al., was one of the earliest studies on personality prediction. In this study, they could accurately predict an individual's personality type by using the information available to them on twitter. This was a step towards where this study is today, as Golbeck proved that you could in fact predict personality types from available social media information. Which in turn now allows various studies to compare the different algorithms that would be most optimal for finding different personality types instead of having to test to see if the particular social media platforms would be able to support personality prediction in the first place. Amirhosseini and Kazemain's most recent article goes into detail about the various studies that have been conducted on textual personality prediction, both in the big five and MBTI personality tests. Their table (shown below) organizes the most recent and relevant studies, at the time of this research being conducted, to personality prediction based on textual information. It also highlights the different personality models each study used and the main method that they focused on to test the different personality model.

Table 1. "Research on personality type prediction and personality models used" – taken from Amirhosseini and Kzaemain's recent study

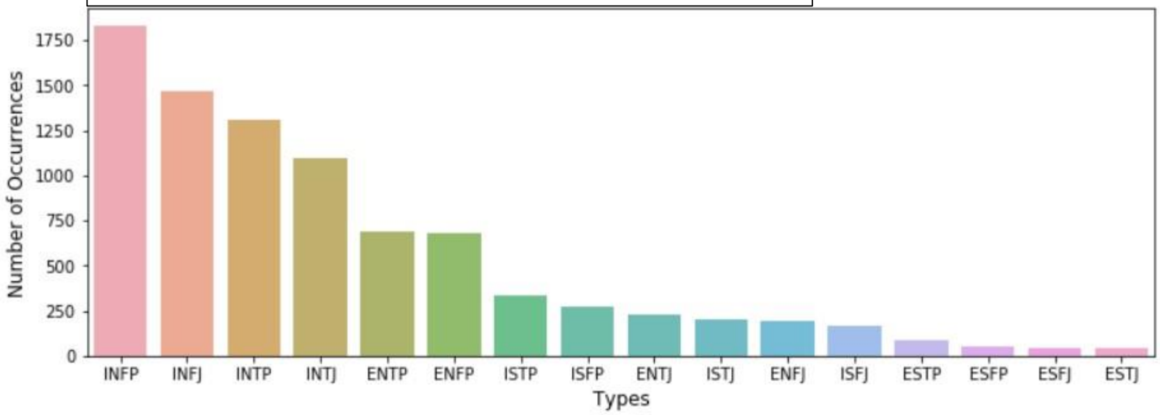
Champa and Anandakumar (2010)	MBTI Network	Artificial Neural
Golbeck and et al. (2011)	MBTI Algorithms	Regression
Komisin and Guinn (2012)	MBTI Bayes and SVM	Naïve
Wan and et al. (2014)	Big Five Naïve Bayes	Logistic Regression
Li, Wan and Wang (2017)	Big Five learning	Multiple Regression and Multi-task
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 Learning

From this table you can see that the study conducted by Cui and Qi in 2017 focused on not only one or two different algorithms, but they attempted to apply 4 different algorithms in their study. Overall, their study obtained 38% overall accuracy with their best model; this being the deep learning approach. However, when we compare the results we obtained from our study with other studies, we will use the individual models instead of overall accuracy to be able to fully evaluate the differences. When using Support Vector Machines, we were able to find similar results to Cui and Qi, however it is surprising that Cui and Qi's study didn't investigate XGBoost, as XGBoost is known as the gold standard for machine learning algorithms; winning a majority of modern machine learning competitions.

2. Data Set and Processing Methods
    a. Data Set Used

The data set that was used for this study is the very popular "(MBTI) Myers-Briggs Personality Type Dataset" on Kaggle. This is the most common data set to use when testing various machine learning models on MBTI prediction as it has 8,675 rows of data that contains a user's personality type as well as a collection of the last 50 items that they have posted, each separated by three pipe characters "|||". This information was gathered through the PersonalityCafe forum, which as Cui and Qi point out, is most likely subject to various biases. Firstly, users were given a personality questionnaire which determined their personality type, then they were free to communicate with others on the forum; mostly conversing about random topics. Unfortunately, this is the biggest public domain data set which means that in the terms of research, unless a massive corporation gives the rights to use their data, most research will have to use this heavily flawed data set. To display and highlight the flaws in this data set, we will use Python merely to visualize the data set and see the general distribution of personality types across the data set.

Figure 4. Kaggle Dataset Personality Type Distributions



As you can see in figure 4, it is quite clear that this data set is anything but uniform. Given that in the world, personality types are not equally distributed, they are also not as skewed as they are in this data set. A quick comparison between the recorded global personality percentages and the personality percentages in this data set showcases this issue, as highlighted in table 2. Table 2 also highlights the vast differences in the personality types; ESFP, ENTP, and INTJ, where each one of those categories has over a 10% difference from what is represented in the data set and what has been recorded over the course of a few years.

Table 2. Personality Percentages in Dataset in comparison to recorded percentages in personality tester database

Personality Type	Percentage in Kaggle Dataset	Percentage estimation from Personality Max Database	Percentage difference
ESTJ	4.09%	13%	-8.91%
ESFJ	3.59%	12%	-8.41%
ISTJ	9.41%	8.5%	0.91%
ISFJ	2.11%	7%	-4.89%
ESTP	2.02%	10%	-7.98%
ESFP	**23.47%**	**11%**	**12.47%**
ISTP	4.81%	6%	-1.19%
ISFP	1.66%	6%	-4.34%
ENTJ	**14.74%**	**4%**	**10.74%**
ENTP	8.28%	4.5%	3.78%
INTJ	**14.75%**	**1.5%**	**13.25%**
INTP	6.81%	2.5%	4.31%
ENFJ	1.42%	4%	-2.58%
ENFP	1.07%	7%	-5.93%
INFJ	0.59%	1%	-0.41%
INFP	1.15%	2%	-0.85%

b. Text Preprocessing

Since this data set has issues, there is a need for preprocessing the information to eliminate as many of the flaws from this data set has as possible. At a brief glance over the data set, the reader can see that there are some URLs included within the data set. Immediately that has to be removed as it could throw off the most frequently occurring words or phrases within the data set as all URLs begin with "https://" (Hernandez & Knight (2016). However, before we could clean any of the text, we needed to transform the data set into a Document Term Matrix (DTM), as that was the container type that our package used for cleaning the textual information. Next, we removed all of the numbers from our data set as much like the URLs, they could disrupt the algorithms and frequency counts. We then removed all stop words. Stop words are words that are essentially meaningless as they are often filler information or words that are only useful to assist grammar and syntax rules, these include words such as: the, and, but, or, etc. To remove the stop words, we again used the tm package in R as it came with a set of words that were already defined as "stop words". After said stop words were removed, we removed the punctuation and took the roots of all of the words. The reason for taking the roots of all of the words was so that in our frequency counter, it would count similar words, referred to as tokens, such as: "run", "running", and "ran", all as the same word so that our frequency counter was guaranteed to have unique tokens instead of different versions of the same tokens. Once this was completed, we then removed certain words that we saw in the data set that could potentially again disrupt the overall performance of our various algorithms in which we tested. These contained all of the various personality types, as in some recorded conversations on the forum, individuals would state how they were one personality when they felt as if they should have been another personality type. This issue was pretty common and in order to increase the legitimacy of this study, we removed all of the personality types from the post's column in our data set. This wrapped up the cleaning of the data set and into the phase of preparing the data set for the various tests that we would put it through. We first used the Naïve Bayes algorithm as our benchmark test, and upon the completion of the first test, decided to split up the personality predictor into four separate models instead of the original multi-classifier model. This meant we also had to adjust our data set, simply adding four different columns to the end of the data set to represent each of the letters contained in the user's assigned personality type. Once we had our individual columns set up and our frequency count was complete and added to our DTM, we saved the data set as a data frame (DF) for easy manipulation and transformation between different container types that the different algorithms used as input.

3. Algorithms Used

The programming language used for this study is R, as although Python offers more frequent usability and slightly easier readability, R makes up for this with its' power and efficiency as it is a programming language designed for statistical computing and graphics. In terms of model creation and tuning, we first had a single classifier and single predictor model but switched to a 4 classifier and 4 predictor model as those gave us better accuracies as well as allowing us to tune the models better. It is very important to note that in machine learning, you should never touch the test set until you have fully fitted and tuned your model. To accomplish this, we used a 75% and 25% split; 75% of the data set being used for training, validation, and tuning, whereas 25% of the data set was used for the testing. Of the 75%, we split it again 70% for initial training and 30% for cross-validation and tuning. It is important to note that when tuning, it is very easy to overfit the model to your data which lowers the accuracy for your model on the test set. To combat this, we reshuffled the training portion of the data to ensure that the models we did tune would not be subject to overfitting. This took a lot of human input and time to be careful and make sure that the values we selected for tuning were indeed the best that they could be.

a. Naïve Bayes

The first algorithm that we used was Naïve Bayes. In order to implement this algorithm, we used the e1071 package for R. We used this package to initially create a single classifier and a single predictor as our benchmark test as previously stated. We used the Naïve Bayes algorithm as our benchmark test as it is a fairly simple algorithm that was proven to be effective with small MBTI data sets as well as the fact that it assumes independency between the features and tokens of the data set. The initial results of this first model were similar to other results found in previously conducted studies which is why it was perfect as our benchmark algorithm as we are attempting to improve upon the results previously gathered. After seeing the accuracy percentages, we then decided to divide the single classifier and single predictor model up into 4 different classifiers and models, one for each of the letters associated with MBTI personality types. Instantly we saw better results when using the four binary model approach instead of using a 16 multiclass classification approach.

b. Logistic Regression

After we tested the Naïve Bayes algorithm, we moved onto Logistic Regression. For this algorithm, we also used the e1071 and the glmnet packages for R. The glmnet package seemed to work better in terms of accuracy as it allowed for more specific tuning compared to just the glm function. In the logistic regression algorithm, we tested both the Lasso (L1) and Ridge (L2) regularizations in order to see which one was better. For both types of regularization, we found the optimal value of lambda and applied that to the model as well as tuning after the initial tests were performed. As stated above, each time we would tune and test, we would reshuffle the data and start the process over again, all the while never touching the test data set. After many iterations, we found that the ridge regularization actually performed better than the lasso regularization when we had tuned the model.

c. SVMs

After logistic regression testing was complete, we moved on to support vector machines or SVMs for short. We used the kernlab package and both the svm and ksvm functions for R to accomplish this. The kernels we tested were the rbfdot, polydot, and vanilladot kernels. The results posted in the results tables only contain the results from rbfdot kernel as it produced the best results. The polydot kernel was secondary to rbfdot, and as expected the vanilladot kernel did not perform well at all for this type of task as it is a linear kernel which doesn't allow for much tuning or transformations. Other studies had claimed that these algorithms were also one of the best for smaller data sets, however they tended to still hold their ground when applied to bigger data sets. Although, SVMs weren't the best algorithm for our study, we found that this statement was generally true as they still held up when applied to a data set of around 8,675 personality types.

d. XGBoost

The final algorithm that we were able to implement in this study with the amount of time given was XGBoost. This algorithm has been often used and mentioned as one of the best models for classification and application problems. It can often be seen winning various machine learning competitions on Kaggle as well as other various machine learning competitions. To use this algorithm, we used the xgboost, pROC, and Matrix libraries to run the algorithm as well as to show and record the results that this algorithm obtained.

4. Results and Comparison to Recent Results

For this study, we used similar algorithms to the recent study conducted by Amirhosseini and Kazemian. In the two images below taken from their study, table 3 shows both their tuning accuracy and in table 4 it shows how well XGBoost performed in comparison to how well their Recurrent Neural Network worked. Overall, these results are interesting as the results from their study indicates that Neural Networks may be better suited for identifying the Feeling and Thinking category. This also confirms the belief that the most optimal model will most likely be a multi-algorithm model as different algorithms are better than others for the different model types. In our study, this wasn't overly apparent as XGBoost came on top for all 4 categories. However, our study didn't investigate deep learning frameworks, which means that we cannot confirm if any deep learning framework would be better for the categories that the other algorithms all struggled with; those being the "Thinking and Feeling" and the "Judging and Perceiving" categories.

Table 3. "Comparison of accuracy prediction before and after configuration" – from Amirhosseini and Kzaemain's recent study

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

Table 4. "Comparison of accuracy of the Extreme Gradient Boosting model and the recurrent neural network model" – also from Amirhosseini and Kzaemain's recent study

Binary Class	MBTI Personality Type	Accuracy of Extreme Gradient Boosting	Accuracy of Recurrent Neural Network
IE	Introversion (I) – Extroversion (E)	78.17%	67.6%

NS	Intuition (I) – Sensing (S)	86.06%	62%
FT	Feeling (F) – Thinking (T)	71.78%	77.8%
JP	Judging (J) – Perceiving (P)	65.70%	63.70%

Our Best results are posted below, followed by the other various accuracies from our different models that we were able to implement and use throughout this study. Initially, one will notice how 3 of the 4 models in table 5 use the ROC metric whereas the other metric type was accuracy. After investigating the rationale onto why this occurred, we determined that it was due to the package handling: as for the intuition and sensing category, the accuracy metric had better natural handling for this specific category, whereas for the other models, the ROC metric was the best metric. We came to this conclusion as there weren't any other clear signs as to why accuracy was only better in 1 category or why the ROC metric dominated the other 3 categories.

Table 5. Best model accuracies recorded

MBTI Personality Type	Algorithm/Package Used	Accuracy of Model	Metric
Introversion (I) – Extroversion (E)	XGBoost	81.38%	ROC
Intuition (I) – Sensing (S)	XGBoost	88.04%	Accuracy
Thinking (T) – Feeling (F)	XGBoost	87.97%	ROC
Judging (J) – Perceiving (P)	XGBoost	78.28%	ROC

Table 6. All model accuracies for Extroversion and Introversion

Model	Algorithm/Package Used	Metric	Accuracy of Model
Binary Model	Naïve Bayes	N/A	0.7510
L1 Regularization	Logistic Regression	Lasso Regression	0.7137
L2 Regularization	Logistic Regression	Ridge Regression	0.7732
Regularization Hybrid	Logistic Regression	Hybrid Regression	0.7603
SVM	Support Vector Machine	N/A	0.7713
Cross-Validated SVM	Support Vector Machine	Cross-Validation (10 – fold)	0.7736
ROC XGBoost	XGBoost	ROC	0.8138
Accuracy XGBoost	XGBoost	Accuracy	0.8065
Kappa XGBoost	XGBoost	Kappa	0.8031

Table 7. All model accuracies for Sensing and Intuition

Model	Algorithm/Package Used	Metric	Accuracy of Model
Binary Model	Naïve Bayes	N/A	0.7953
L1 Regularization	Logistic Regression	Lasso Regression	0.7326
L2 Regularization	Logistic Regression	Ridge Regression	0.8603
Regularization Hybrid	Logistic Regression	Hybrid Regression	0.8598
SVM	Support Vector Machine	N/A	0.8608
Cross-Validated SVM	Support Vector Machine	Cross-Validation (10 – fold)	0.8603
ROC XGBoost	XGBoost	ROC	0.8329
Accuracy XGBoost	XGBoost	Accuracy	0.8804
Kappa XGBoost	XGBoost	Kappa	0.8775

Table 8. All model accuracies for Thinking and Feeling

Model	Algorithm/Package Used	Metric	Accuracy of Model
Binary Model (I & E)	Naïve Bayes	N/A	0.7805
L1 Regularization	Logistic Regression	Lasso Regression	0.4822
L2 Regularization	Logistic Regression	Ridge Regression	0.5399
Regularization Hybrid	Logistic Regression	Hybrid Regression	0.5284
SVM	Support Vector Machine	N/A	0.5408
Cross-Validated SVM	Support Vector Machine	Cross-Validation (10 – fold)	0. 5422
ROC XGBoost	XGBoost	ROC	0.8797
Accuracy XGBoost	XGBoost	Accuracy	0.8019
Kappa XGBoost	XGBoost	Kappa	0.7993

Table 9. All model accuracies for Judging and Perceiving

Model	Algorithm/Package Used	Metric	Accuracy of Model
Binary Model (I & E)	Naïve Bayes	N/A	0.6524
L1 Regularization	Logistic Regression	Lasso Regression	0.5740
L2 Regularization	Logistic Regression	Ridge Regression	0.6040
Regularization Hybrid	Logistic Regression	Hybrid Regression	0.6007
SVM	Support Vector Machine	N/A	0.6030
Cross-Validated SVM	Support Vector Machine	Cross-Validation (10 – fold)	0.6035
ROC XGBoost	XGBoost	ROC	0.7827

Accuracy XGBoost	XGBoost	Accuracy	0.7212
Kappa XGBoost	XGBoost	Kappa	0.7224

5. Conclusions

This research has investigated the newly proposed machine learning method for automating MBTI type prediction as well as improving upon the accuracies obtained from Amirhosseini and Kazemain's study. It also highlights 4 different algorithms that can be used for this type of machine learning task, as well as showing which algorithm and model performed the best with a certain metric or parameter specified. This study concludes that XGBoost is the best classifier out of the various algorithms used, however a promising method would be to use deep learning (not just basic recurrent neural networks). The various packages and libraries used for this study are as follows; tm, caret, data.table, SnowballC, tidyverse, Matrix, xgboost, pROC, e1071, kernlab, keras, glmnet, and dplyr. These may not have been the most popular packages as during the time that this research was conducted, the programming language R, underwent an update, meaning that many of the popular and easy to use packages were not the most quickly updated. Since not all of the packages that we wanted to use were updated, we had to find a couple of workarounds. This mostly meant that the tuning was not always completed through caret, on some occasions it had to be done manually. In regard to the content in which this project adds to the community, it clarifies prior research as well as improves the accuracies obtained in recently released research. It also confirms that the multi – model approach for determining Myers-Briggs personality types opposed to the multi-classification approach yields better results.

6. Research in this field going forward

Going forward, there is a need for a deep dive of research into two categories; semi-supervised learning as well as deep learning framework. As of the time of this paper being written, there isn't much information on either category. Deep learning framework would be useful as it has potential to improve upon the results gathered from this study and prior studies. Semi-supervised learning would be very useful as well, as it would provide a new perspective or method on the task of classifying Myers-Briggs personality types. However, in order to help benefit all research which uses the Myers-Briggs type indicator, there is a need for a public database to use for research purposes. The best public database for various research studies on this topic is the Kaggle database that was used in this study, however, as mentioned above in this paper, there are flaws with this database. This problem isn't as easy to solve as investigations of semi-supervised learning and deep learning framework would be. Despite being the area of utmost importance, this is the hardest task and will most likely be the last task to be completed due to the limited budget of research conducted and published publicly.

7. References

a. Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H., & Seligman, M. E. P. (2014, November 3). Automatic Personality Assessment

Through Social Media Language. Journal of Personality and Social Psychology. Advance online publication. http://dx.doi.org/10.1037/pspp0000020

b. Vu, V., Labroche, N., & Bouchon-Meunier, B. (2010). Active Learning for Semi-Supervised K-Means Clustering. *2010 22nd IEEE International Conference on Tools with Artificial Intelligence*. doi:10.1109/ictai.2010.11

c. Weston, J., Ratle, F., Mobahi, H., & Collobert, R. (2012). Deep Learning via Semi-supervised Embedding. *Lecture Notes in Computer Science Neural Networks: Tricks of the Trade,* 639-655. doi:10.1007/978-3-642-35289-8_34

d. Gjurković, M., & Šnajder, J. (2018). Reddit: A Gold Mine for Personality Prediction. *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*. doi:10.18653/v1/w18-1112

e. Hernandez, R., & Knight, I. S. (2016). Predicting Myers-Briggs Type Indicator with Text Classification. Retrieved May 25, 2020, from https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1184/reports/6839354.pdf

f. Amirhosseini, M. H., & Kazemian, H. (2020). Machine Learning Approach to Personality Type Prediction Based on the Myers–Briggs Type Indicator®. *Multimodal Technologies and Interaction, 4*(1), 9. doi:10.3390/mti4010009

g. Bachrach, Y., Kosinski, M., Graepel, T., Kohli, P., & Stillwell, D. (2012). Personality and patterns of Facebook usage. *Proceedings of the 3rd Annual ACM Web Science Conference on - WebSci '12*. doi:10.1145/2380718.2380722

h. Plank, B., & Hovy, D. (2015). Personality Traits on Twitter—or—How to Get 1,500 Personality Tests in a Week. *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. doi:10.18653/v1/w15-2913

i. Golbeck, J., Robles, C., Edmondson, M., & Turner, K. (2011). Predicting Personality from Twitter. *2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing*. doi:10.1109/passat/socialcom.2011.33

j. Wald, R., Khoshgoftaar, T., & Sumner, C. (2012). Machine prediction of personality from Facebook profiles. *2012 IEEE 13th International Conference on Information Reuse & Integration (IRI)*. doi:10.1109/iri.2012.6302998

k. Zhao, Z., & Liu, H. (2007). Spectral feature selection for supervised and unsupervised learning. *Proceedings of the 24th International Conference on Machine Learning - ICML '07*. doi:10.1145/1273496.1273641

l. Kern, M. L., Eichstaedt, J. C., Schwartz, H. A., Dziurzynski, L., Ungar, L. H., Stillwell, D. J., . . . Seligman, M. E. (2013). The Online Social Self. *Assessment, 21*(2), 158-169. doi:10.1177/1073191113514104

m. Sindhwani, V., Niyogi, P., & Belkin, M. (2005). Beyond the point cloud. *Proceedings of the 22nd International Conference on Machine Learning - ICML '05*. doi:10.1145/1102351.1102455

n. Bleidorn, W., & Hopwood, C. J. (2018). Using Machine Learning to Advance Personality Assessment and Theory. *Sage*. doi:10.31234/osf.io/ctr5g

o. Cui, Brandon and Calvin Qi. "Survey Analysis of Machine Learning Methods for Natural Language Processing for MBTI Personality Type Prediction." (2017)

p. Kaggle Dataset - https://www.kaggle.com/datasnaek/mbti-type