Myers–Briggs Type Indicator (MBTI) Classification using Text data

Dr Huzefa Rangwala
rangwala@gmu.edu
George Mason University

Janit Bidhan
jbidhan@gmu.edu

11th December 2021

Sashank Kakaraparty
skakarap@gmu.edu

Abstract

The Myers–Briggs Type Indicator (MBTI) which is based Neuro Linguistic Programming (NLP) on is one of the most popular and reliable methods to predict personality. MBTI is a self-report instrument and is non-judgmental. It is an indicator of preferences, and it does not test a person's Intelligence quotient (IQ) or Maturity, etc. In this project, a new machine learning approach has been developed for personality type prediction. We were able to create a fast and effective way for analyzing the personality type using the text data extracted from social media platforms which could be either internal or external like Twitter, Facebook, slack etc. The data was not in a usable format and required a good amount of pre-processing since it contained a lot of special characters, numbers, and hyperlinks. Instead of discarding the hyperlinks we devised an approach to extract data from these links. Once cleaned, we tried different machine learning models and one of the gradient boosting machine learning models was able to identify the personality type of the user correctly, more than 67% of the times, out of the available 16 MBTI types. We created a demo application by integrating Twitter API and extracting text data from a given user's tweets.

1. Introduction

Neuro-linguistic programming (NLP) [1] is a psychological method that incorporates evaluating successful people's strategies and using them to achieve a personal objective. It establishes a link between taught thoughts, language, and behavioral patterns and specific outcomes. MBTI is one such tools of determining the personality traits.

Each MBTI personality type has its own set of strengths and weaknesses and understanding them may help the people understand their own personality and traits, which helps in eliminating conflicts, improving communication and effective teamwork. The MBTI also encourages the people to understand what roles they could excel in. The MBTI is concerned with the significant variations in people that arise from where they like to focus their attention, how they prefer to process information, how they prefer to make decisions, and the lifestyle choices they make.

When working as a team, it is important to play to people's strengths and minimize the impact of their weaknesses to enable the team to perform better. Thus, it's crucial for an organization/university to have a deep understanding of the people/students and aid them with requisite professional training. There are several ways of evaluating the personality of a person.

Knowing the personality type one can have Increased self-awareness. The person could discover how people differ in terms of energy sources, information gathering, decision-making, and lifestyle. They could develop a respect for each person's unique talents and abilities. They may learn how to improve your team's performance by using their own and others' skills. They will be able compile a list of areas or possibilities for personal or professional growth.

The problem statement, a literature review including the history of Neuro-linguistic programming (NLP) and how the MBTI was developed, and research on Automated Personality Prediction will be discussed in the following parts. After that, we'll examine at the project's technical aspects, including approaches, procedures, experiments, evaluation metrics, and deployment.

2. Problem Statement:

The MBTI is administered as a 100-item questionnaire that takes a considerable time to complete [23]. However, rather than administering a questionnaire exam to assist them understand their own personality, the individual may provide social media handles as part of the procedure. Our solution would look at text rather than questionnaire replies to estimate a user's personality using the MBTI scale, which is faster than filling out a questionnaire. The text might originate from social media accounts like twitter, Facebook, slack, and other platforms with the user's consent.

2.1. Notations

MBTI: Myers–Briggs Type Indicator, **INTJ**: **Introversion Intuition Thinking Judging, ESFP**: **Extraversion Sensing Feeling Perceiving** and similarly different combinations …

3. Literature Review

Personality refers to individual differences in characteristic patterns of thinking, feeling, and behaving. The study of personality focuses on two broad areas: One is understanding individual differences in particular personality characteristics, such as sociability or irritability. The other is understanding how the various parts of a person come together as a whole. [14]

"People are not nouns, but processes..." Richard Simons, 1997

The MBTI is the world's most widely used and popular personality tool. It is based on Carl Jung's psychological type theory [3] where "type" meant typical and not a particular kind of person. Kathryn Briggs and daughter Isabel Briggs Myers extended and expanded Jung's descriptions. They developed a structure of 16 types based on four opposites. These types are shown below with a famous personality. [4]

3.1. History of Neuro Linguistic Programming (NLP)

Neuro Linguistic Programming (NLP) is a collection of techniques that can help to identify how people think, how they communicate and how they behave. Meta programmes can have a major influence on behaviors as well as how people communicate with others. The initial list of NLP meta programmes included 60 different patterns. Many of these meta programmes have been combined by subsequent researchers to form a much smaller and more useful set. [5]

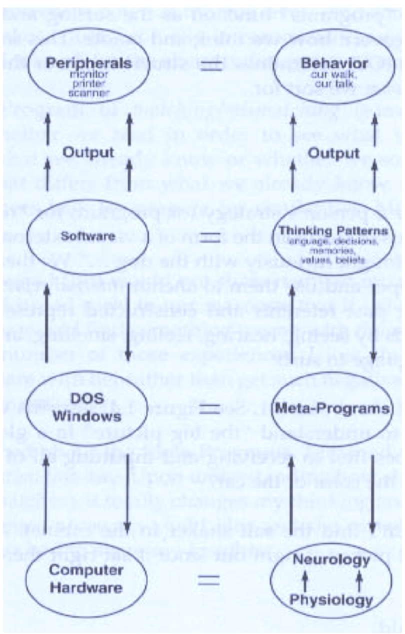


Figure 1: Meta-Programs

Dr. Michael Hall [6] depicts our human brain as an information processing mechanism in his book "Figuring Out People." [7] It would have its own hardware in our nervous system, brain, blood, chemistry, and neurotransmitters, among other things. All these biological aspects contribute to the world's expressions'(information or messages) intake, processing, and output. Our thinking patterns, ideational categories (we think and reason via "categories" Lakoff [8]), belief notions, valuational importance (or values those ideas that we perceive as extremely significant), programs for functioning, and so on make up our human "Software." We need this software that tells us how to process our ideas and emotions to run them. This operating system connects the hardware and software so that the neurology of the brain and body may input, process, and output information such as thoughts, ideas, beliefs, and so on. And this operating system is called the Meta-Programs. (Figure-1)

"Dr Wyatt Woodsmall" mentioned in his book "Metaprograms" [9] says "The Meta-Programs refer to those programs in our "minds" that lies above and beyond ("meta") the specific words determines their perspective, way of valuing, style of thinking and emoting, and pattern of choosing and behaving". He developed these meta programmes for use in business and therapy and integrated them with the Myers–Briggs Personality Inventory. [10]

He reduced the number of patterns 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.[11].

An individual's preferences are grouped into four dimensions, and different combinations of the personality type key in these categories reflect 16 different personality types. Figure 2 depicts the 16 personality types that come from an individual's preferences interacting, and each MBTI personality type's cognitive functions are described. The primary function of each kind is represented by the background color, while the auxiliary function is represented by the color of the text.

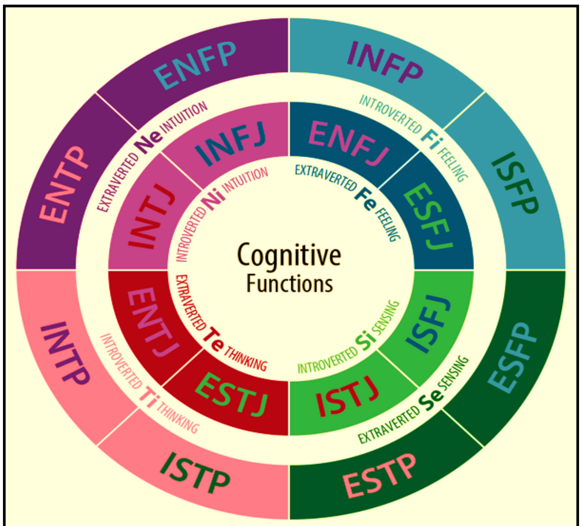


Figure 2: Cognitive Functions [15]

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) and these determine which one of 16 personalities are you:

- Extraversion (E) and Introversion (I). Extroverts are more action-oriented and feel energized after social interactions, whereas introverts are more thought-oriented and they "charge batteries" when they're alone.

- Sensing (S) and Intuition (N). This refers to the way we gather information from the world around us. Those who prefer sensing gather information from reality, facts, and experience, while the ones who prefer intuition prefer impressions, possibilities, abstract thinking.
- Thinking (T) and Feeling (F). After we gather data, we process them in two ways: thinking (logic-based) and feeling (emotion-based).
- Judging (J) and Perceiving (P). This refers to our attitude toward an outside world. Judgers prefer structure and order; Perceivers prefer being flexible and spontaneous.

Taking about other methods, Extroversion, agreeableness, conscientiousness, neuroticism, and openness are the five major trait qualities that make up the Big Five personality paradigm [12]. The four key characteristic traits that make up the Predictive Index (PI) Behavioral Assessment are dominance, patience, extraversion, and formality [13].

3.2. Research on Automated Personality Prediction

Researchers in the domains of Natural Language Processing and Social Science are becoming increasingly interested in automated personality prediction utilizing social media. Traditional personality tests have largely been used in clinical psychology, counseling, and human resource management up until now. Automated personality prediction from social media, on the other hand, offers a larger range of uses.

Most of the 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. There is little research on predicting personality types from textual data. For predicting MBTI personality types, traditional machine learning algorithms and neural networks have proved successful. Anand et al. [16] predicted the personality of a person from the baseline, the pen pressure, and the letters as found in an individual's handwriting. These parameters are the inputs to the Artificial Neural Network which outputs the personality trait of the writer. Golbeck et al. [17] published one of the first research on personality prediction using machine learning techniques using twitter data. The Nave Bayes and Support Vector Machine (SVM) approaches were utilized by Komisin and Guinn [18] to determine an individual's personality type based on their word choice. Wan et al. [19] employed a machine learning algorithm to predict the Big Five personality types of users based on their words on Weibo, a Chinese social media platform. Li, Wan, and Wang [20] employed the grey prediction model, multiple regression model, and multi-tasking model to predict the user personality type on the Big Five Learning Model. Tandera et al. [21] employed the Big Five personality model and deep learning architecture to estimate a person's personality based on their Facebook data and used MLP and LSTM+CNN 1D architectures. Cui and Qi [22] employed Baseline, Logistic Regression, Naive Bayes, and SVM to predict MBTI personality type from a social media post.

Research done on personality type prediction and personality models used the classification techniques as mentioned in the table 1.

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

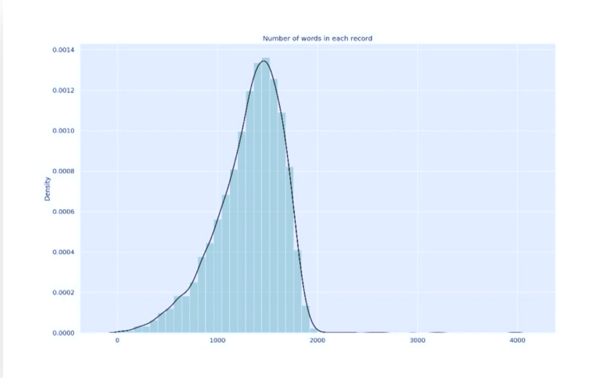
Table 1: Past research on personality type prediction and classification techniques

4. Methods and Techniques

4.1. Analysis and Preprocessing text data

The first task we took up, was a deep dive into understanding our text data, looking at the distribution of the words, the distribution of classes and identifying potential issues with our dataset. As and when we found issues, we integrated tools into our pre-processing pipeline to address them. This subsection will detail the results of the analysis along with the pre-processing steps performed based on the analysis.

Figure 3: Distribution of text HERE



4.1.1. Distribution of text data:

Before we headed to visualize the text, we cleaned the data by stripping out the symbols, special characters, and numbers from the user posts to leave only words for analysis.
One of the key observations we made was the presence of several hyperlinks in the posts made by users. We'll explain our approach to handling them, in detail in the next subsection.

4.1.2. Vectorization:

We require text strings to be represented in numerical format as vectors to perform machine learning on this data. Given a choice of several options such as N-Grams, Term Frequency – Inverse Document Frequency (TD-IDF) and Count Vectorization, we went with TD-IDF for it was more well-rounded. We applied Term Frequency – Inverse Document Frequency vectorization on our training dataset. TD-IDF is an improvement on the Count Vectorization, while building on its core concept. TD-IDF builds the vector representation based on two fundamental metrics: "Term Frequency"– The number of times it appears in a document – and "Document Frequency" – The number of times it appears when presented with a set of documents. It performs better than plain term frequency because it identifies only the words that are statistically relevant, instead of just picking commonly occurring words. The below formula represents the relationship between Term frequency and the 'inverse' document frequency.

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$ [30]

We used the TF-IDF implementation available from Scikit-Learn – TfidfVectorizer [29] with only the top 5000 features being considered. We also removed stop-words from the vectorization process by using TfidfVectorizer's inbuilt stop-word dictionary.

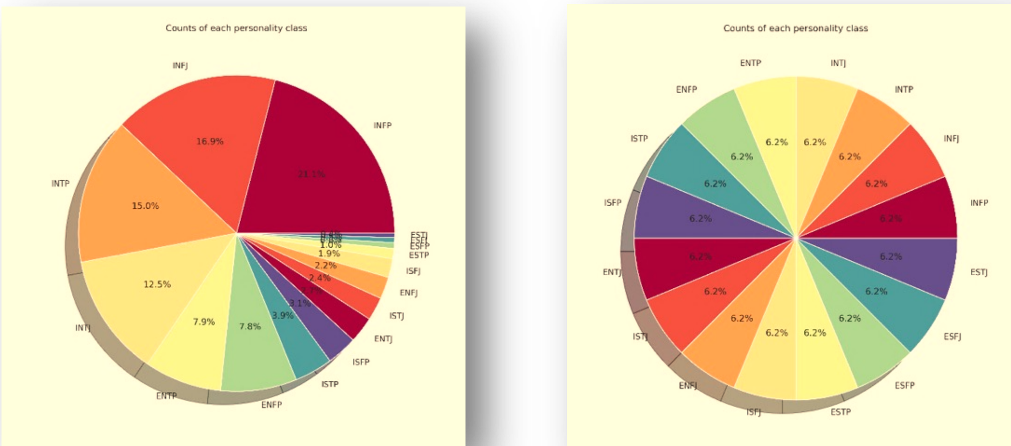


Figure 4: Distribution of classes – Before and After SMOTE techniques

We observed high imbalance in the dataset, which could lead to deteriorated performance on minority classes during model building. So, we set out to oversample the minority classes using Synthetic Minority Oversampling Technique (SMOTE). SMOTE performs data augmentation by synthetically creating data points for minority classes to balance the data in terms of class labels. Internally, it picks an example for a minority class, then another point is randomly selected from the K nearest neighbors of the initial point. Now, it creates a synthetic data point in the feature space between these two points, hence proceeding to balance the minority class. We used the packaged version of SMOT through the imbalanced-learn library [28].

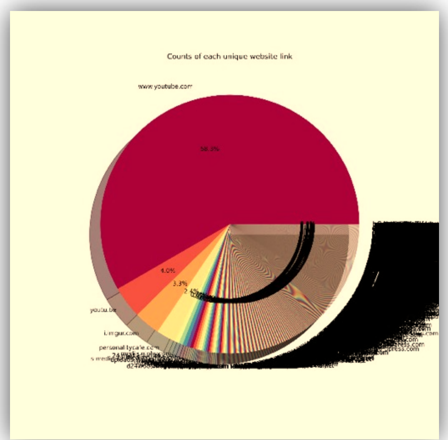
After balancing the classes using data augmentation, we were left with 1831 samples for each of the 16 classes. Now, we had to encode the class labels, as they were in the string format (Ex: 'INTP', 'ENTJ', etc.), we created two types of encoding, one being the standard text-to-numerical encoding which lies between 0 to n_classes-1 and the other being one-hot-encoding of labels which follow the one-of-K encoding scheme. We did this because some model estimators required categorical data in one-hot-encoded format. The encoding tasks were performed using the LabelEncoder and OneHotEncoder objects from Scikit-learn's preprocessing [29] module.

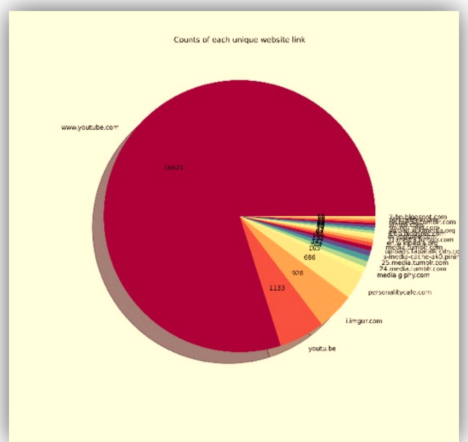
4.2. Exploring hyperlinks present in the text to get text information:

Very early into our analysis, we realized that the text contained a lot of hyperlinks posted by the users, along with some text accompanying it. In earlier approaches, attempts were made to remove the links from the text or in other cases remove the records containing links from the dataset. However, we viewed this differently. In our view, the content that users share in the form of hyperlinks are items which might also be in-line with the user's personality and started trying to figure out how these could be used. We decided to scrape text from these links and add them to the records. In this regard, we analyzed all the different websites that were mentioned in the data.

This analysis was important for us to develop the scraping mechanisms since every website is not structured the same way. The results were favorable, since YouTube was the overwhelming majority with close to 62% of all links in our data being YouTube videos which meant that we could focus our scraping only on YouTube. We decided on scraping the YouTube video titles and appending them to our records, using BeautifulSoup public package. The scraping process in a nutshell consisted of sending a GET request and extracting the title from the returned markup response. We scraped close to 18,000 YouTube video hyperlinks to get data, but also realized that a good 10-20% of these videos did not exist anymore and returned empty pages. Overall, we were able to extract value out of these hyperlinks instead of just discarding them.

Figure 5: Distribution of links from websites





4.3. Exploring various ML algorithms for multi-class classification

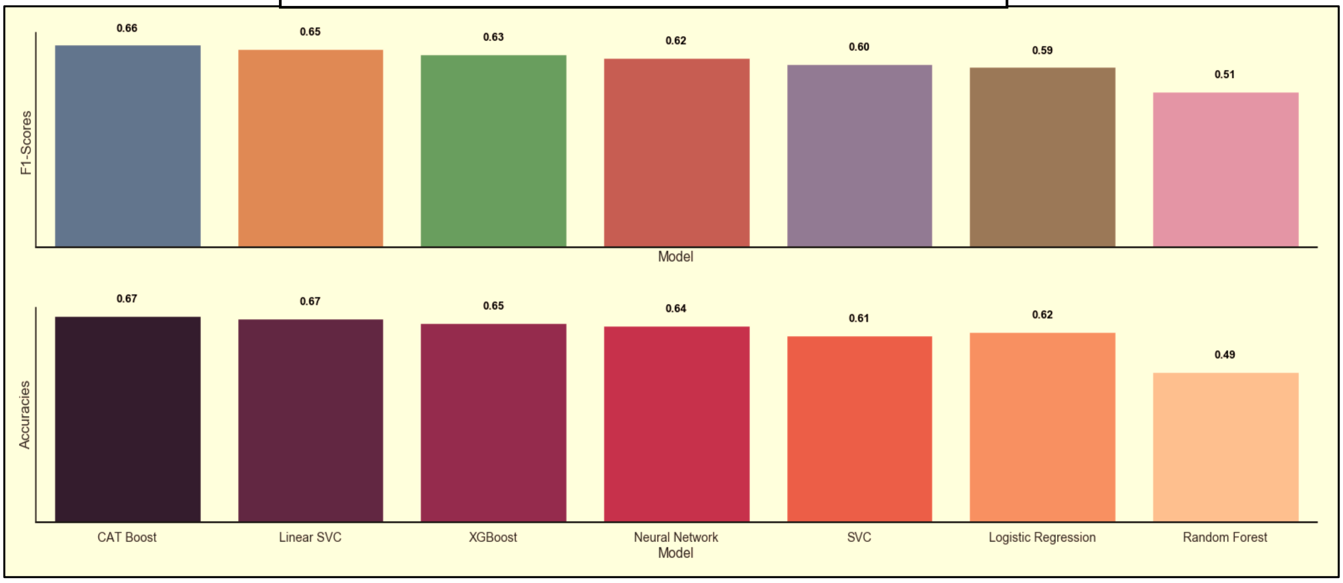
Once we had all the descriptive analysis, scraping and pre-processing done and dusted, we moved to the predictive analysis part of our project – Model building! In this section, we will document the different machine learning algorithms we have explored and will go into detail about the top performing algorithms.

We started with a relatively basic linear classifier algorithm, Logistic Regression. We fit the regressor on our data with a limit of maximum 4000 iterations for convergence, and 0.5 regularization strength and using all CPU cores. The F1 Score and Accuracy averaged at 0.59 and 62 over multiple runs.

Then we tried a random forest classifier which internally fits decision trees, with the maximum depth of every tree at 100 and worked our way up to 500, but the best results we got were – F1-score: 0.51, Accuracy: 0.49
We also tried a Support Vector Machine based Classifier, with no maximum limit for the iterations for convergence. F1-score: 0.60, Accuracy: 0.61

We also built a sequential neural network classifier with 4 hidden layers and nodes increasing at powers of 2, using categorical cross entropy as the loss function optimized with adam. Its output consists of 16 nodes and uses a Softmax activation function. We ran the neural network for 100 epochs with a batch size of 100 but did not see much improvement in the f1-score and AUC towards the last 20-30 epochs. F1-score: 0.62

Figure 6: Different models and F1-Scores and Accuracy



Now, our top-3 performing models:

Extreme Gradient Boosting – XGBoostClassifier: The next model was decision tree ensembles using Gradient Boosting. The Extreme gradient boosting used as a classifier performed better than the random forest models, which are also ensembles of trees. We manually experimented and fixing of the number of estimators (n_estimators), maximum tree depth (max_depth) and learning rate yielded in better performance of the models, while we went with some generally amicable values for sub-sample column and level ratios. The Extreme Gradient Boosting Classifier gave us an F1-score of 0.63.

Linear Support Vector Classifier: While being much faster than the standard SVC which had massive fitting times, the linear SVC also performed better than the standard SVC with default parameters, giving us an F1-score of 0.64 and Accuracy of 0.66

Categorical Boosting – CatBoost: The CatBoost gradient boosting performs best on our data and gives us the best results so far. Although XGBoost and CatBoost have more in common, Categorical Boosting takes up the top spot. It uses minimal variance sampling which performs stochastic gradient boosting in a weighted samples approach and a better-balanced tree growth. Its better performance can be attributed to its expertise with categorical data, using advanced mean encoding. From our end, we went with the 'MultiClass' loss function and evaluation metrics, with no cap on the number of maximum iterations. We however, experimented with the depth of the tree, which is a prominent parameter, between 5-10 with varied results and stuck with 10 finally. Our best performing set of parameters gave us an F1 or 0.66 and Accuracy of 0.67. We have plotted the AUC ROC for our best performing CatBoost Model.

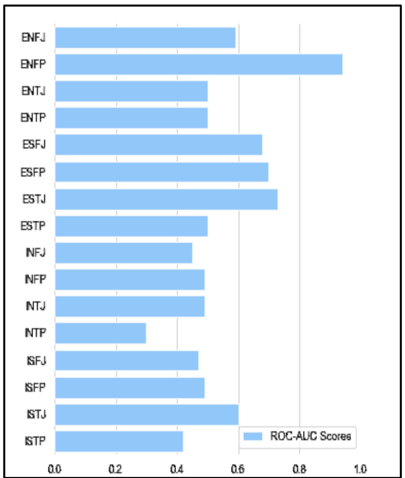


Figure 7: ROC-AUC Scores for each class using CatBoost

Figure 8: Modeling Pipeline in action

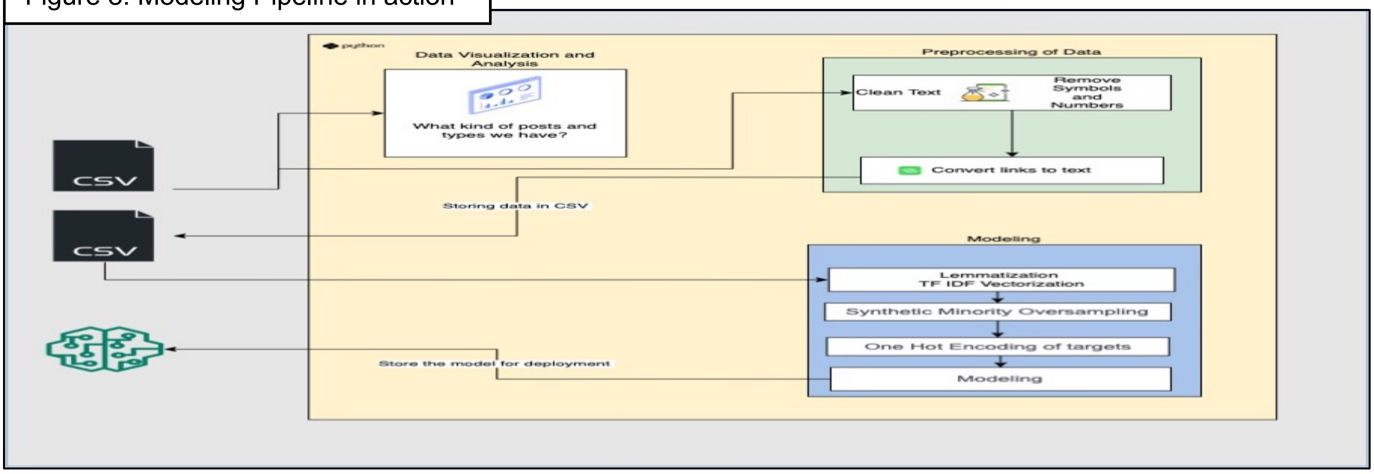


Figure 8: The pipeline image encapsulates the entire modeling process discussed in sections 4.1 to 4.3.

4.4. Deploying ML Model

We wanted to implement a webapp for getting the insight how our prediction model performs on the real-world data. So, we tried train different models varying various hyperparameters and saving them in form of python objects. Using pickle module, we could save them as ".pkl" file and these could be used for deployment. The technique we used here is pickling and unpickling. Pickling is the process of converting a Python object into a byte stream, while unpickling is the process of converting a byte stream (from a binary file or bytes-like object) back into an object hierarchy. [24]

After the models were ready to be tested, we are a script that would use flask as a backend server and have 2 frontend scripts in 2 HTML files. In the "questionnaire.html" file, we'll need to provide a form for the user to submit text or a twitter handle for usage as input to the model to categorize MBTI type. For this, we'll use the POST method. We're gathering all the form values in the /prediction route, as you can see. Because our model requires text input rather than simply the twitter handle, we rearrange the features in the backend script to suit the model. We used Tweepy wrapper for the Twitter API to retrieve the twitter ID of the user, and then used the twitter ID to extract the text from the latest 100 tweets. The text data collected is cleaned, links are handled, vectorized and passed to the model created in the model development. The "model.pkl" pickle file is then loaded and used to estimate the MBTI type based on the input characteristics. The predicted string is returned to the route and placed in the "prediction.html" html file. Following the prediction, the user may view a description of the personality type, its attributes, and their work habits, as well as suggestions for how that individual might be more productive. The page also included information about other MBTI Types as well as a link to a famous questionnaire [25] for manually testing MBTI. If the user gives his or her consent, there is a feedback feature where you can give us a rating from 1 to 5, which is sent to the backend script via the POST method and may be utilized for additional research and data collection to enhance the model. The Figure 10 shows the pipeline that was used, as explained above.

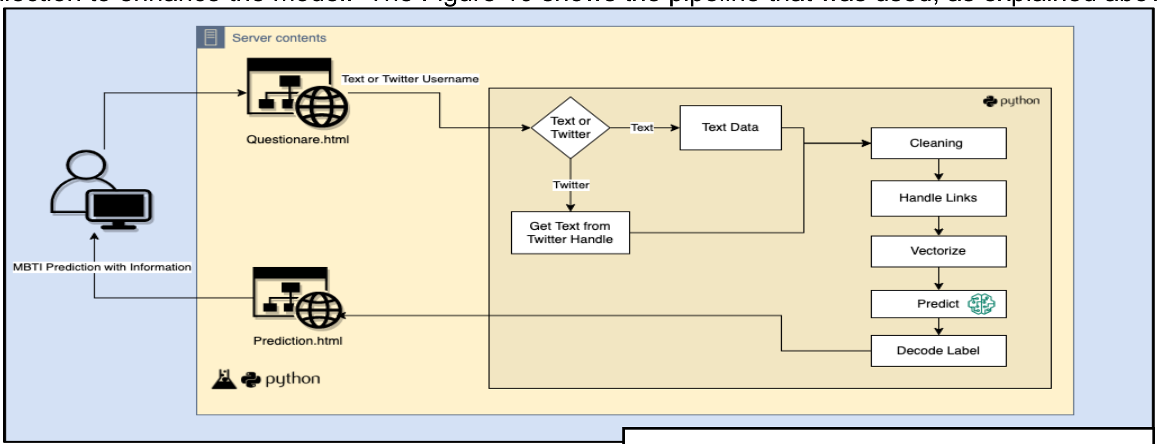


Figure 9: Deployment Pipeline in action

5. Discussion and Results

5.1. Datasets

We are using (MBTI) Myers-Briggs Personality Type Dataset [26]. This data was gathered from the "PersonalityCafe forum" [27], which has a significant number of individuals, their MBTI personality types, and what they have posted. This dataset contains over 8600 rows of data, on each row is a person's Type (This persons 4 letter MBTI code/type) and a section of each of the last 50 things they have posted (Each entry separated by "|||" (3 pipe characters)). The dataset was explored in the in the analysis and preprocessing section (4.1).

5.2. Evaluation Metrics

The evaluation metrics which have been the basis for our model understanding are:

5.2.1. F-1 Score: We always prefer classifiers with high precision and recall score, so F1-Score is a harmonic mean of precision and recall. It helps us bring together two key metrics into a single numerical. [29]

5.2.2. AUC Score: AUC-ROC indicates how well the probabilities of the different classes are separated. We compute the Area Under the Receiver Operating Curve from prediction scores in a one-vs-all context. [29]

5.2.3. Accuracy: The ratio of correctly classified samples by all the samples is the accuracy of a model, it is not usually used in classification due to misleading imbalanced classification issues, however since our dataset is perfectly balanced due to oversampling, the accuracy has swayed in the same tune as the F-1 score, hence making it a metric of consideration. [29]

6. Experimental Results

We have experimented in several aspects of the project, to improve its usability and make it more usable, all of which has been merged into their respective sections. Starting with extracting hyperlinks from data, which was detailed in 4.2. We also worked on improving the models by tuning the parameters by hand, which has been mentioned in the modelling section at 4.3. However, we were unable to automate the parameter searching and selection process due to computational power restrictions, where we intended to use cross validation with searching under constraints.

7. Ethics:

As Dr Isabel Briggs Myers says, "It is up to each person to recognize his or her true preferences."
The solutions should benefit the users of the application and the users, not the organizations, should have control over the application. The user should have complete control over whether his data is shared with the organization.

7.1. Ethical Concerns?

We understand the ethical questions that might arise from the use of technology like this, and we would like to address a few noted concerns:
The usage of machine learning techniques to predict the personality type of an individual is merely an alternative to filling the questionnaire but does not move away from the value of “self-reporting”, this technique must and should only be used with the full consent of the individual who is being assessed. For instance, the employee must be requested for their consent to use his social media data, if the employer plans on assessing the individual’s personality, if not they can simply choose to answer the questionnaire.

In terms of reporting on demographics, neither our dataset, nor other related dataset seem to contain any information on the demographic identities of the surveyed population, which makes it impossible to comment on this aspect of the application.

7.2. What can we do from our end?

After having introspected from our end on the possible privacy breaches that can happen, which violate the ‘self-reporting’ aspect of the MBTI model. A small change to what we showed in the demo would be, that the user will have to login with their social media handle and can run the assessment only on the account they have logged into (i.e., their own account).

8. Conclusion:

In terms of predictive analysis, our final model was able to identify the personality type of the user correctly, more than 67% of the times, out of the available 16 MBTI types. This performance can be further improved with more labeled data from social media platforms and with enough computing infrastructure, the use of Deep Learning methods such as recurrent neural networks, Long Short-Term Memory networks which are more capable of processing and learning from sequential data.
In conclusion, this can be a very useful tool for assessing MBTI personality types without having to spend an hour on a questionnaire, however as pointed out it comes with its set of ethical considerations and must be used carefully.

8.1. How is our approach different from other approaches?

Our application powers the user with the control over their data and seeks their consent before using the data for betterment of the algorithm. We will be using Machine Learning to predict the MBTI type based on tweets from social media accounts from text written by user, or the content shared by the user for effective self-reflection. We're also using a different strategy to be preprocessing the data, relying on the user's video links to figure out what sort of material they're sharing.

8.2. Directions for Future Work

Collection of more labeled data in terms of quantity. More data with information on demographics etc. will help us report any biases. We need to build on the feedback section with the agreement of the users and gather data because there is very little data currently. We can deploy more advanced deep learning algorithms such as recurrent neural networks.

9. References

1. Team, G. E. (2018, December 02). Neuro-Linguistic Programming (NLP). Last Checked December 11, 2021, from https://www.goodtherapy.org/learn-about-therapy/types/neuro-linguistic-programming
2. Insights on building and managing virtual teams. (n.d.). Last Checked December 11, 2021, from http://www.thisisxy.com/blog/insights-on-building-and-managing-virtual-teams.
3. The Myers & Briggs Foundation - C G Jungs Theory, Last Checked December 11, 2021, https://www.myersbriggs.org/my-mbti-personality-type/mbti-basics/c-g-jungs-theory.htm.
4. The Myers & Briggs Foundation - MBTI® Basics, Last Checked December 11, 2021, https://www.myersbriggs.org/my-mbti-personality-type/mbti-basics/.
5. Hoag, John David. NLP Meta Programs, by John David Hoag, 4 Feb. 2008, Last Checked December 11, 2021, http://www.nlpls.com/articles/metaPrograms.php.
6. “Michael Hall About.”, Last Checked December 11, 2021, Michael Hall, https://www.neurosemantics.com/michael-hall/.
7. Hall, L. Michael. Figuring out People: Design Engineering with Meta-Programs. Crown House Pub., 2003.

8. Lakoff, George. “Explaining Embodied Cognition Results.” Wiley Online Library, John Wiley & Sons, Ltd, 7 Sept. 2012, https://onlinelibrary.wiley.com/doi/full/10.1111/j.1756-8765.2012.01222.x. Last Checked December 11, 2021
9. Woodsmall, Wyatt Lee. Metaprograms. Advanced Behavioral Modeling, 1988.
10. “Mbti Metaprograms.” NLP Mentor, Kim, 25 Feb. 2020, https://nlp-mentor.com/mbti-metaprograms/. Last Checked December 11, 2021
11. The Myers & Briggs Foundation - the 16 MBTI® Types, https://www.myersbriggs.org/my-mbti-personality-type/mbti-basics/the-16-mbti-types.htm. Last Checked December 11, 2021
12. Lim, Annabelle G.Y. “The Big Five Personality Traits.” Big Five Personality Traits | Simply Psychology, 15 June 2020, https://www.simplypsychology.org/big-five-personality.html.
13. “What's The Predictive Index (PI) Test?” 123test, https://www.123test.com/assessment-training/predictive-index-test/. Last Checked December 11, 2021
14. “Personality.” American Psychological Association, American Psychological Association, https://www.apa.org/topics/personality. Last Checked December 11, 2021
15. “CognitiveFunctions.” Https://Hy.wikipedia.org/Wiki/%D5%8A%D5%A1%D5%BF%D5%AF%D5%A5%D6%80:CognitiveFunctions.png . Last Checked December 11, 2021
16. Artificial Neural Network for Human Behavior Prediction ... https://grafos.cl/Articulos/Artificial_Neural_Network_for_Human_Behavior_Prediction_through_Handwriting_Analysis.pdf.
17. “Predicting Personality from Twitter.” IEEE Xplore, https://ieeexplore.ieee.org/document/6113107.
18. Komisin, Michael C., and Curry I. Guinn. "Identifying personality types using document classification methods." Twenty-Fifth International FLAIRS Conference. 2012.
19. Wan, Danlin, et al. "Personality prediction based on all characters of user social media information." Chinese national conference on social media processing. Springer, Berlin, Heidelberg, 2014.
20. Personality Prediction of Social Network Users. (n.d.). Retrieved from https://ieeexplore.ieee.org/document/8253041?part=1 Last Checked December 11, 2021
21. Tandera, Tommy, et al. "Personality prediction system from facebook users." Procedia computer science 116 (2017): 604-611
22. Survey Analysis of Machine Learning ... - Stanford University. http://cs229.stanford.edu/proj2017/final-reports/5242471.pdf. Last Checked December 11, 2021
23. “MBTI® Questionnaire.” The Myers & Briggs Foundation - VERSIONS OF THE MBTI® Questionnaire, https://www.myersbriggs.org/using-type-as-a-professional/versions-of-the-mbti-questionnaire/. Last Checked December 11, 2021
24. “Pickle — Python Object Serialization — Python 3.10.1 Documentation.” https://docs.python.org/3/library/pickle.html. Accessed 11 Dec. 2021
25. “‘It's so Incredible to Finally Be Understood.".” 16Personalities, https://www.16personalities.com/. Last Checked December 11, 2021
26. J, Mitchell. “(MBTI) Myers-Briggs Personality Type Dataset.” Kaggle, 22 Sept. 2017, https://www.kaggle.com/datasnaek/mbti-type. Last Checked December 11, 2021
27. “Personality Cafe.” Personality Cafe, https://www.personalitycafe.com/ Last Checked December 11, 2021
28. Scikit-Learn-Contrib. “Scikit-Learn-Contrib/Imbalanced-Learn: A Python Package to Tackle the Curse of Imbalanced Datasets in Machine Learning.” GitHub, https://github.com/scikit-learn-contrib/imbalanced-learn. Last Checked December 11, 2021
29. “Scikit API Reference.” Scikit, https://scikit-learn.org. Last Checked December 11, 2021
30. Sengupta, Anirban. “TF IDF SEO: How Does TF-IDF Work for SEO: Top TF IDF Seo Tools.” Digital Marketing Chef, Publisher Name Digital Marketing Chef Publisher Logo, 22 Feb. 2020, http://www.digitalmarketingchef.org/what-is-tfidf-and-why-does-it-matter-for-seo/. Last Checked December 11, 2021