International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

Personality Prediction based on Myers Briggs type Indicator Using Machine Learning

Prof. Ankita Gandhi[1], Vinay Talaviya[2*], Lakshmikrishnasai.Alle[3], Tanu Yadav[4], Abhishek Yadav[5]

[1]Deputy HOD & Assistant Professor,
Computer Science & Engineering Department,
Parul Institute of Engineering & Technology,
Parul University, Vadodara, India
E-mail: ankita.gandhi@paruluniversity.ac.in

[2*]Student, Computer Science & Engineering Department,
Parul Institute of Engineering & Technology,
Parul University, Vadodara, India
E-mail: vinaytalaviya4955@gmail.com

[3]Student, Computer Science & Engineering Department,
Parul Institute of Engineering & Technology,
Parul University, Vadodara, India
E-mail: lakshmikrishnasai2003@gmail.com

[4]Student, Computer Science & Engineering Department,
Parul Institute of Engineering & Technology,
Parul University, Vadodara, India
E-mail: tanu2681@gmail.com

[5]Student, Computer Science & Engineering Department,
Parul Institute of Engineering & Technology,
Parul University, Vadodara, India
E-mail: a20hekyadav@gmail.com

Abstract—In this study, we leverage a combination of machine learning algorithms, including classification and regression models, along with natural language processing techniques, such as NLP and spacy, to predict user personality types from their social media posts. We focus on utilizing the Myers-Briggs Type Indicator (MBTI) to identify a user's unique personality among sixteen possible types *[1]*. This research aims to establish a correlation between individuals' social media content and their personality traits. Our approach involves extensive preprocessing of textual data, employing techniques like text tokenization, regular expressions, lemmatization, sentiment analysis, and part-of-speech tagging, followed by dimensionality reduction *[2]*. We evaluate several machine learning models, including logistic regression, SVM, Naive Bayes, lasso regression, and random forest classifiers, with logistic regression delivering the most accurate results. We deploy this trained model on a web page connected to a Flask app, allowing users to input a brief description of themselves and receive their predicted personality type. This research explores the intersection of text analysis and personality prediction, shedding light on the hidden dimensions of human personality revealed through digital traces in the age of social media *[4]*.

Keywords-Natural Language Processing (NLP), Textual Data Preprocessing, Tokenization, Personality

I. INTRODUCTION

In recent years, natural language processing and machine learning advancements have made it possible to predict MBTI personality types from textual data, such as social media posts, essays, or emails. This innovative method employs machine learning algorithms to estimate a person's MBTI personality type by analyzing linguistic patterns [12].

The Myers-Briggs Type Indicator (MBTI) is a well-known framework for categorizing individuals into 16 personality types, providing valuable insights into their motivations, behaviors, and communication preferences [9]. In this research, we explore various NLP techniques to understand verbal clues indicative of different personality types, shedding light on how these types influence individual traits and behaviors.

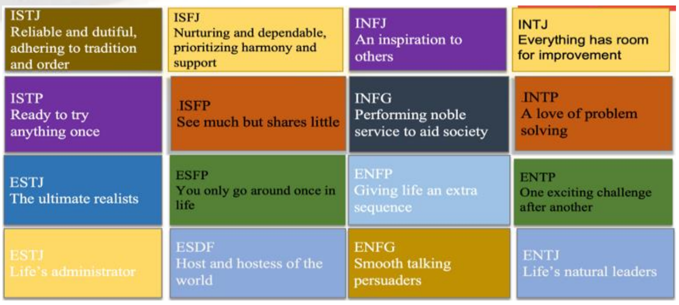

Figure 1: 16 Personality Types

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

II. LITERATURE

The Myers-Briggs type indicator is a well-known personality evaluation tool that identifies individual features and groups them into one of the 16 different personality types. However, it would be much more beneficial to design a system that allows users to forecast their personality type. Knowing someone's MBTI type can be extremely insightful into how they behave, communicate, and make decisions. However, because the MBTI is normally administered using a questionnaire, it takes time. The primary aim of the MBTI personality prediction is to develop a data-driven system that can accurately predict an individual's Myers-Briggs type indicator personality type from the given text by leveraging the most efficient machine learning models and NLP techniques [20]. Our research objective is to create a well-trained machine-learning model that can predict an individual's MBTI type from written text, such as social media posts or textual data sources, using natural language processing (NLP) methods.

III. METHODOLOGY

In this paper, each process is divided into steps which help to maintain the systematic flow of the processes.

A. Data Source

Dataset: The publicly available MBTI dataset provided by the psychological cafe used in this project. This dataset contains 8675 rows and two columns which are posts and type. Type includes the (4 letter MBTI personality type) A section of each of the last 50 things they have posted in which each entry is separated by the given below symbol.
'|||'

B. Preparation of Data

Identifying whether data contains any null or duplicates. Initially whole dataset is imbalanced with raw data which needs to be sorted according to the 4 classes. There we divided 8 dimensions of personality type into 4 classes [7]. These are Extrovert-Introvert, sensing-intuition, Thinking-Feeling, and Judging-perceiving. Summing the count of each type of class.

C. Cleaning the Data

Cleaning of data is a crucial role in which it plays a vital role in deciding the accuracy of the model. For cleaning the raw data we applied the regular expressions including cleaning white spaces, dropping mail, dropping punctuations, and removing the Stop words. Words with one to two character lengths were dropped.

	type	is_extrovert	is_sensing	is_thinking	is_judging	posts	clean_posts
0	INTP	0	0	1	0	'The main reasons why I visit someone's profil...	the main reasons why i visit someone s profil...
1	ISTJ	0	1	1	1	'After reading the op, first Hot For Teacher b...	after reading the op first hot for teacher b...
2	INFJ	0	0	0	1	I got used to not caring about what other peop...	i got used to not caring about what other peop...
3	INFP	0	0	0	0	Something interesting I discovered when I was ...	something interesting i discovered when i was ...
4	ENTP	1	0	1	0	'Eh, I got INFJ and I'm an ENTP. Needs work, b...	eh i got and i m an needs work but i can...

Figure 2: Cleaning the data

D. Tokenization

Applied tokenization to the data which separates the sentence into lists that are considered tokens [11]. We have used 'punkt' which is downloaded by importing the natural language tool kit (NLTK) which is a collection of computer modules to NLP [18]. Included word tokenization and sentence tokenization for converting the sentences into small tokens to identify the patterns within them, and later these tokens are joined to apply the stemming and lemmatization [19].

E. Lemmatization

Lemmatization contains various lramatization techniques, In this project we used a wordnet lemmatization with POS (parts of speech). Wordnet is used to find the semantic relation between the words, here the base wordnet results in missing some base words and has not reached up to mark. So we have used the wordnet lemmatizer with POS in which these words are treated as a noun in the given sentence rather than a verb [15]. Cleaned the data generated after executing the above steps was lemmatized using NLTK word net lemmatizer.

F. Sentiment Analysis

The sentiment intensity analyzer is used to find the intensity score from the sentence which estimates whether the meaning of the sentence is positive, negative, compound, or neutral sentence. The Sentiment Intensity analyzer is provided by the Vader sentiment library [17]. We obtained the intensity score of the sentence which ranges from 0 to 1. Later we applied the perceptron tagger which finds the tagged words contained which are required to be counted separately from the sentence. Parts of speech (POS) are also called grammatical tagging. Here mainly counts the adjectives, adverbs, nouns, verbs, pronouns, and entities from the sentences [5].

	type	is_extrovert	is_sensing	is_thinking	is_judging	posts	clean_posts	compound_sentiment	pos_sentiment	neg_sentiment
0	INTP	0	0	1	0	'The main reasons why I visit someone's profil...	main reason visit someone profile misclic...	0.998449	0.315930	0.234742
1	ISTJ	0	1	1	1	'After reading the op, first Hot For Teacher b...	reading op first hot teacher van halen poppe...	0.998849	0.348059	0.205008
2	INFJ	0	0	0	1	I got used to not caring about what other peop...	got used caring people think friend long time ...	0.999900	0.497992	0.143975
3	INFP	0	0	0	0	Something interesting I discovered when I was ...	something interesting discovered transitioning...	0.999650	0.338688	0.057903

Figure 3: Sentiment Analysis

From the above2 obtained the sentiment analysis of posts which are compound sentiment, positive sentiment as "pos_sentiment", and negative sentiment as "neg_sentiment".

G. Vectorization

The (TF-IDF) Term frequency – Inverse document frequency vectorization and count vectorization are two commonly used techniques for converting text data into numerical representations.

TF-IDF vectorization: In the TF-IDF vectorization, is applied to represent textual features in the dataset related to personality traits. The resulting TF-IDF matrix serves as input features for machine learning models, capturing the importance of words.\\ Term frequency is calculated by dividing the number of times a term occurred in the document by the total number of terms in that document.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

$$\mathrm{TF(t,d)} = \frac{\text{Total number of terms in document}}{\text{Number of times term t appears in document}}$$

Inverse Document Frequency (IDF): Measures the importance of a term in the entire dataset. It is calculated by dividing the total number of documents by the number of documents containing the term and taking the logarithm.

$$\mathrm{IDF(t,D)} = \frac{\text{log (Total number of documents in the dataset } |D|}{\text{Number of documents containing term } t+1}$$

TF-IDF score: It is the product of TF and IDF.

$$\mathrm{TF-IDF(t,d,D) = TF(t,d) \times TDF(t,D)}$$

H. Count vectorization

It is also similar to TF-IDF, as it represents the frequency of a term in a document. Instead of considering the importance of a term in the entire dataset as like TF-ID.

```
count_vectorized_data.head()
```

	ab	abandon	abandoned	abhor	ability	able	abnormal	aboard	abortion	abrasive	...	zelda	zen	zero	zodiac	zombie	zone	zoned	zoning	zoo	zoom
0	0	0	0	0	0	2	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	0	0	0	1	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	1	0	0	0	1	...	0	0	0	0	0	0	0	0	0	0

5 rows × 8636 columns

Figure 4: Count Vectorized

However, We have applied it along with TF-IDF vectorization with the higher priority of TF-IDF because based on requirements the frequency of term plays a major role in this paper, and count vectorization is mostly used for baseline models as it simply counts the occurrences of each term in a document.

posts	clean_posts	compound_sentiment	pos_sentiment	neg_sentiment	...	ADV_avg	CONJ_avg	DET_avg	NOUN_avg	NUM_avg	PRT_avg	PRON_avg	VERB_avg	._avg	X_avg
'The main reasons why I visit someone's profi...	main reason visit someone profile misclic...	0.998449	0.315930	0.234742	...	151.0	214.0	142.0	285.0	15.0	7.0	145.0	352.0	228.0	3.0
'After reading the op, first Hot For Teacher b...	reading op first hot teacher van halen poppe...	0.998849	0.348059	0.205008	...	77.0	140.0	75.0	312.0	17.0	3.0	114.0	193.0	120.0	1.0
I got used to not caring about what other peop...	got used caring people think friend long time ...	0.999900	0.497992	0.143975	...	150.0	204.0	114.0	310.0	15.0	6.0	204.0	325.0	252.0	8.0
Something interesting I discovered when I was ...	something interesting discovered transitioning...	0.999650	0.338688	0.057903	...	130.0	196.0	124.0	361.0	24.0	7.0	180.0	298.0	254.0	0.0
'Eh, I got INFJ and I'm an ENTP. Needs work, b...	eh got need work see would said lol thi...	0.999600	0.409639	0.178404	...	115.0	180.0	90.0	266.0	19.0	12.0	192.0	292.0	212.0	3.0

Figure 5: Average Vectorized

I. Counting

As per user average counts were taken for several question marks, exclamations colons, emojis, words unique words, upper case words, links, ellipses, and images. These counts were our additional features for the machine learning models covered.

posts	clean_posts	compound_sentiment	pos_sentiment	neg_sentiment	...	em	colons	emojis	word_count	unique_words	post_length_var	link_count	upper	ellipses	img_count
'The main reasons why I visit someone's profi...	main reason visit someone profile misclic...	0.998449	0.315930	0.234742	...	0.00	0.16	0.00	29.06	15.24	92.608400	0.06	1.14	0.02	0.00
'After reading the op, first Hot For Teacher b...	reading op first hot teacher van halen poppe...	0.998849	0.348059	0.205008	...	0.02	0.12	0.00	20.00	11.18	169.520818	0.08	1.24	0.04	0.00
I got used to not caring about what other peop...	got used caring people think friend long time ...	0.999900	0.497992	0.143975	...	0.70	0.32	0.12	29.04	14.14	121.891600	0.04	1.68	0.26	0.00
Something interesting I discovered when I was ...	something interesting discovered transitioning...	0.999650	0.338688	0.057903	...	0.04	0.50	0.00	28.30	13.98	124.731600	0.04	1.98	0.06	0.00
'Eh, I got INFJ and	eh got need														

Figure 6: Counting

J. Modelling

We split the data into x and y which are input features and target features. Thereby, x is compromised of clean posts, compound sentiment score, pos tag counts, and various other counts. 'y' was set to four target features which are Extrovert, Sensing, Thinking, Judging [14].

The preprocessing step was designed to vectorize the clean posts to select k best features out of the other features using a column transformer.

Since the data was imbalanced imbalance learning pipeline was used to create the models. The pipeline was composed of preprocessing, random under sampler, and the classification model [8].

K. TF-IDF Logistic Regression

The obtained vectorized data of TF-IDF was split into train and test data with 0.70 as the size of trained data and 0.30 test data. Importing the logistic regression and fitting it with the data in the pipeline [16]. Creating the TF-IDF logistic regression model involves using the TF-IDF vectorization technique to represent text data and then training a logistic regression classifier on the transformed features.

L. Count Vectorized Logistic Regression

The count vectorizer was applied with the logistic regression by adjusting parameters to control the dimensionality of the count vectors.

M. TF-IDF Logistic Lasso

The regularization method of lasso regression was taken from the model selection of sklearn, in which we applied the TF-IDF logistic lasso regression with pipelining of preprocessor, Random under sampler, and Logistic regression CV. The acquired ROC-AUC score for Extrovert vs. Introvert is 0.76, and Sensing vs. Intuition is 0.73, Thinking vs. Feeling score is 0.86, Judging vs. Perceiving is 0.66.

N. TF-IDF Random Forest

The TF-IDF vectorizer random forest model is applied to compare with the previous models applied, in which the parameters pipelined are Preprocessor tf, Dense Transformer, Random forest classifier with number of estimators, and max depth. The acquired ROC-AUC score for Extrovert vs. Introvert is 0.72, and Sensing vs. Intuition is 0.68, Thinking vs. Feeling score is 0.82, Judging vs. Perceiving is 0.59.

Count Vectorized SVM: The count vectorized SVM model we applied with the following parameters which are pipelined preprocessor method, Dense Transformer, Random under

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

sampler, svc with linear kernel. The acquired ROC-AUC score for Extrovert vs. Introvert is 0.72, and Sensing vs. Intuition is 0.75, Thinking vs. Feeling score is 0.71, Judging vs. Perceiving is 0.67.

- **Evaluation metrics** – Accuracy, precision, recall, ROC-AUC, and average precision Recall Score were used as the measures to evaluate the models and select the one performing best.
- **Final model:** Based on the scores for the evaluation metrics used, the TF-IDF logistic regression model was selected as the final model for predicting MBTI type.

The selected model with appropriate accuracy is stored in the joblib file separately having 4 classes of model. The accuracy obtained combined for extroverts and Introverts was 68.45 percentage, sensing and intuition got 60 percentage, Thinking, and feeling accuracy was 70.98 percentage, and Judging and perceiving accuracy was 65 percentage.

O. Framework

A. Webpage

The web pages are designed with respect to making proper interaction of the developed system with the user dynamically. Initially index page contains a simple introduction to the personality test. The index page contains the text box that provides the option user to enter the text form data which can be brief of character or related to any social media post. once the data, is directed to the response page, it displays the result of the given input although it provides the reference link for the particular predicted type output, and the user can provide the data multiple times with a minimum of 50 words. systems Once the user is satisfied with the accurate output, the user can check their type which contains how many people all over the world relate to his personality, these all things are provided on the analysis page with dynamic usage of Power BI. This type of interaction is helpful for the user to check and compare with the other personality types people with the ratio which we have provided in the form of bar charts, pie charts, and count plots. Another one is the methodology page, which contains a complete brief of how the project works with the complete processing steps from the initial to the final result. The methodology page helps us to improve our performance with the user feedback and also proof of working system which includes the starting point of a dataset to the final prediction of result. The main theme is to present the model to the users with appropriate results in which the user input and response page plays a key role.

B. Flask App

Flask allows us to render HTML templates dynamically by embedding the Python code within the HTML files. Flask helps us to make the dynamic use of HTML templates with data from our backend logic and also the separation of concern between the frontend and backend has become essential for building interactive web pages. We have defined the routes using decorators such as "@app.route('/index')", and when the user accesses the specified URL, flask calls the associated view function. Although used the ngrok for the local development of server to the internet. Postman is to interact for testing with HTTP requests. The Flask app provided the routes that accept GET and POST method requests from the portman. We have used Postman to send HTTP requests to our Flask app routes. When Postman sends a request to Flask, flask processes the request and performs the actions on the response page which obtains the result.

IV. RESEARCH FLOW

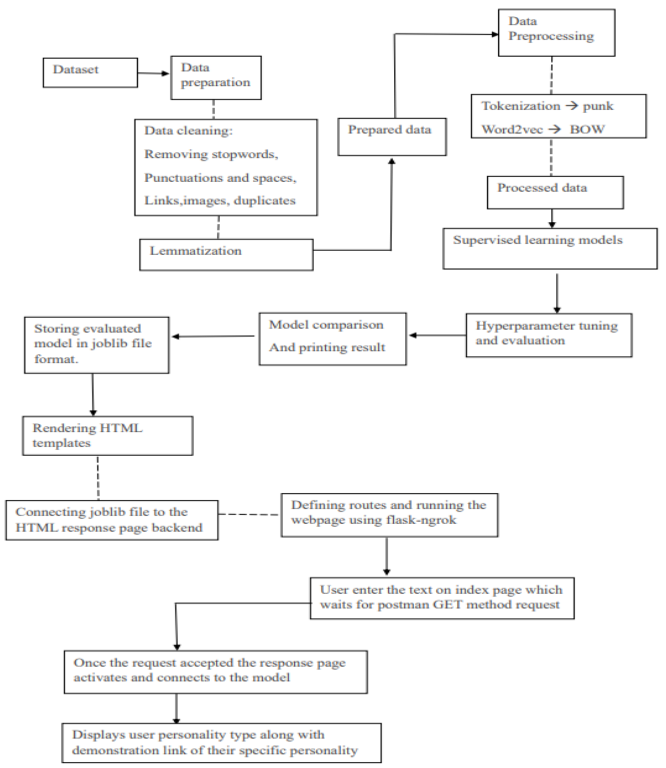

Figure 7: Research Flow

A. Flow Chart

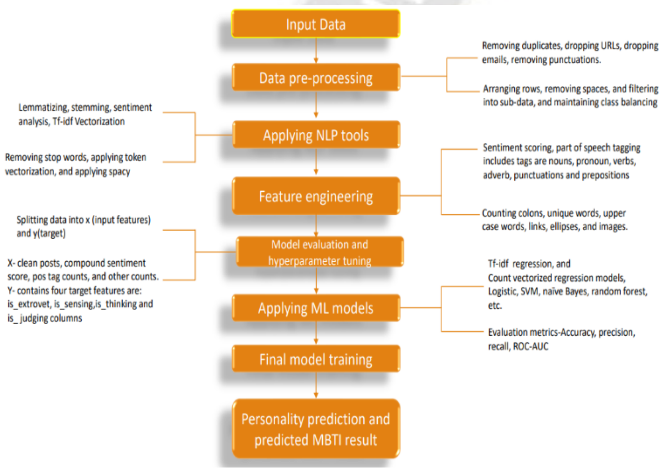

Figure 8: Flow Chart

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

V. RESULTS AND FINDINGS

A. Data analysis and visualizations

The MBTI personality dataset, which has 8675 rows and 2 columns, is the source of the data. Posts and type columns representing each MBTI type are included. The 16 different personality types identified by the MBTI are based on four axes: thinking (T) and feeling (F), judging (J) and perceiving (P), introversion (I) and extroversion (E), intuition (N) and sensing (S). The dataset contains no null or missing values. Since INFPs have the highest frequency, they will have a lot more data, but ESTJs have the lowest frequency, hence they will have the least quantity of data overall. The dataset does not contain any duplicate posts. There is an imbalance in the dataset across the various classifications.

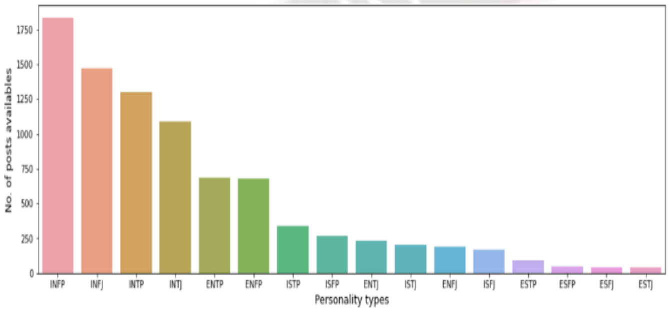

Figure 9: Personality Types

From the below visualization, the data is visualized from the dataset where most of the people in the social media are Introverts (INFP) with the highest count.

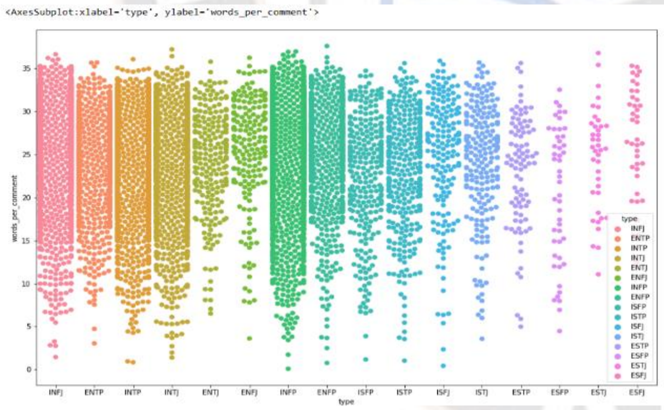

Figure 10: words per comment

	is_Extrovert	is_Sensing	is_Thinking	is_Judging
is_Extrovert	1.0000	-0.0458	0.0697	-0.1614
is_Sensing	-0.0458	1.0000	0.0814	-0.0146
is_Thinking	0.0697	0.0814	1.0000	-0.0046
is_Judging	-0.1614	-0.0146	-0.0046	1.0000

Figure 11: Confusion Matrix

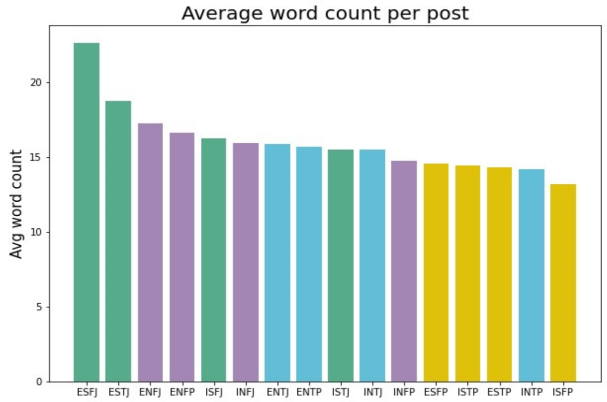

Figure 12: Average Word Per Count

	type	posts	result
0	INTP	'1. Yes, the Feminist movement combined with t...	ENTJ
1	INFP	'Congrats! :)|||I have two younger sisters and...	ISFP
2	ISFJ	'Ill be twenty nine in two weeks.|||I wouldn't...	ESTP
3	ENTP	'Well...That all depends on the question: What...	ESTJ
4	INFP	'No, I said its Fe thing to do would be to fol...	ESFJ
...	...	...	...
2164	INFJ	'Thank you. There's no way he's an INFP. Phil ...	ESTJ
2165	INFP	'MBTI Type- infp Enneagram/Socionics/other (o...	ISTP
2166	INFP	'It's all about being humble, having no ego. B...	ISFP
2167	INTJ	'Friend: So, you haven't been to the new offic...	ENTJ
2168	ENTP	'Learning to drive oscillated between my mom h...	ENTP

2169 rows × 3 columns

Figure 13: Results Output

The outcomes of this research aim to present a machine-learning model capable of predicting MBTI personality types with accuracy and interpretability. Insights into the linguistic features influencing these predictions are revealed, offering a deeper understanding of the relationship between language and personality [10].

TABLE I. ACCURACY SCORE

	Accuracy of Logistic regression	**Accuracy of SVM**	**Accuracy of Random-forest classifier**	**Accuracy of Multinomial Bayes**
Extroversion-Introversion	0.79	0.72	0.76	0.76
Sensing-Intuition	0.76	0.75	0.85	0.73
Feeling-Thinking	0.88	0.71	0.78	0.86
Judging-Perceiving	0.70	0.67	0.63	0.66

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

Based on the models mentioned above, logistic regression performs well on the data and produces accurate results.

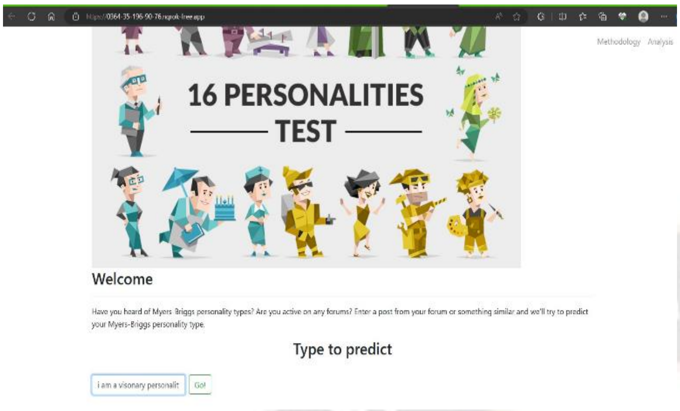

Figure 14: Output 1

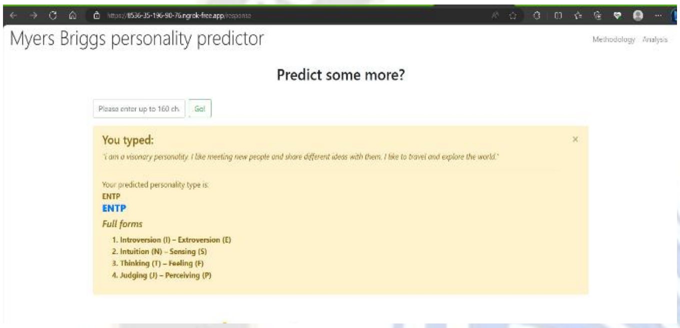

Figure 15: Output 2

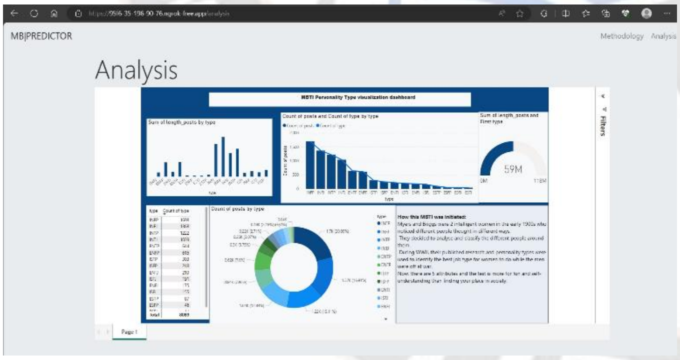

Figure 16: Output 3

VI. Discussion

The significance of this research extends beyond the realm of personality prediction. It advances the fields of personality psychology and NLP, shedding light on the intricate connection between language use and personality traits. Furthermore, the creation of an intuitive application or user interface for MBTI personality predictions serves to democratize and popularize the use of these insights, enabling users to gain a better understanding of themselves and others.

VII. Conclusion

In conclusion, the MBTI Personality Prediction is a testament to the effectiveness of data-driven innovation and insights in the field of personality research. We have employed cutting-edge methods throughout this project to achieve a wide range of objectives, resulting in an impressive fusion of technology and psychology [6]. Our initial goal was to automate personality assessment, freeing ourselves from the constraints of self-reported tests and leveraging the wealth of information embedded in textual data.

Our highly sophisticated algorithm can accurately determine a user's Myers-Briggs Type Indicator (MBTI) personality type based on their spoken and written natural language expressions. The interplay between Machine Learning (ML) and Natural Language Processing (NLP) has been the cornerstone of our endeavor [13]. We have harnessed the ability to extract subtle linguistic patterns, emotions, and behavioral traits from text through the harmonious integration of these two domains, offering users a unique insight into their personalities.

Our user-friendly Flask-based web application allows individuals to effortlessly submit their textual data and receive real-time predictions regarding their MBTI personality type. By democratizing personality testing, this interactive tool empowers users to gain insightful knowledge about their psychological profiles.

Ultimately, the MBTI personality prediction model serves as a prime example of how technology, psychology, and user-centric design can coexist successfully [3]. It exemplifies the countless possibilities that can be realized when data-driven innovation meets the intricacies of human behavior. With each prediction generated, we advance our understanding of personality, laying the foundation for a future where self-discovery is as simple as pressing a few keys.

VIII. Future Work

The current focus of the project is on personality prediction using text-based data. Future research endeavors might explore the incorporation of additional modalities, such as images, audio, and video, to obtain a more comprehensive understanding of an individual's personality. Maintaining and updating NLP models and procedures will result in even more accurate future predictions. Integrating sentiment analysis and emotion detection can add depth to personality predictions, encompassing not just personality traits but also emotional states.

Expanding the project to create user profiles over time could provide insight into how an individual's personality evolves. Users who undergo this kind of analysis may receive valuable feedback on the changes in their personalities.

Exploring the variation in personality predictions across different cultures and languages can offer insights into cultural differences in communication styles and behavior.

The ultimate goal for future work is to transition from a research project to real-world applications such as mental health monitoring, HR assessments, or personalized education. This transition represents a significant step forward, demonstrating the practical and meaningful implications of our research.

ACKNOWLEDGMENT

This work was supported by **Assistant Professor Ankita Gandhi** for her invaluable suggestions, scholarly guidance, constructive feedback, inspiration, and practical approach. Her unwavering support provided us with the motivation we needed to work diligently.

REFERENCES

[1] N. a. S. J. Ahmad, "Personality assessment using Twitter tweets," *Procedia computer science,* vol. 112, pp. 1964--1973, 2017.

[2] F. a. S. E. A. a. R. G. Alam, "Personality traits recognition on social network-facebook," *Proceedings of the international AAAI conference on web and social media,* vol. 7, pp. 6--9, 2013.

[3] C. Basto, "Extending the abstraction of personality types based on MBTI with machine learning and natural language processing," *arXiv preprint arXiv:2105.11798,* 2021.

[4] S. a. O. T. M. a. G. F. M. a. H. D. a. S. T. Breuer, "Aviation and personality: Do measures of personality predict pilot training success? Updated meta-analyses," *Personality and Individual Differences,* vol. 202, p. 111918, 2023.

[5] P. a. M. S. a. K. P. Dandannavar, "Social media text-a source for personality prediction," *2018 international conference on computational techniques, electronics and mechanical systems (CTEMS),* no. IEEE, pp. 62--65, 2018.

[6] K. a. E.-K. R. A. a. S. M. A. I. a. A. S. El-Demerdash, "Deep learning based fusion strategies for personality prediction," *Egyptian Informatics Journal,* vol. 23, pp. 47--53, 2022.

[7] I. a. I. Z. a. Y. {. Ergu, "Predicting personality with twitter data and machine learning models," *2019 innovations in intelligent systems and applications conference (ASYU),* no. IEEE, pp. 1--5, 2019.

[8] N. a. G. S. a. A. J. a. K. Z. H. a. M. G. Fatima, "A rule-based machine learning model for career selection through MBTI personality," *Mehran University Research Journal Of Engineering \& Technology,* vol. 41, no. Mehran University of Engineering \& Technology Jamshoro, Pakistan, pp. 185--196, 2022.

[9] S. J. a. M. M. R. a. o. Johnson, "Machine Learning Approach to Improve Data Connectivity in Text-based Personality Prediction using Multiple Data Sources Mapping," *Journal of Scientific \& Industrial Research,* vol. 82, pp. 109--119, 2023.

[10] D. a. S. P. a. G. D. a. K. K. a. D. I. a. V. A. Karanatsiou, "My tweets bring all the traits to the yard: Predicting personality and relational traits in Online Social Networks," *ACM Transactions on the Web (TWEB),* vol. 16, no. ACM New York, NY, pp. 1--26, 2022.

[11] M. a. R. H. U. a. S. A. J. a. S. N. Karnakar, "Applicant personality prediction system using machine learning," *2021 2nd global conference for advancement in technology (GCAT),* no. IEEE, pp. 1--4, 2021.

[12] K. a. S. M. a. R. R. Karunarathna, "Ensemble Learning Approach for Identifying Personality Traits based on Individuals' Behavior," *EDITORIAL NOTE,* p. 107, 2023.

[13] R. a. S. J. a. M. F. S. a. o. Kurniawan, "A Framework for Determining the Big Five Personality Traits Using Machine Learning Classification through Graphology," *Journal of Electrical and Computer Engineering,* vol. 2023, no. Hindawi, 2023.

[14] V. a. B. N. a. S. H. A. Lynn, "Hierarchical modeling for user personality prediction: The role of message-level attention," *Proceedings of the 58th annual meeting of the association for computational linguistics,* pp. 5306--5316, 2020.

[15] W. a. E. V. Maharani, "Big five personality prediction based in Indonesian tweets using machine learning methods.," *International Journal of Electrical \& Computer Engineering (2088-8708),* vol. 12, 2022.

[16] R. a. B. K. a. R. R. a. K. R. Masika, "Classification Algorithm for Career Recommendation System".

[17] A. S. a. K. B. S. a. R. R. a. C. S. a. S. A. a. P. R. K. K. Rao, "Use of artificial neural network in developing a personality prediction model for career guidance: A boon for career counselors," *International Journal of Control and Automation,* vol. 13, pp. 391--400, 2020.

[18] G. a. B. K. R. a. H. R. a. P. B. a. K. M. M. Rohit, "Machine intelligence based personality prediction using social profile data," *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT),* no. IEEE, pp. 1003--1008, 2020.

[19] G. a. S. K. a. N. D. a. S. S. a. o. Sudha, "Personality prediction through CV analysis using machine learning algorithms for automated e-recruitment process," *2021 4th international conference on computing and communications technologies (ICCCT),* no. IEEE, pp. 617--622, 2021.

[20] M. M. a. L. H. a. X. B. a. Y. L. Tadesse, "Personality predictions based on user behavior on the facebook social media platform," *IEEE Access,* vol. 6, no. IEEE, pp. 61959--61969, 2018.