

# Personality Prediction Using Social Media Platform

Swapnali Shipankar<sup>1</sup>, Gayatri Sawale<sup>2</sup>, Rajashree Shelke<sup>3</sup>, Prof. Ashwini Khairkar<sup>4</sup>

<sup>123</sup>Student, Department of Information Technology, Bharati Vidyapeeth's College of Engineering for Women, Pune, Maharashtra, INDIA

<sup>4</sup>Professor, Dept. of Information Technology Engineering, Bharati Vidyapeeth's College of Engineering for Women, Pune, Maharashtra, INDIA

\*\*\*

**Abstract** - Personality is an critical aspect that influences someone's opinions, like-dislike, thoughts, and the way someone behaves in exceptional situations. Nowadays social media is a platform wherein human beings explicit their view, how they feel, what they're doing. Every 2nd on average, round 6000 tweets are generated. This records may be analyzed nicely to are expecting the character of the consumer. In this project, we'll display you a way to use gadget getting to know to examine someone's character primarily based totally on their real-time Twitter feed. Specifically, we'll first use the Twitter API you purchased our enter records earlier than going for walks a naive Bayes classifier to educate and check our version, so as to then be capable of categorize new Twitter profiles stay right into a Myers-Briggs Type Indicator (MBTI). We're already acquainted with the famous metric, which makes use of a four-letter end result (consisting of INFJ or ENFP) to summarize exceptional character traits in phrases of ways people understand the sector and make decisions. Normally, that is derived via questionnaires and psychometric checks administered to every person, however right here we'll robotically get a end result at the pressing of a button. Models anticipated every character trait with the accuracy 76.231%.

**Key Words:** ML, Pkl, HTML, CSS, Flask.

## 1. INTRODUCTION

With over 3.5 million tests performed every year, MBTI is the maximum broadly used character indicator globally. The Myers Briggs Type Indicator (MBTI) is a character kind device that divides absolutely each person into sixteen awesome personalities primarily based totally on 4 dimensions, namely: Introversion (I) — Extroversion (E), Intuition (N) — Sensing (S), Thinking (T) — Feeling (F), Judging (J) — Perceiving (P). MBTI expected character tendencies preserve important houses of the conventional character traits. Researchers broadly use gadget getting to know and deep getting to know algorithms to are expecting character and mental tendencies from virtual records. We're growing an MBTI character classifier that makes use of gadget getting to know fashions to are expecting someone's character primarily based totally at the 50 latest social media posts in line with consumer as enter. We locate correlations among someone's MBTI character kind and writing style. The classifier additionally demonstrates the validity of the MBTI check. We have used a respectable

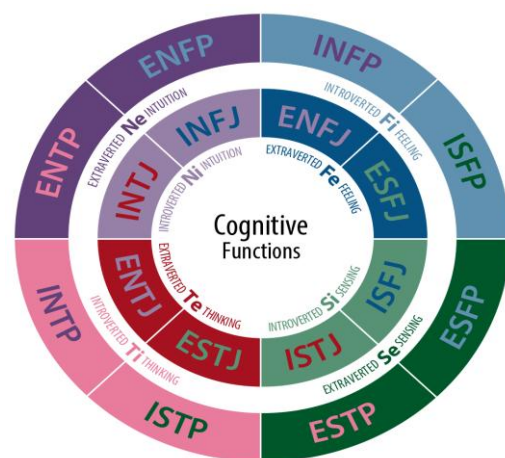
quantity of mined character annotated records from social media. Furthermore, our version might run on greater records than that furnished in a traditional character check, which serves as a affirmation device and facilitates human beings depend greater at the results.

### 1.1 Web scraping User's data and Preprocessing:

Twitter permits builders to scrap publicly to be had records via a fixed of personal keys i.e., Twitter API keys. These keys permit developer to request records on behalf of App's owner. Tweepy is a Python library for gaining access to the Twitter API. The closing 50 tweets of consumer are extracted, after which preprocessed with the assist of regex.[2]

### 1.2 Myers- Briggs Type indicator (MBTI):

The MBTI is one of the maximum famous character checks within side the world. The Myers Briggs Type Indicator (or MBTI for short) is a character kind device that divides absolutely each person into sixteen awesome character kinds throughout four axes: Introversion (I) – Extroversion (E); Intuition (N) – Sensing (S); Thinking (T) – Feeling (F); Judging (J) – Perceiving (P). The processed records could be analysed via every of character kind indicator and maximum favorable out of sixteen character kind could be allotted to the consumer.[2]



### 1.3 GOALS AND OBJECTIVES

**AIM:** The purpose is to categorise the character of a consumer from the enter textual content through making use of gadget studying algorithms specifically Naïve Bayes the usage of MBTI version.

**GOAL:** The typical aim is to create the internet UI if you want to expect the character of the consumer on the premise of the given enter.

**OBJECTIVES:**

- To construct the version with most accuracy the usage of the proper Machine studying algorithms.
- To store the very last constructed version.
- Design internet pages the usage of HTML and CSS
- Expose the internet pages the usage of Flask framework

### 1.4 MOTIVATION

In today’s era, character is one of the closely researched and charming subjects in psychology. Personality popularity of customers is broadly utilized in studies domain names like advice structures and human-robotic interaction. Traditional advice structures stumble upon troubles like loss of statistics approximately the alternatives of the consumer, free-riders problem, and the statistics sparsity problem. The recognized consumer character tendencies assist apprehend customers’ alternatives. MBTI’s test-retest reliability hovers round a 0.5 mistakess rate. On retest, human beings pop out with 3–four kind alternatives 75–90% of the time. Our method can help with extra accuracy than presently current tests, permitting customers to depend upon their outcomes. Personality class primarily based totally on virtual statistics has proved to be a extra green opportunity to conventional mental tests.

### 2. LITERATURE SURVEY

Randall Wald [1] said that Twitter, a micro blogging site, is used by it number of users to share their experiences and thoughts about their day-to-day life. Although researchers have often discarded the method of predicting personality by analyzing the tweets because there are of the view that it contains very little content to predict significant information, but these tweets can be confirmed to make a larger picture of the user who is posting them, select RUSBoost a new form of ensemble learning has been used t predict psychopath using twitter, which uses four classification learners and four feature selection techniques.

J. Golback [2] said that Social Network is a platform where the users tend to reveal themselves to the outside world, sharing their personal details and giving insight to other people into their lives. Personality plays an important role in many types of interaction among people, it can be used to predict the job satisfaction. Until now, in order to accurately

predict user’s personality, they surveyed among individuals with the help of personality test. However, the was highly unfeasible while collecting data from social media platforms and hence correct analysis of personality was problem.

Bayu Yudha Pratama[3] Social Media is place where uses express themselves to the world. Posts or comments made by users on social media can be analyzed to obtain their personal information. It uses text classification to predict personality based on text written by twitter users. The languages used are English and Indonesian. Classification methods implemented are Naïve Bayes, K-Nearest (KNN), SVM. Testing results showed Naïve Bayes slightly outperformed the other methods.

### 3. SYSTEM ARCHITECTURE

Paper Title	Year	Seed Idea
Facebook Profiles Reflect Actual Personality, Not Self-Idealization	2010	-People express their original thoughts on social media instead of an idealized version of themselves
Personality traits recognition on social network—Facebook	2013	-Correlation between user’s personality and social media is strong
The development and psychometric propertiesof LIWC2015	2015	-the work of personality extraction from the text.
Neural Networks in Predicting Myers Brigg Personality Type from Writing Style	2017	-LSTM with other model gives more accuracy. LSTM haspotential of More information to beencoded in parameters

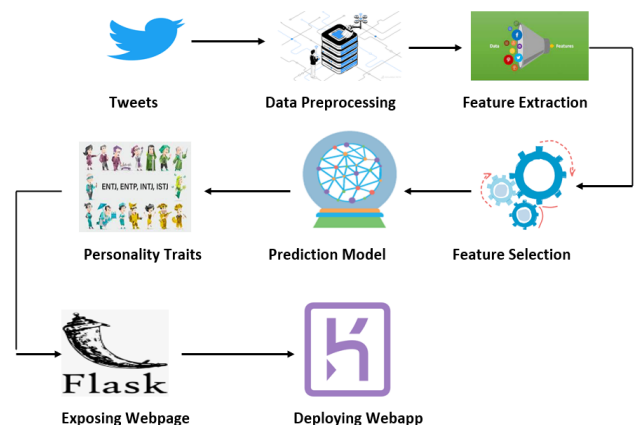


Chart 1. System Architecture

#### 4. PERSONALITY CLASSIFICATION

To set up the personality type of a user, we use Myers-Briggs. The MBTI is the maximum usually used for man or woman categorization, which has been utilized in numerous departments to research the individual's persona. MBTI classifies the man or woman primarily based totally on four-class, namely:

- Introversion (I) or Extraversion (E)
- Intuition (N) or Sensing (S)
- Feeling (F) or Thinking (T)
- Perceiving (P) or Judging (J)

Introversion (I) or Extraversion (E):

People and Things within the outside world is mainly centered with the aid of using an Extraverts, while an introverts look into internal perspectives and ideas.

Intuition (N) or Sensing (S):

Sensing refers to understand items through vision, noise, flavor, tap, and smell, even as instinct refers to presiding kind which include searching into the preceding enjoy and additionally pay attention greater on their wondering.

Feeling (F) or Thinking (T):

The wondering and feeling come into the class called judgment where in we will decide people. Thinking uses experience to determine the society, wherein emotions lead us to outlook items base on what sentiment they evoke.

Perceiving (P) or Judging (J):

Every individual evaluates and figure, however individual who're judging presiding are regarded to be nicely ordered and outcome-driven, wherein figure presiding are bendy and multitasking personality.

Type	Expansion
ISTJ	(Introverted, Sensing, Thinking, Judging)
ISFJ	(Introverted, Sensing, Feeling, Judging)
INFJ	(Introverted, Intuitive, Feeling, Judging)
INTJ	(Introverted, Intuitive, Thinking, Judging)
ISTP	(Introverted, Sensing, Thinking, Perceiving)
ISFP	(Introverted, Sensing, Feeling, Perceiving)
INFP	(Introverted, Intuitive, Feeling, Perceiving)
INTP	(Introverted, Sensing, Thinking, Perceiving)
ESTP	(Extraverted, Sensing, Thinking, Perceiving)
ESFP	(Extraverted, Sensing, Feeling, Perceiving)
ENFP	(Extraverted, Intuitive, Feeling, Perceiving)
ENTP	(Extraverted, Intuitive, Thinking, Perceiving)
ESTJ	(Extraverted, Sensing, Thinking, Judging)
ESFJ	(Extraverted, Sensing, Feeling, Judging)
ENFJ	(Extraverted, Intuitive, Feeling, Judging)
ENTJ	(Extraverted, Intuitive, Thinking, Judging)

#### 5. ALGORITHMS

##### Step 1: Create Twitter Developer Account

To predict the Personality, we have to login into the Twitter Developer Account to fetch the Tweets.

In this project, we have taken real time Twitter data and based on that we have predicted the personality of an individual person.

##### Step 2: Data Preprocessing

Pre-Processing of data is done in order to remove non-words, punctuations, repeated letters and stop-words, so as to clean the data in order to increase. Pre-processing consists of selective word removal, lemmatization and conversion of MBTI 4 letter code to binary. In Selective Word Removal certain words which might cause the machine to cheat. Lemmatization brings all the words to their root form. For example, running, playing and thinking becomes run, play and think respectively.[4] Converting the MBTI personality to binary consists of assigning each aspect with 0 or 1 so that it can be easily interpreted by the model. Here I, N, F, J are all assigned with 0, whereas E, S, T, P are all assigned with 1. So, the personality type ENFP is converted to [1,0,0,1].

##### Step 3: Feature Extraction

Feature Extraction goals to lessen the wide variety of functions or features in a dataset through developing new functions from the prevailing ones (after which discarding the unique functions). These new decreased set of functions

must then be capable of summarize maximum of the statistics contained withinside the unique set of functions.

#### Step 4: Feature Selection

Feature selection is the system of decreasing the wide variety of input variables while making a predictive model. It is ideal to lessen the wide variety of input variables to each lessen the computational value of modeling and, in a few cases, to enhance the overall performance of the model.

#### Step 5: Predicting Model

In this paper, supervised learning approach is used for personality prediction. The model will take snippet of post or text as an input and will predict and produce personality trait (I-E, N-S, T-F, J-P) according to the scanned text. Mayer's Briggs Type Indicator is used for classification and prediction [5]. This model categorize an individual into 16 different personality types based on four dimensions, namely, (i) Attitude → Extroversion vs Introversion: this dimension defines that how an individual focuses their energy and attention, whether get motivated externally from other people's judgement and perception, or motivated by their inner thoughts, (ii) Information → Sensing vs intuition (S/N): this aspect illustrates that how people perceive information and observant(S), relying on their five senses and solid observation, while intuitive type individuals prefer creativity over constancy and believe in their guts, (iii) Decision → Thinking vs Feeling(T/F): a person with Thinking aspect, always exhibit logical behavior in their decisions, while feeling individuals are empathic and give priority to emotions over logic, (iv) Tactics → Judging vs Perceiving (J/P): this dichotomy describes an individual approach towards work, decision-making and planning. Judging ones are highly organized in their thoughts.[6] They prefer planning over spontaneity. Perceiving individuals have spontaneous and instinctive nature. They keep all their options open and good at improvising opportunities.

## 6. CONCLUSION

In this paper, we have shown that a users' MBTI personality traits can be predicted from the public information they share on Twitter. Our subjects completed a personality test and through the Twitter API, we collected publicly accessible information from their profiles. After processing this data, we found many small correlations in the data. While compare to other algorithms that are used previously like naive Bayes. The results show that the methodology presented in this research has better accuracy and reliability in comparison to other existing methods. Regarding the knowledge contribution in this paper, the presented methodology significantly improved the accuracy of recognizing the Intuition (I)-Sensing (S) and Introversion (I)-Extroversion (E) personality categories, as well as slightly better accuracy for the Judging (J)-Perceiving (P) personality category.

## REFERENCES

- [1] R. Wald, T. M. Khoshgoftaar, A. Napolitano Using Twitter Content to Predict Psychopathy R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.
- [2] Golbeck, J., Robles, C., and Turner, K. 2011a. Predicting Personality with social media. In Proc of the 2011 annual conference extended abstracts on Human factors in computing systems.
- [3] Bayu Yudha Pratama, Riyanarto Sarno "Personality Classification based on twitter text," International Conference on Data and Software Testing, 2015.
- [4] M. H. Amirhosseini and H. Kazemian, "Machine Learning Approach to Personality Type Prediction Based on the Mayers-Briggs Type Indicator".
- [5] A. S. Khan, H. Ahmad, M.Z. Asghar, "Personality Classification from Online Text using Machine Learning Approach".
- [6] Tommy Tandra, Hendro, Derwin Suhartono, "Personality Prediction System from Facebook Users".

## BIOGRAPHIES



Ms. Swapnali Shipankar is pursuing B.E(IT) from Bharati Vidyapeeth's College of Engineering for Women, Pune and presently is in Fourth Year



Ms. Gayatri Sawale is pursuing B.E(IT) from Bharati Vidyapeeth's College of Engineering for Women, Pune and presently is in Fourth Year



Ms. Rajashree Shelke is pursuing B.E(IT) from Bharati Vidyapeeth's College of Engineering for Women, Pune and presently is in Fourth Year



Prof. A. D. Khairkar is working as a professor in department of IT, Bharati Vidyapeeth's College of Engineering for Women, Pune.