



OPEN Predicting MBTI personality of YouTube users

Luisa Stracqualursi[✉] & Patrizia Agati

For several decades, the standard method of measuring personality has been through self-report questionnaires. In this work, we apply a model based on the analysis of the words used by people. Using Machine Learning (ML) and Natural Language Processing (NLP) techniques, we created a classifier that categorizes people into their Myers-Briggs Type Indicator (MBTI) personality type based on text samples extracted from their social media comments. Then, we applied this classifier to determine the MBTI personality types of YouTube users who comment on videos related to conspiracy theories, spirituality, and travel. The results highlight that a sufficient estimate can be made with only 100 words per user, as long as all 100 words are related to the same topic. The most common MBTI type among YouTubers who comment on these types of videos is INFP, followed by INTP, INFJ, and INTJ, indicating a predominance of Introverted and Intuitive individuals. While the predominant personality type in YouTube comments on videos related to conspiracy theories is INTP, in spiritual and travel topics, the most common personality type is INFP.

People differ greatly in how they think, feel, and behave, which forms the foundation of what is commonly referred to as personality. For several decades, self-report questionnaires have been the standard approach for measuring personality. Traditionally, such questionnaires were seen as capturing individuals' beliefs about their own traits—what is often referred to as the 'self-disclosure' view of personality¹. However, it is important to note that other perspectives exist. For example, the 'self-presentation' view suggests that responses to personality questionnaires can also be seen as performative acts, reflecting how individuals wish to present themselves in different contexts rather than being mere passive self-descriptions^{2,3}. This broader understanding aligns with current approaches to analyzing individuals' language use as a means of personality assessment. In fact, the 'self-presentation' and 'language-based personality assessment' both reflect how individuals present themselves through their expressions, capturing not only their self-perceptions but also the ways in which they perform and communicate their identity to others. Moreover, research has shown that well-constructed personality measures, such as Costa and McCrae's NEO Personality Inventory and the Hogan Personality Inventory, display strong convergent correlations with ratings provided by knowledgeable acquaintances and are predictive of significant life outcomes^{4–6}. This evidence highlights that personality questionnaires, when properly designed and validated, can offer robust and objective insights into personality traits beyond an individual's self-perception.

In recent years, personality measurement has garnered increasing interest due to the growing demand for personalized content on the Internet. Accurately inferring users' personalities has become crucial for enhancing the effectiveness of various applications, such as targeted advertising. Shanahan et al. (2019)⁷ point out that social media marketing expenditures have more than quadrupled in the past decade, underscoring the importance of personality inference for business and marketing strategies.

Since the foundational work of Pennebaker and Graybeal⁸, numerous studies have focused on evaluating personality through written language. These studies demonstrate that personality traits influence language use, providing valuable insights into how individuals express their thoughts and emotions^{1,9–12}. Language can serve as a rich source of personality cues, particularly because personality tends to remain stable over time, allowing consistent patterns to emerge in large bodies of text¹³.

More recently, large language models (LLMs) such as ChatGPT have been employed in the assessment of personality traits, offering new avenues for research. For example, Hussain et al. (2024) demonstrate how open-source LLMs can be used in behavioral science applications¹⁴, while Peters and Matz (2024) show that these models can accurately infer psychological dispositions from social media data¹⁵. These advances underscore the growing utility of LLMs in personality research.

Historically, research on personality prediction has typically focused on two popular models: the Myers-Briggs Type Indicator (MBTI)¹⁶ and the Big Five personality traits¹⁷. There is substantial evidence suggesting correlations between these two models, with MBTI classifications aligning with several Big Five traits^{18,19}. Although MBTI remains widely used, particularly in organizational contexts, there is considerable debate within the academic community regarding its utility compared to the Big Five model. The MBTI's popularity may stem from its ease

Department of Statistics, University of Bologna, Bologna 40126, Italy. ✉email: luisa.stracqualursi@unibo.it

of use and brand recognition, rather than any inherent superiority in measuring personality^{20–22}. Nonetheless, it is highly valued by practitioners, especially for activities such as group exercises and workshops^{23,24}.

In the context of social media, an ever-growing number of individuals share details about their lives, thoughts, and emotions online. This has led to a surge in studies using Machine Learning (ML) techniques and Neural Networks (NNs) to predict personality traits based on user-generated content. These studies have explored diverse personality models, datasets, and learning algorithms. Choong and Varathan (2021)²⁵ provide an in-depth examination of these methods, while He and de Melo (2021)¹³ perform correlation analyses on MBTI data, revealing compelling patterns between personality dimensions and other traits, and demonstrating the robustness of their data.

Despite the progress made, automatic personality prediction remains a complex and challenging task. Two key issues persist: the limited availability of publicly accessible labeled datasets for training classifiers, and the small sample sizes that characterize many existing datasets, often providing only a few sentences per sample¹³. These challenges underscore the need for further research to refine and improve automatic personality assessment methodologies.

The MBTI type indicators

The MBTI, or Myers-Briggs Type Indicator¹⁶, is a psychometric tool designed to measure individual preferences across four dichotomous type pairs through a series of forced-choice questions. This instrument, based on Jung's theory of types, suggests that each person naturally gravitates towards one end of the spectrum for each pair of psychological preferences measured by the MBTI, emphasizing that neither preference is inherently superior to the other. In other words, there is no inherent value judgment placed on being, for example, extraverted versus introverted, or sensing versus intuitive. This contrasts with some other psychological assessments that may label certain traits as positive or negative. It utilizes single letters from the alphabet to represent eight preferences, each of which are across 4 axes and corresponds closely to Jung's original definitions²⁶:

- **Introversion (I) vs Extraversion (E):** Introversion (I) reflects a preference for the inner world of ideas and personal reflections. Introverts draw energy from their internal thoughts and feelings, directing their attention inward and often valuing solitude. Extraversion (E) denotes a propensity to focus on the external world, including people and events. Extraverts derive energy from external stimuli, directing their attention outward towards interactions and experiences.
- **Intuition (N) vs Sensing (S) :** Intuition (N) represents the ability to perceive information by discerning patterns and connections, focusing on the broader picture rather than specific details. Intuitive individuals excel at envisioning possibilities and exploring alternative perspectives. Sensing (S) signifies a preference for gathering information through the senses, such as sight, sound, and touch. Individuals with a sensing preference are attentive to concrete details and practical aspects of their environment.
- **Thinking (T) vs Feeling (F) :** Thinking (T) reflects a preference for making decisions based on logical consequences, considering the cause-and-effect relationships. Those who prefer thinking tend to detach themselves emotionally to analyze situations objectively. Feeling (F) reflects the inclination to make decisions based on personal values and emotions, considering what is important to oneself and others involved. Individuals with this preference empathize with others to understand their perspectives before making decisions.
- **Judging (J) vs Perceiving (P) :** Judging (J) indicates a preference to organize life in a planned, orderly way, with a desire to regulate and control it. People who prefer judging make decisions, achieve closure, and appreciate an environment that is structured and organized. Perceiving (P) denotes a preference for flexibility and adaptability, embracing spontaneity and openness to new experiences. Individuals with this preference enjoy exploring various options and trust their ability to adapt to changing circumstances. The MBTI, in contrast to Jung's theory of principal and subordinate functions²⁶, treats each preference equally, determining sixteen different personality types based on four sets of dichotomous preferences. Each personality type is represented by a unique combination of four letters indicating dominant preference areas²⁷.

Outline of the paper

In this study, we began with the ML method utilized for MBTI personality type prediction by Amirhosseini and Kazemian (2020)²⁸, which is based on Extreme Gradient Boosting (XGBoost). We then made some adjustments to this method. After confirming that these adjustments increased the model's predictive capacity, we applied our classifier to scrape data from YouTube.

Specifically, we employed our classifier to determine the MBTI personality types of YouTube users who comment on videos related to conspiracy theories, spirituality, and travel. In this article, we aim to address three main questions:

- What is the minimum number of words per user to have an appropriate MBTI personality prediction? What are the most common MBTI types among YouTubers who frequently comment on videos?
- Is there a predominant MBTI type among YouTubers who comment on conspiracy videos? Does this pattern vary among the three conspiracy theories examined? Is there a specific characteristic that influences this type of interest?
- How do the personality types of individuals who comment on spiritual, travel, and conspiracy videos differ? We addressed these questions using our MBTI estimator.

The data

In this paper, the well-known MBTI Kaggle dataset²⁹ was used to train our classifier. This dataset is sourced from the 'Personality Cafe' forum³⁰ and collects 50 posts in natural English for each of the 8,675 individuals, whose

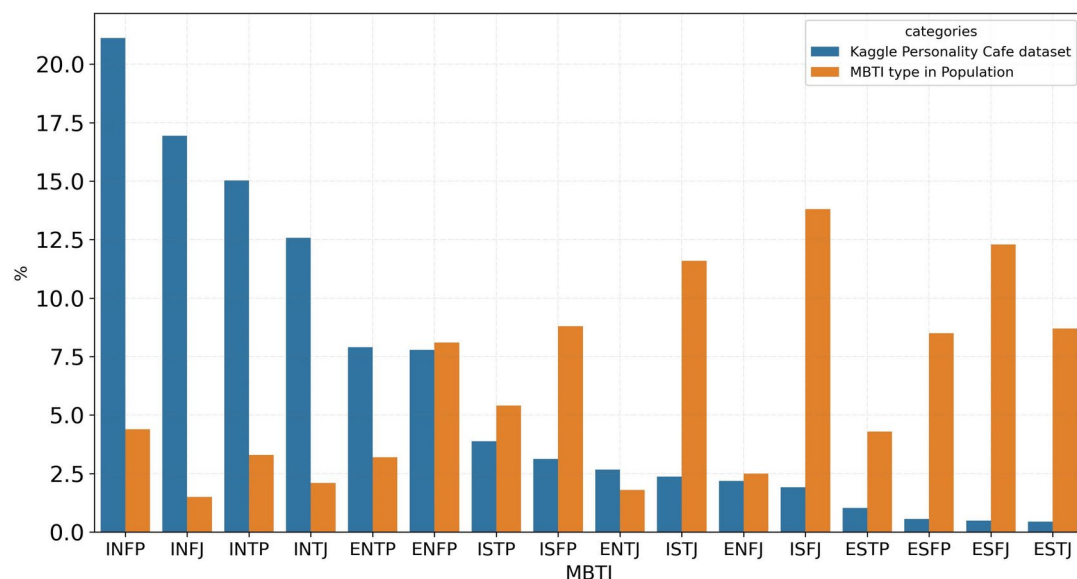


Fig. 1. Distribution of 16 MBTI personality types. The figure shows the percentage distribution of personality types in the Kaggle dataset and in the MBTI reference population.

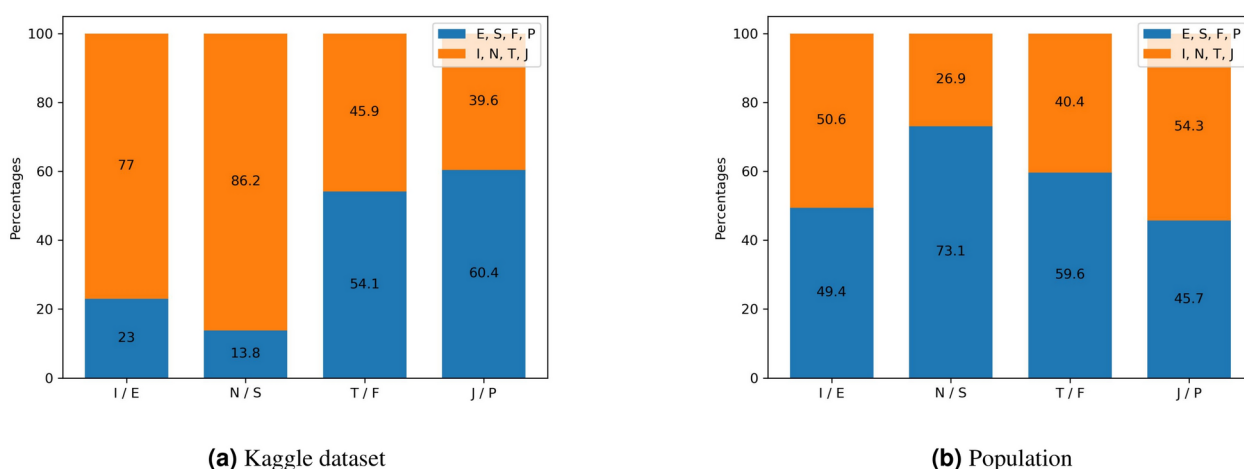


Fig. 2. Distribution across the 4 MBTI dichotomies: Kaggle vs Population. The figure shows the percentage distribution across the 4 MBTI dichotomies for the Kaggle dataset (a) and for the MBTI population (b).

MBTI personality type is known. Here, we present two graphs from Exploratory Data Analysis (EDA) of this dataset, illustrating the data distributions of the 16 personality types (Fig. 1) and the distribution within each of the 4 MBTI binary dichotomies (Fig. 2).

As shown in Fig. 1, these results indicate a non-uniform representation of MBTI types in the dataset, which does not align with the proportions of MBTI types in the reference population. Particularly, the reference population refers to the U.S. population, estimated through a stratified random sample (N=3009) as explained in the Myers and McCauley manual (1986)¹⁶.

To understand the distribution of type indicators within the dataset, four distinct personality pairs were established. The first dichotomy distinguishes between Introversion (I) and Extroversion (E), while the second categorizes iNtuition (N) against Sensing (S). The third dichotomy differentiates between Thinking (T) and Feeling (F), and the fourth dichotomy delineates Judging (J) from Perceiving (P). Consequently, each trait yields one letter, resulting in a combination of four letters that correspond to one of the 16 personality types in the MBTI. For example, if the first trait yields 'E', the second yields 'S', the third yields 'T', and the fourth yields 'P', the resulting personality type would be ESTP. Fig. 2 shows the percentages for the 4 MBTI dichotomies in the Kaggle dataset (a) and in the population of Myers-Briggs¹⁶ Type Index (b). It can be observed that the percentages of these dichotomies between the reference population and our dataset are quite similar, except for the Introverts-Extroverts (I/E) and Intuitives-Sensing (N/S) dichotomies. In the Kaggle dataset, the I/E distribution is 77% vs. 23%, whereas in the reference population, it is approximately 50% vs. 50%. Additionally, there is a complete

reversal in the Intuitives-Sensing (N/S) dichotomy, where the N/S distribution in the Kaggle dataset is 86.2% vs. 13.8%, while in the reference population, it is 26.9% vs. 73.1%.

After training a suitable classifier the analysis was then carried out on 140,933 posts (see Supplementary File ‘S1’) coming from the scraping of YouTube users’ comments on some of the most common videos relating to the following topics:

- 1. *Conspiracy theories*:
 - a) Extraterrestrials - Among the foremost concerns of conspiracy theorists are questions of alien life. They argue that various governments and in particular the United States government, are suppressing evidence that aliens exist and that Earth governments are in communication or cooperation with extraterrestrial visitors despite public disclaimers.
 - b) Vaccines - Conspiracy theorists claim that the Coronavirus was engineered for specific goals, including to force mass vaccinations as a part of a population control scheme. Throughout 2020, an alleged correlation between Coronavirus and 5G mobile technology was also widely promoted and disseminated on social media, often in the form of maps overlaying the distribution of COVID-19 cases on the installation of 5G towers.
 - c) New world order - This theory states that a group of international elites control governments, industry, and media organizations, with the goal of establishing global hegemony. This group is alleged to be implicated in most of the major wars of the last two centuries, to carry out secretly staged events, and to deliberately manipulate economies.
- 2. *Spiritual* topics: A dataset obtained by scraping YouTube users’ comments to some most common spiritual videos.
- 3. *Travel* topics: A dataset obtained by scraping YouTube users’ comments to some most common travel videos. Table 1 shows the number of posts collected for each examined topic.

Pre-processing

To clean the data and prepare it for training a classifier, we applied the following preprocessing steps using NLP techniques implemented in Python. Using the NLTK library, the posts were lemmatized, i.e., inflected forms of the words were transformed into their root words.

We used customized stop words. Starting from English stop words included in the NLTK library, we excluded terms such as ‘myself’, ‘ourself’, ‘themselves’, etc. and ‘own’ and ‘against’, which in our opinion, might give some idea of some MBTI dimensions. Then the corpus was lower-cased and we cleaned data by removing:

- customized stop words,
- URL links,
- no words,
- multiple spaces and multiple letter repeating words,
- mentioning of personality types. We used TF-IDF (Term Frequency-Inverse Document Frequency) representation. TF-IDF for feature engineering evaluates how relevant/important a word is to a document in a collection of documents or a corpus. As we train individual classifiers here, it is very useful for scoring words in machine learning algorithms for NLP. The ‘Sklearn’ library was used to recognize the words appearing in 10% to 70% of the posts. In general, TF-IDF representations downweigh words that appear universally across many documents, as these are less likely to be sufficiently discriminative in personality prediction¹³. In the first step, posts were placed into a matrix of token counts. In the next step, the model learned the vocabulary dictionary and returned a term-document matrix. The count matrix then transforms into a normalized TF-IDF representation, which can be used for the Extreme Gradient Boosting classifier. After removing all irrelevant words, multiple spaces, and repeated words, as well as customizing the stopwords, 596 words were found in 10–70% of the posts, while were 791 in Amirhosseini and Kazemian (2020)²⁸. A more accurate approach to data cleaning - by removing noise and less relevant information - allows the model to focus on more meaningful features. While this cleaning process might lead to a slight reduction in the classifier’s immediate performance, we believe it ensures greater model robustness and improves its ability to generalize to new data, reducing the risk of overfitting on unhelpful information.

Topic	No. of comments
Conspiracy theories:	
- Extraterrestrials	28,627
- New world order	28,431
- Covid, vaccines and 5G	23,307
Spiritual	24,103
Travel	36,783
	141,251

Table 1. Number of comments collected for each topic.

Methods
The classifier

The classification task was segmented into four distinct binary tasks, aligning with the structure of each MBTI type, which comprises four binary classes. These binary classes signify specific facets of personality based on the MBTI model: I/E, N/S, T/F, J/P. Consequently, four separate binary classifiers were trained, each specializing in one aspect of personality. Thus, at this stage, a distinct model was developed for each type indicator. Term Frequency-Inverse Document Frequency (TF-IDF) was applied, and MBTI type indicators were transformed into binary representations. Therefore, our data (X variable) was the posts in TF-IDF format, while the response variable Y was the binarized MBTI type indicator.

Using the *train_test_split* function from 'Sklearn' library, 70% of the Kaggle dataset was used as the training set and 30% of the data was used as the test set. The model was fitted into the training data and the predictions were made for the testing data. The XGBoost Python API was used in this step to create the XGBoost Model in order to predict the MBTI personality of each user. MBTI classifiers were trained individually, XGBoost, short for eXtreme Gradient Boosting, is a powerful machine learning algorithm that has achieved considerable success in a wide range of practical applications because it is highly customizable to the particular needs of the application; it is also efficient, speedy, and accurate³¹. It belongs to the family of boosting algorithms and is a method based on creating a very accurate prediction rule³², using ensemble learning techniques that combine the predictions of multiple weak learners. Fundamentally, XGBoost builds a strong predictive model by aggregating the predictions of several weak learners, usually decision trees. It uses a boosting technique to create an extremely accurate ensemble model by having each weak learner after it correct the mistakes of its predecessors. The optimization method (gradient) minimizes a cost function by repeatedly changing the model's parameters in response to the gradients of the errors³³. A number of configuration heuristics have been published in the original gradient boosting papers such as: '*learning_rate*' in XGBoost should be set to 0.1 or lower, and smaller values will require the addition of more trees or '*max_depth*' in XGBoost should be configured in the range of 2-to-8, where not much benefit is seen with deeper trees.

Using scikit-learn we performed a grid search of principal XGBoost parameters, finding the best results in order to improve generalization and prevent overfitting. Here are the results that are in line with that of Amirhosseini and Kazemian (2020)²⁸:

- *learning_rate* = 0.2: The learning rate, also known as shrinkage, quantifies each tree's contribution to the total prediction. Since each tree has less of an influence, an optimization process with a lower learning rate is more resilient.
- *n_estimators* = 200: This indicates the number of trees in an XGBoost model.
- *nthread* = 8: This refers to the number of parallel threads used to run XGBoost.
- *max_depth* = 2: The maximum depth of a tree, is used to control over-fitting as higher depth will allow model to learn relations very specific to a particular sample but increase the overfitting risks.

Table 2 displays the classifiers results in terms of accuracy and Roc AUC scores for both the training and test set. We can see that the model suffers from overfitting: the testing accuracy is lower than the training accuracy, and the classifiers have low prediction ability since the AUC ROCs are very low, especially in some dimensions of MBTI. This is partly due to the unbalanced distribution of MBTI dimensions in our dataset. We therefore decided to balance the data by applying an oversampling technique.

Synthetic minority oversampling technique

As depicted in Fig.2, the training dataset exhibits an imbalance between MBTI dimensions. To address this issue, resampling techniques are employed to mitigate the discrepancy in dimension sizes. Two approaches are commonly considered: under-sampling, which involves excluding instances from the majority class, and over-sampling, which entails generating examples for the minority class³⁴. Given that random under-sampling has demonstrated subpar outcomes in similar studies, over-sampling strategies, such as the Synthetic Minority Over-sampling Technique (SMOTE), are often preferred in data balancing methods³⁵.

SMOTE aims at balancing class distribution by generating synthetic instances to represent minority classes. The technique operates by inserting artificial samples along the line connecting the current instance and some of its k-nearest neighbors from the same class for each minority class sample. Depending on the desired level of oversampling, the algorithm randomly selects some of the k-nearest neighbors and constructs pairs of vectors to generate synthetic samples. These newly created instances expand decision boundaries, rendering them more extensive and denser. Consequently, classifiers can gain more insights from the smaller classes within

MBTI dichotomies	Training		Test	
	Accuracy	Roc AUC	Accuracy	Roc AUC
Introversion (I) - Extroversion (E)	82.2%	92.3%	76.9%	65.7%
Intuition (N) - Sensing (S)	88.4%	94.9%	86.2%	64.5%
Feeling (F) - Thinking (T)	84.2%	92.0%	70.7%	77.2%
Judging (J) - Perceiving (P)	79.2%	89.0%	65.5%	64.8%

Table 2. Results of MBTI classifiers: training and test accuracy and ROC AUC scores.

MBTI dichotomies	Accuracy	ROC AUC
Introversion (I) - Extroversion (E)	75.6%	83.0%
Intuition (N) - Sensing (S)	79.1%	87.5%
Feeling (F) - Thinking (T)	71.8%	81.2%
Judging (J) - Perceiving (P)	66.11%	71.9%

Table 3. Test results of MBTI classifiers on real data after oversampling.

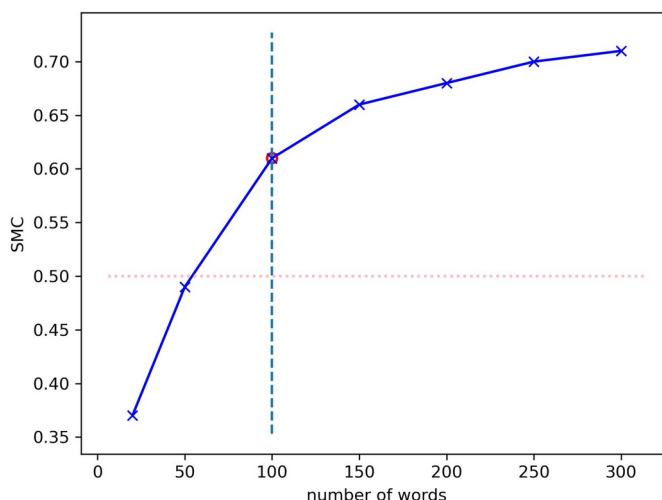


Fig. 3. Average SMC according to the number of words per users. This measure is obtained as the average of the Simple Matching Coefficient of the 4 MBTI dichotomies.

these decision regions, as opposed to predominantly focusing on the larger surrounding classes. According to Wang et al., SMOTE surpasses other oversampling algorithms in performance³⁶. By generating new synthetic samples, instead of simply duplicating existing samples, SMOTE can help to reduce the risk of overfitting which commonly accompanies random oversampling. Hence, SMOTE is meant to be an improvement over random oversampling.

The results in terms of accuracy are very similar to table 2, but by applying SMOTE we reduce overfitting improving ROC AUC on the test set (table 3), thus increasing the predictive capacity of the model.

Simulation setting for a new dataset

Often real data sets only provide a limited number of posts per user, making it difficult to apply an automated model for personality prediction. To apply our model to a new real dataset, we need to determine the minimum number of words per user required for a satisfactory prediction.

We selected subsets of 20–50–100–150–200–250–300 words for each user in the test set and tested our model on each subset. Once we chose the Simple Matching Coefficient (SMC) as a measure of our model's ability to match the 4 MBTI dimensions, we plotted the SMC as a function of the number of words in the posts. Finally, we used the well-known elbow³⁷ method: that is, we chose the elbow of the curve as the minimum number of words to use. Fig. 3 represents the (pink dashed line) minimum threshold for the SMC for a match, values above 0.5 indicate similarity. The elbow point displayed in red indicates the minimum value for a match, which is 100 words per user.

The relationship between the ability to predict personality traits and the number of words required by the classifier is complex and certainly sensitive to the nature of the content. Different topics, such as conspiracy theories, spiritual discussions, and travel experiences, evoke varied emotional responses and levels of engagement from users.

For instance, comments on conspiracy theories often involve more emotionally charged and elaborate reasoning, which may provide deeper insights into personality traits compared to more straightforward travel-related comments.

Results

Assess MBTI dimensions on YouTube dataset

Since video comments can be written in different languages, any comments in languages other than English were translated into English using Google Translate, a free Python library that utilizes the Google Translate API³⁸. To identify the MBTI dimensions of each Youtuber in our dataset, we grouped YouTube comments by user for every topic and according to our simulation settings (see par. 3.3) we pruned data by placing a minimum limit

Topic	No. of users	No. of users selected
Conspiracy theories:		
- Extraterrestrials	20,297	1,389
- New world order	19,071	1,482
- Covid, vaccines and 5G	13,611	1,033
Spiritual	17,926	1,263
Travel	27,422	868
	98,327	6,035

Table 4. Number of comments and user selected for each topic.

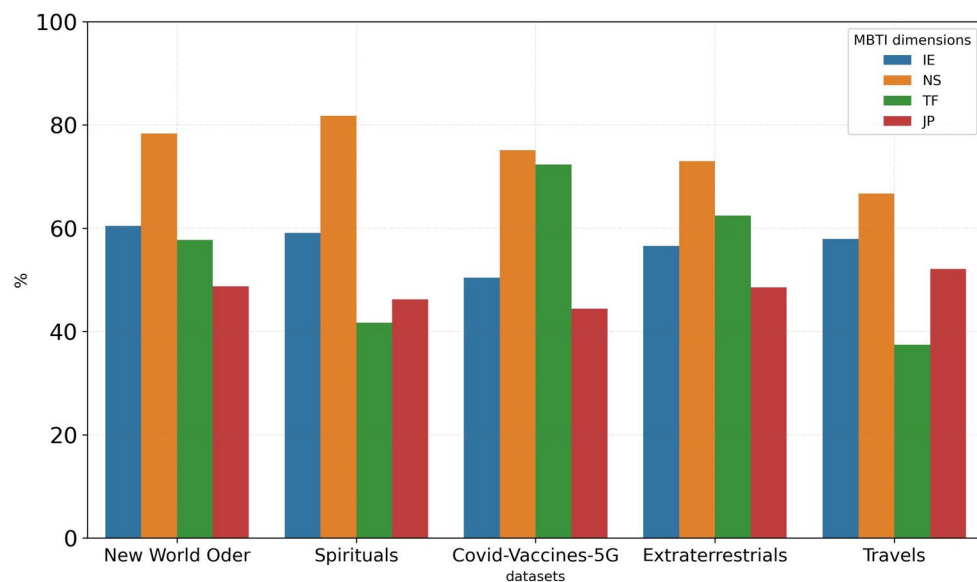


Fig. 4. Percentage distributions of the 4 MBTI dichotomies in the YouTube dataset. The figure illustrates the percentage distribution of the four MBTI dichotomies across the various video categories from which comments were collected.

of 100 words (e.g., users with fewer than 100 words were eliminated). Finally, we applied our classifier. Table 4 shows the number of comments scraped for each type of topic video before and after pruning. Then we applied pre-processing cleaning and prepared our corpus, for applying the trained XGBoost classifier.

Fig. 4 shows the percentage distributions within each dimension in the different MBTI datasets relative to different video comments. For each dimension in a dichotomy, the first type is coded as 1, and the second type is coded as 0. For example, for I-E, 'Introvert' is represented as 1, and 'Extravert' is represented as 0. Regarding the I-E dimension, the only dataset where the two percentages are balanced is the Covid-Vaccine-5G one. A (slight) prevalence of 'Introverted' individuals over 'Extroverts' can instead be noted in all other datasets: this result aligns with literature findings, which suggest that 'Introverted' individuals prefer online communication^{13,39,40}.

Observing the behavior of the N-S dimension in the examined datasets, a marked prevalence of 'iNtuitive' individuals over 'Sensing' individuals stands out, with percentages ranging from 67% to 81%. As regards the remaining dimensions (T-F and J-P), no homogeneity of behavior emerges in the five datasets. 'Thinking' individuals prevail over 'Feeling' ones when debating issues such as the New world order (59%), Covid vaccines and 5G (72%), or Extraterrestrials (62%). 'Judging' and 'Perceiving' individuals are approximately balanced, except in the 'Spiritual' and 'Covid-Vaccine-5G' datasets, where 'Perceiving' individuals outnumber 'Judging' ones (with percentages around 55% for both subsets).

The 4 MBTI dichotomies from the YouTube dataset show some differences in percentage composition compared to the original MBTI population¹⁶ (Fig. 5), but only N-S dichotomy is completely reversed in proportion. This indicates a clear predominance of iNtuitive individuals among the YouTubers in our dataset.

Interestingly, there are some differences in MBTI types among users in YouTube datasets (see Fig. 5). The most frequent MBTI type in the 'Extraterrestrial' category are INTP, which cover approximately 16% of users, while the type most represented in the 'Covid-Vaccines-5G' category is INTP (approximately 19%). In the 'Spirituals' category, the largest percentages is INFP that reach the maximum value on all types equal to 21%, while in the 'Travels' sector, the INFP and INFJ types (with percentages about 18% and 15.5%, respectively) prevail over all others. In the 'New world order' category, the most present types are INFPs and INTPs, with coverage percentages of 15% and 14% respectively. It is worth noting that the distribution of MBTI types in each

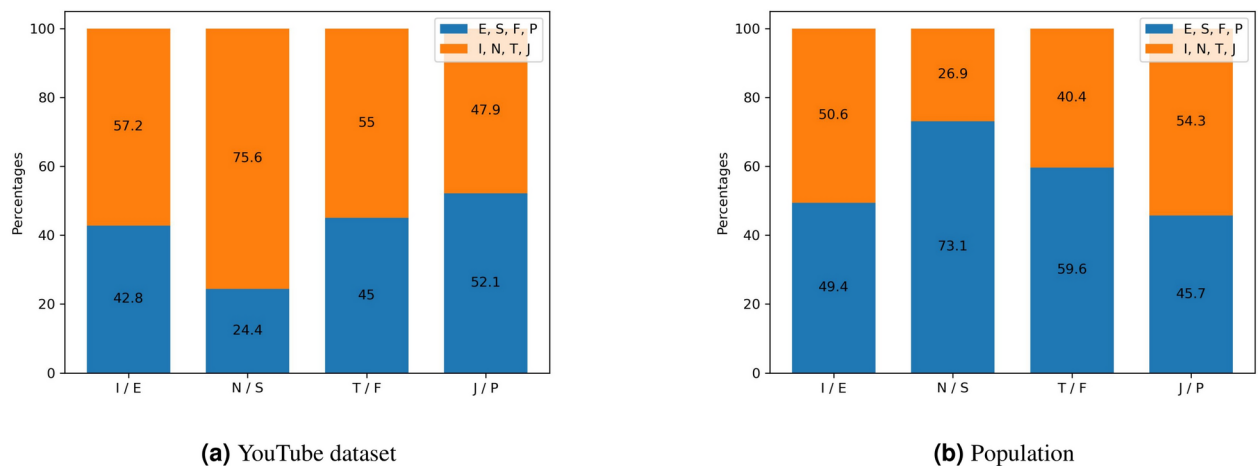


Fig. 5. Distribution across the 4 MBTI dichotomies: YouTube vs Population. The figure shows the percentage distribution across the four MBTI dichotomies for the YouTube comments dataset (a) and for the MBTI reference population (b).

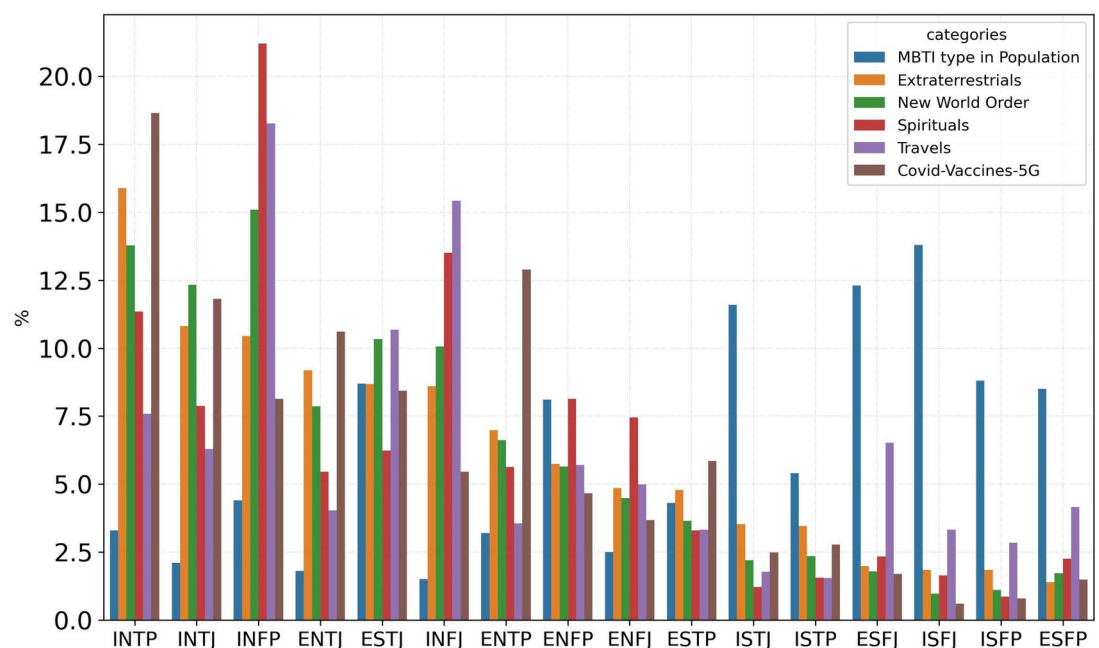


Fig. 6. Distributions of the 16 MBTI types by topics. The figure shows the personality types in the YouTube comments dataset broken down by video categories, also comparing these types to those in the MBTI reference population.

of the five datasets examined appears very different compared to the distribution in the Population (displayed in blue in Fig. 6), where the ISFJ, ESFJ and ISTJ types prevail, with percentages not too different from each other (about 13.8%, 12.3% and 11.6% respectively). Therefore the predominant personality types in videos related to conspiracy theories is INTP, while in spiritual and travel topics, the most common personality type is INFP.

Finally Fig. 7 displays the percentages of 16 MBTI type across the whole YouTube dataset indicating a prevalence of INFP types, followed by INTP, INFJ, and INTJ. This indicates a predominance of Introverted and Intuitive individuals between YouTuber who comments such video.

Strengths and limitations

One advantage of this study over other research using the same training dataset²⁹ to develop an MBTI classifier is its use of a model trained on topic-specific conversations for similar data. Our extracted YouTube comments are tied to specific topics, aligning well with the trained classifier. In contrast, other research often applies classifiers

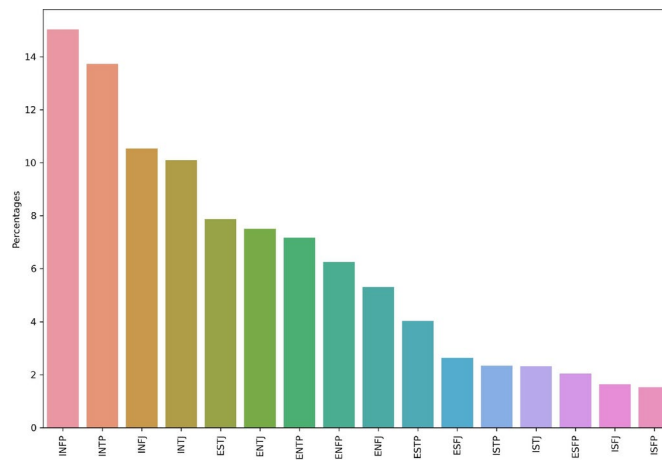


Fig. 7. Percentage distributions of the 16 MBTI types in YouTube dataset.

to predict the personality of social media users discussing diverse topics, which may not match the classifier's training data as closely⁴¹.

Other studies indicate that at least 25 tweets⁴² are needed to make a significant prediction of personality. Since a tweet is limited to 280 characters, or about 55 words⁴³, 25 tweets should contain approximately 1375 words.

Our work introduces the notion that merely 100 words per user can yield adequate personality predictions for the first time. Even if training a classifier with a limited dataset of only 100 words per user may seem risky and inaccurate for determining the correct personality of individual users, its limitations become significantly less pronounced when applied to user groups as a whole. This is especially true if you have a dataset capturing the nuances and variability inherent in group behavior, such as a dataset containing discussions on specific topics. Furthermore, since YouTube comments are made in response to videos, it is reasonable to assume that the experience is emotionally more engaging, as viewers are stimulated by both sight and sound, rather than just reading text or viewing images. In such cases, the classifier becomes more effective in detecting trends and behavioral patterns within the group, even though individual accuracy may be just sufficient.

The limitations of the present work are common to all studies that have used our same training dataset or have employed the datasets currently available to train a classifier. Here is a detailed summary of these limitations:

- **Imbalance of user classes:** The user classes represented in the Kaggle corpus are largely unbalanced and do not closely match the population distribution. This imbalance could influence the results of the classification model, as some user categories may be overrepresented or underrepresented compared to the demographic reality.
- **Self-administered MBTI labels:** The large corpora currently available (included Personality Cafe dataset) have MBTI (Myers-Briggs Type Indicator) labels from self-administered MBTI assessments. This means that there is no information on which version of the MBTI assessment was used, leading to inconsistent and potentially unreliable data.
- **Trait personality assessments in binned form:** In psychometric assessments of personality traits, personality is measured in continuous scores. However, the available benchmark datasets primarily present personality trait scores in artificially binned form, i.e., divided into categories rather than represented as continuous data. This may affect the accuracy and completeness of the analysis of personality traits and their relationships with other factors.

Discussion and conclusions

Language reveals key psychological traits, and in the digital age, understanding personality is critical for personalized services, such as targeted ads and content recommendations. However, predicting personality from short online comments, especially on platforms like YouTube, is still challenging.

Our study used a pre-trained XGBoost model, initially trained on the 'Personality Cafe' forum data³⁰, to classify MBTI personality types from 140,933 YouTube comments related to conspiracy theories, spirituality, and travel. After filtering users by selecting only those with a minimum of 100 words, we analyzed 5,782 users (Table 4). This approach demonstrates (see section 4) how machine learning can effectively predict personalities even from short-form content on specific topics. The predominant personality among the YouTubers in the entire dataset turned out to be the INFP, commonly known as 'Mediator', which identifies a personality that is calm, open-minded, and imaginative. This type is followed by INTP, INFJ, and INTJ, highlighting a predominance of 'Introverted' and 'Intuitive' individuals. Indeed, 'Introverted'(I) individuals reflect a preference for the inner world of ideas and personal reflections, drawing energy from their internal thoughts and feelings and often appreciating solitude. On the other hand, this result is consistent with the literature, where it emerges that 'Introverted' individuals prefer online communication^{13,39,40}. Therefore, it is not surprising that they are also predominant among the YouTubers. Instead, the Intuitive (N) subjects present the ability to perceive information

by discerning patterns and connections, focusing on the broader picture rather than specific details. What better way to get an overview of certain topics than by watching videos?

However, if we distinguish between specific video categories within the dataset, the predominant personality, as well as the traits, change. For videos related to conspiracy theories, the prevalent personality type is INTP, also known as ‘Logicians’. INTPs are characterized by their flexible thinking and inclination toward unconventional approaches in life, coupled with a penchant for experimentation and personal creativity. Conversely, videos discussing spiritual and travel topics attract comments predominantly from individuals of the INFP personality type (Mediator).

The analysis of personality types among individuals who engage with spiritual, travel, and conspiracy-related videos reveals a notable distinction primarily centered around one of the MBTI dichotomies: Thinking (T) versus Feeling (F). Indeed, as shown above, the most prevalent personality type in conspiracy theory topics is INTP, while in spiritual and travel topics it is INFP. Thinking (T) individuals, characterized by a preference for logic and rationality, tend to approach situations with a detached emotional stance, prioritizing objective analysis. Their decision-making process is often guided by logical reasoning rather than personal emotions or values. On the other hand, Feeling (F) individuals exhibit a propensity to make decisions based on personal values and emotions, considering not only their own perspectives but also empathizing with others involved. Their emotional connection plays a significant role in shaping their responses and interpretations. Despite the limitations discussed, we hope this initial study can contribute to the understanding of online behavior and personality dynamics in digital environments.

Data availability

All data generated or analyzed during this study are included in the Supplementary Information Files of this published article: **S1** Dataset of scraped YouTube comments. The dataset for train our classifier came from Kaggle: “<https://www.kaggle.com/datasnaek/mbti-type>”. The YouTube scraping dataset is included in Supplementary Information of this published article.

Received: 29 June 2024; Accepted: 1 January 2025

Published online: 28 February 2025

References

- Boyd, R. L. & Pennebaker, J. W. Language-based personality: a new approach to personality in a digital world. *Curr. Opin. Behav. Sci.* **18**, 63–68. <https://doi.org/10.1016/j.cobeha.2017.07.017> (2017).
- Johnson, J. A. The, “self-disclosure” and “self-presentation” views of item response dynamics and personality scale validity. *Journal of Personality and Social Psychology* **40**, 761–769. <https://doi.org/10.1037/0022-3514.40.4.761> (1981).
- Johnson, J. A. Seven social performance scales for the California psychological inventory. *Human Performance* **10**, 1–30. https://doi.org/10.1207/s15327043hup1001_1 (1997).
- Hofstee, W. K. Who should own the definition of personality? *European journal of Personality* **8**, 149–162. <https://doi.org/10.1002/per.2410080302> (1994).
- Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A. & Goldberg, L. R. The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science* **2**, 313–345. <https://doi.org/10.1111/j.1745-6916.2007.00047.x> (2007).
- Soto, C. J. How replicable are links between personality traits and consequential life outcomes? the life outcomes of personality replication project. *Psychological Science* **30**, 711–727. <https://doi.org/10.1177/0956797619831612> (2019).
- Shanahan, T., Tran, T. P. & Taylor, E. C. Getting to know you: Social media personalization as a means of enhancing brand loyalty and perceived quality. *Journal of Retailing and Consumer Services* **47**, 57–65. <https://doi.org/10.1016/j.jretconser.2018.10.007> (2019). <https://www.sciencedirect.com/science/article/pii/S0969698918300055>.
- Pennebaker, J. W. & Graybeal, A. Patterns of natural language use: Disclosure, personality, and social integration. *Curr. Dir. Psychol. Sci.* **10**, 90–93. <https://doi.org/10.1111/1467-8721.00123> (2001).
- Hirsh, J. B. & Peterson, J. B. Personality and language use in self-narratives. *J. Res. Pers.* **43**, 524–527. <https://doi.org/10.1016/j.jrp.2009.01.006> (2009).
- Schwartz, H. A. et al. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS One* **8**, e73791. <https://doi.org/10.1371/journal.pone.0073791> (2013).
- Boyd, R. L. & Pennebaker, J. W. Did Shakespeare write double falsehood? identifying individuals by creating psychological signatures with text analysis. *Psychol. Sci.* **26**, 570–582. <https://doi.org/10.1177/0956797614566658> (2015).
- Boyd, R. et al. Values in words: Using language to evaluate and understand personal values. In *Proceedings of the International AAAI Conference on Web and Social Media* **9**, 31–40. <https://doi.org/10.1609/icwsm.v9i1.14589> (2015) ((Association for the Advancement of Artificial Intelligence (AAAI)).
- He, X. & de Melo, G. Personality predictive lexical cues and their correlations. In *Proceedings of the Conference Recent Advances in Natural Language Processing - Deep Learning for Natural Language Processing Methods and Applications*, https://doi.org/10.26615/978-954-452-072-4_058 (INCOMA Ltd. Shoumen, BULGARIA, 2021).
- Hussain, Z., Binz, M., Mata, R. & Wulff, D. U. A tutorial on open-source large language models for behavioral science. *Behavior Research Methods* 1–24. <https://doi.org/10.3758/s13428-024-02455-8> (2024).
- Peters, H. & Matz, S. C. Large language models can infer psychological dispositions of social media users. *PNAS nexus* **3**, pgae231, <https://doi.org/10.1093/pnasnexus/pgae231> (2024).
- Myers, I. B. & McCaulley, M. H. *MBTI Manual: A Guide to the Development and Use of the Myers-Briggs Type Indicator* (Consulting Psychologists Press, 1986).
- Goldberg, L. R. An alternative “description of personality”: The Big-Five factor structure. *J. Pers. Soc. Psychol.* **59**, 1216–1229. <https://doi.org/10.1037/0022-3514.59.6.1216> (1990).
- McCrae, R. R. & Costa Jr, P. T. More reasons to adopt the five-factor model. *American Psychologist* **44**, 451–452. <https://doi.org/10.1037/0003-066X.44.2.451> (1989).
- Tobacyk, J. J., Livingston, M. M. & Robbins, J. E. Relationships between Myers-Briggs type indicator measure of psychological type and neo measure of big five personality factors in polish university students: a preliminary cross-cultural comparison. *Psychol. Rep.* **103**, 588–590. <https://doi.org/10.2466/pr0.103.2.588-590> (2008).
- Jr. Barbut, J. E. A critique of the Myers-Briggs type indicator and its operationalization of Carl Jung’s psychological types. *Psychological Reports* **80**, 611–625. <https://doi.org/10.2466/pr0.1997.80.2.611> (1997).

21. Furnham, A. The big five versus the big four: the relationship between the myers-briggs type indicator (mbti) and neo-pi five factor model of personality. *Personality and Individual Differences* **21**, 303–307. [https://doi.org/10.1016/0191-8869\(96\)00033-5](https://doi.org/10.1016/0191-8869(96)00033-5) (1996).
22. Furnham, A., Moutafi, J. & Crump, J. The relationship between the revised neo-personality inventory and the myers-briggs type indicator. *Social Behavior and Personality: an international journal* **31**, 577–584. <https://doi.org/10.2224/sbp.2003.31.6.577> (2003).
23. Kerwin, P. L. Creating clarity: addressing misconceptions about the mbti assessment, [white paper] from the myers-briggs company (2018). <https://www.themyersbriggs.com/en-US/Resources/Creating-Clarity-Addressing-Misconceptions-of-MBTI>.
24. Lake, C. J., Carlson, J., Rose, A. & Chlevin-Thiele, C. Trust in name brand assessments: The case of the myers-briggs type indicator. *The Psychologist-Manager Journal* **22**, 91. <https://doi.org/10.1037/mgr0000086> (2019).
25. Choong, E. J. & Varathan, K. D. Predicting judging-perceiving of Myers-Briggs type indicator (MBTI) in online social forum. *PeerJ* **9**, <https://doi.org/10.7717/peerj.11382> (2021).
26. Jung, C. G. *Psychological type* (H. G. Baynes, Trans. London: Kegan Paul, 1921).
27. Randall, K., Isaacson, M. & Ciro, C. Validity and reliability of the myers-briggs personality type indicator: A systematic review and meta-analysis. *Journal of Best Practices in Health Professions Diversity* **10**, 1–27 (2017). <https://www.jstor.org/stable/26554264>.
28. Amirhosseini, M. H. & Kazemian, H. Machine learning approach to personality type prediction based on the myers-briggs type indicator. *Multimodal Technologies and Interaction* **4**, <https://doi.org/10.3390/mti4010009> (2020).
29. Kaggle Personality Cafe MBTI dataset. <https://www.kaggle.com/datasets/datasnaek/mbti-type>.
30. Discussion forum of the Personality Cafe website. <https://www.personalitycafe.com/>.
31. Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Annals of statistics* 1189–1232, <https://doi.org/10.1214/aos/1013203451> (2001).
32. Freund, Y. & Schapire, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* **55**, 119–139. <https://doi.org/10.1006/jcss.1997.1504> (1997).
33. Natekin, A. & Knoll, A. Gradient boosting machines, a tutorial. *Frontiers in neurorobotics* **7**, 21. <https://doi.org/10.3389/fnbot.2013.00021> (2013).
34. Ramos-Pérez, I., Álvarez Arnaiz-González, Rodríguez, J. J. & García-Orsorio, C. When is resampling beneficial for feature selection with imbalanced wide data? *Expert Systems with Applications* **188**, 116015, <https://doi.org/10.1016/j.eswa.2021.116015> (2022).
35. Joloudari, J. H., Marefat, A., Nematollahi, M. A., Oyelere, S. S. & Hussain, S. Effective class-imbalance learning based on smote and convolutional neural networks. *Applied Sciences* **13**, <https://doi.org/10.3390/app13064006> (2023).
36. Wang, S., Dai, Y., Shen, J. & Xuan, J. Research on expansion and classification of imbalanced data based on SMOTE algorithm. *Sci. Rep.* **11**, 24039. <https://doi.org/10.1038/s41598-021-03430-5> (2021).
37. Thorndike, R. L. Who belongs in the family?. *Psychometrika* **18**, 267–276 (1953).
38. Googletrans python library. Python Package Index. <https://pypi.org/project/googletrans/> (2015).
39. Plank, B. & Hovy, D. Personality traits on twitter—or—how to get 1,500 personality tests in a week. In *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, <https://doi.org/10.18653/v1/W15-2913> (Association for Computational Linguistics, Stroudsburg, PA, USA, 2015).
40. Goby, V. P. Personality and online/offline choices: MBTI profiles and favored communication modes in a singapore study. *Cyberpsychol. Behav.* **9**, 5–13. <https://doi.org/10.1089/cpb.2006.9.5> (2006).
41. Li, C. *et al.* Feature extraction from social media posts for psychometric typing of participants. In *Augmented Cognition: Intelligent Technologies*, Lecture notes in computer science, 267–286, https://doi.org/10.1007%2F978-3-319-91470-1_23 (Springer International Publishing, Cham, 2018).
42. Arnoux, P.-H. *et al.* 25 tweets to know you: A new model to predict personality with social media. **11**, 472–475, <https://doi.org/10.1609/icwsm.v11i1.14963> (Association for the Advancement of Artificial Intelligence (AAAI), 2017).
43. Word-counter.com. How many words in a tweet? <https://word-counter.com/how-many-words-in-a-tweet-a-word-count-guide/> (2022).

Author contributions

Luisa Stracqualursi and Patrizia Agati contributed equally to this work. Particularly, Luisa Stracqualursi wrote the main text of the manuscript, contributed to the formal analysis, conceptualization, investigation and creation of the software in Python. Patrizia Agati contributed to the formalization of the applied methodology, to the data management, to the preparation of the figures, to the supervision and to the revision and modification of the manuscript.

Declarations

Competing interests

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-85183-z>.

Correspondence and requests for materials should be addressed to L.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025