KTH Royal Institute of Technology

Research methodology and scientific writing
II2202

Extraction of user's behavioural insights from social media

STEFANO PERENZONI Hiranmayi Pechetti
stefano.perenzoni | pechetti.hiranmayi @gmail.com

January 15, 2021

Fall semester 2020

Abstract

Personality type refers to the psychological description of various groups of persons. Also, recognizing these traits helps one develop a personal development plan, create a tailored career improvement plan, eliminate and settle disputes, aid in the recruiting process, and much more. The Myers-Briggs Type Indicator (MBTI) is a personality model that measures psychological interests on how people perceive the world and make decisions. Generally, we need to fill out detailed questionnaires to assign personality traits. Nevertheless, by studying the data from a person's social media, we can identify their personality traits. Machine-learning algorithms can be used to predict a person's personality type from the data of their social media. Initially, this paper offers a brief historical analysis of previous studies undertaken on this subject and then provides a research survey in predicting psychological types from social media evidence. It proceeds from a study of Twitter data with the proposal of a method to predict psychological styles. From an annotated dataset containing 490 Italian Twitter profiles, we tried different approaches to compute the classification. In particular, first, we classified every user using aggregation of their activities on the social as features. We later classified each activity singularly and merged the partial results to generate the final insights at the user level. We concluded saying that the approach we proposed represent a good alternative to the state of the art since it allows both good performance and a higher number of applications

Contents

List of Acronyms and Abbreviations

BIG5 Big Five Personality traits

CLiPS Computational Linguistic & Psycholinguistic

GDPR General Data Protection Regulation

KNN K-nearest neighbours

LIWC Linguistic Inquire and Word Count

MBTI Myers-Briggs Type Indicator

NLP Natural Language Processing

NLPLAB Natural Language Processing Laboratory

SVM Support Vector Machine

1 Introduction

According to neuroscientists Adelstein et al., personality describes human behavioural responses to wide classes of external stimuli [1]. It works as an adaptive system for taking in, organizing information and driving the response to inner and outside demands [2]. The parameters of the adaptive system represent the variation of the same from person to person and, therefore, characterize uniquely every individual. These parameters are also referred to as personality traits in several different personality models studied over the years. Each model includes its range of traits which combinations describe several personality types. Researchers have shown clear connections between general personality traits and many types of behaviour.

Some fundamental traits describe the type of relationship a person has with the outside world and the way he or she communicates [3]. Thus, to facilitate communication, recently, businesses are using personality models to gain a better understanding of what drives the interests of a person. This approach is showing clear benefits in many different applications. In the field of Human-Computer Interaction, users prefer interfaces designed to represent personalities that most closely matched their own [4]. Some studies have also suggested connections between customer personality and marketing. Through techniques more focused on the target audience, it is possible to profile individuals, and tailor advertisement automatically displayed based on their personality [5]. Therefore, the ability to identify people's personality or, even better, details of their personality traits through well-defined models is a significant competitive advantage since we would have a precise representation of the customer's reasoning process.

Using a personality model to catch these behavioural aspects, the extraction of personality from social media activities is a *machine learning* problem. Precisely, with a categorical model, such as the Myers-Briggs Type Indicator (MBTI), it consists of numerous classification tasks, one for each variable of the taxonomy.

A classification task has the goal of assigning a belonging class to a given object. The input is composed by a tuple of *features* that characterize the object, usually made by numbers, and the output is a categorical variable, such as a "yes/no" label. In other words, it can be seen as a mathematical function, that maps a vector $x \in \mathbb{R}^n$ to an answer $y \in C$

$$f : \mathbb{R}^n \to C$$
$$f : x \mapsto y$$

where C is a set of possible categories. For example, in one of the four classifiers for this problem, x represents a user and her activities on the social media, and $C = \{\texttt{Introvert}, \texttt{Extrovert}\}$

1.1 Myers-Briggs Type Indicator

The Myers-Briggs model, also called *Myers-Briggs Type Indicator, or MBTI*, is the most common alternative to the Big-Five model. Contrarily to the former, there are discussions about the MBTI and its limitations in reflecting the whole personality system. Boyle and Barbuto are two of the scientists that presented a number of psychometric limitations pertaining to the validity and reliability of this model [6, 7]. However, many of their arguments have been proved wrong by Furnham who demonstrated several correlations between the dimensions defined by Myers and the big five factors [8].

The MBTI is a categorical model, based on the conceptual theory of Jung and developed by Katharine Briggs and Isabel Myers who used four different dichotomies to evaluate the personality of people [9]. A first one differentiates a person's attitude in either extraversion (E) or introversion (I). These two preferences describe if one focusses on external stimuli, such as action and interaction with other people or internal ones like self-reflection. Two perceiving functions, sensation (S) and intuition (N) describe the process of gathering new information. On the one hand, people who trust tangible and concrete facts; on the other hand, those who tend to find patterns and meaning also regarding future possibilities. The third cognitive function is that of decision-making which can be thinking (T) or feeling (F). While thinkers make reasonable and consistent choices and reflect over consequences applying a rigid set of rules, feelers tend to emphasize with the situation considering the needs of people involved Finally, there is the lifestyle preference function dichotomy, judging (J) or perceiving (P). Judging types like the outside world to be structured; according to Myers, they prefer to "have matters settled". On the contrary, perceiving personalities like it flexible and spontaneous and tend to "keep decisions open" [10]. There are 16 different types of personality given by the combination of these 4 cognitive functions identified by 4-characters codes such as "INFJ" or "ENFP".

1.2 Research questions

As Section 2 will explain in details, this paper differs from previous studies and make a step further compared to the actual state of the art because of two reasons.

First of all, all the previous studies did not explore different techniques that can be used to organize the data. For example, instead of treating the online profiles as a single and unique document to classify, an alternative is classifying singular activities or group of activities and then put together the partial results to generate the final insights about the user. The application of this, or others, methodologies rather than the standard one can results in better results both for classification accuracy and result's flexibility. Indeed, it would be easier and effortless remove from the final results those activities that do not satisfy specific filters. For example, keeping into consideration only activities that talk about sports.

Secondly, no research in the current literature takes into consideration different components of the online activities at the same time. Indeed, either text components or multimedia files are used for the classification of the personality types. We explore and include different aspects into the final classification.

So, the question this research answers is: *can the accuracy of personality extraction through social media be improved using alternative techniques for the feature extraction and the data organization?*

1.3 Aims and goals

Aim of this study is to generate models through alternative techniques and compare their result. The MBTI personality model will be used to extract the personality type of users. So, different classifiers will be implemented, trained and tested with different setups. Starting from those introduced by the current state of the art, such as Support Vector Machines (SVMs), decision trees and naive Bayes classifiers, we will apply various classification algorithms.

Finally, we will develop a system that allows the profiling of behavioural characteristics through multiple social media profiles of a person.

2 State of the art

This section presents the current state of the art regarding behavioural traits extraction on social media.

It is a common practice inferring behaviour through a variety of personality models. Two of the most important ones are the Big Five Personality traits (BIG5)[11] and the MBTI[9] which divide the personality type of an individual respectively in 5 and 4 different traits. If the first one is used, four different classification problems have to be solved, one for each of the four categorical traits. Similarly, for the latter, personality extraction can be decomposed into five regression tasks because of the numerical continuous nature of the model[12]. Even though, Sumner et al. experimented a binary classification for each aspect using as classes the two extremes of the trait[13].

Many different social networks have been explored as well as many of their components including, for example, text, images and social interactions. However, many of the systems proposed for social media analysis use as fundamental component features of the online activities that describe interactions between users, such as the number of followers, mentions, likes, and comments.

Almost all models presented work on social user composed by the totality, or a portion, of their timeline rather than single activities since linguistic information contained by a single short activity is not enough to accurately predict personality aspects[14]. It means that all the features extracted by the different online activities of a user are put together before the actual classification of the personality type. We think that this is limiting and trying different approaches could bring to improved performances. These approaches were briefly described in Section 1.2.

The feature extraction shares some fundamental aspects in the majority of systems. Research has shown a strong correlation between discussed topics and personality aspects of a person[15]. Guntuku et al. proved that studying semantic concepts contained in posted images can give a significant performance gain in predicting personality traits with respect to the BIG5[16]. However, the literature contains a very few number of proposals that considered the content of the activity and are usually confined to hashtags and keywords in the text[17].

Regarding features that describe the social presence of a person. These are usually included by the majority of models. Although some are limited to basic information such as the number of followers, following or friends, the number of activities, and their frequency[18]. Over time, the literature presented the application of more complex features, obtained as results from further analysis of the user's network such as interaction patterns by a person towards the author of the post[19]. For example, significative patterns could be a high retweet ratio by users who do not retweet many other sources by or an elevate number of interactions by users with many followers. However, these last observations need the permission of each person belonging to the analysed network to be respectful of the General Data Protection Regulation (GDPR) requirements. Thus, even though they could give great results, their lawful application in the market is quite intricate.

Then, there is a third fundamental group of features which is probably the most important one. Since psychological studies proved that there is an effective relationship between linguistic style and personality aspects, understanding detailly how an individual writes is a crucial step[20]. Some of the most common and basic features are word counts, sentences per activity, word per sentence, and punctuation count. These have been applied by the majority of models with great results in many different environments. For example, Farnadi recognized personality of YouTube vloggers using the script of their videos to extract this linguistic information[21]. Furthermore, more recent studies have tested features from specialized and complex tools for text analysis. These can reveal precisely thoughts, feelings, and motivations of the text's author. The Linguistic Inquire and Word Count (LIWC) developed by Tausczik and Pennebaker is certainly the most used one[22]. Other services that have been tested are the MRC Psycholinguistic Database[1] and the NLPRO, developed by Natural Language Processing Laboratory (NLPLAB)[2][23, 24]. Lima et al. tested the three of them concluding with the first one as the most performing one[25].

[1] https://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm

[2] https://www.ida.liu.se/divisions/hcs/nlplab/

3 Research methodologies

For this paper, we use data that has already been collected and made available by the Computational Linguistic & Psycholinguistic (CLiPS) research center[1].

3.1 Data

The dataset we used is the TwiSty Corpus[26]. It contains a totality of 18168 Twitter profiles. Each is labelled for MBTI personality type and gender. It includes six different languages: Dutch, German, French, Italian, Portuguese and Spanish. For each entry, the ID of the profile and the IDs of their available tweets are provided. For each author, a native language is identified and then tweets are divided into two categories: belonging to the language in which the person is situated and other tweets.

The user annotation was done by looking for self-assessed data about MBTI through TwitterAPI. They looked, for each language, for text containing both a language-specific content word and a MBTI personality type.

Starting from the IDs, we retrieved the activities using the official TwitterAPI through the TweePy library for Python. Then, some general statistics about the data were evaluated. Such as distribution of each trait, distribution of the gender label, average number of tweets, etc. Then, the data were analysed and preprocessed to retrieve significant feature for the following steps. Some of the analysis may include Natural Language Processing (NLP) and network analysis, for example. Finally, the extracted features will be used by the different classifiers for the extraction of the final personality type.

To analyze the data we used Python, Numpy and Pandas as analysis tool. Indeed, as said before, we had to retrieve the actual content of the social profiles starting only from lists of IDs. So, for each user, we used endpoints for downloading both the profiles and the tweets. Some information we retrieved for users includes:

- Name
- Number of followers
- Number of following
- Number of times the users has been listed
- Location
- Description
- Profile picture

Same has been done for the tweets. Here we obtained information such as:

- Text
- Creation date time
- Attachments
- Language
- Type (Normal, reply or quote)
- Number of retweets, likes, replies and quotes

[1] `https://www.clips.uantwerpen.be/clips.bak/datasets/twisty-corpus`

At this point, we have all the information we need to run our analysis. While all the data mentioned before represents the starting point of our independent variables, the dependent features are the MBTI labels and the gender annotation. We are particularly interested in the MBTI one.

We used only the users and their tweets that belonged to the Italian language since we were limited by the TwitterAPI's free plan. It is composed of 490 users. 370 are female while only 170 are male. Among these 490 users, 432 posted more than 500 tweets. The total number of tweets is 932786 and 658332, the 70.58% are confirmed to be in Italian. So, the dataset contains an average of 1904 tweets per user.

The personality information is given through a single annotation characterizing the MBTI personality type. These consist of 4-characters code such as "INFJ" or "ENFP". However, we are interested in every single component of the personality types, also called personality traits. So, we divided the personality annotation into four different labels, one for each trait.

We discovered that not all the four properties are balanced.

E/I	109 (22.2%)	381 (77.8%)
N/S	421 (85.9%)	79 (14.1%)
T/F	227 (46.3%)	263 (53.7%)
J/P	260 (53.1%)	230 (46.9%)
Female/Male	320 (65.3%)	170 (34.7%)

Table 1: Ratio between the MBTI traits

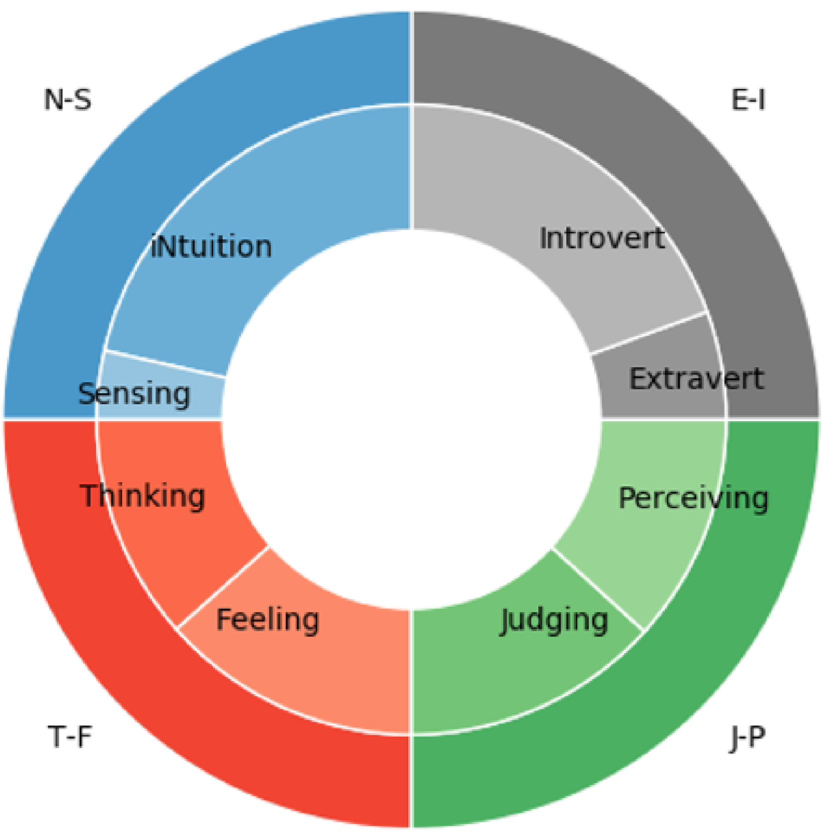

Figure 1: MBTI labels distribution

Indeed, while two of the four properties are well-balanced the remaining two are strongly unbalanced. These last two probably reflect a particular aspect of the Twitter population. Indeed, findings have shown that online communication is easier and more accessible for introverts individuals.

3.2 Ethical issues

The main ethical issues of this research are related to user profiling through Social Media. Personality traits of individuals represent sensitive information that requires us to guarantee anonymity and confidentiality. So, results will be presented in a way that it is not possible to identify data of a specific person.

Profiling of users through their online profiles is a practice that received a lot of criticism. However, we think profiling is ethical since recently it has been used in some good social applications. Indeed, the extraction of personality traits can help companies improve their platforms, communicate to their clients in the most proper way and improve therefore their User Experience.

In Section 5 reasoning is done about users' data and it explains why our solution also represents an improvement regarding this point of view. Indeed, when a system like that is launched into the production environment, a series of principles have to be met and our alternative helps in doing that

3.3 Sustainability

From the 17 SDGs, this study meets three sustainable development goals. The following goals are:

- **Goal 3: Good health and well-being**

Out research incorporates the target "3.D: Improve early warning systems for global health risks", Improving early warning systems for global health risks. We know that mental health issues like depression, anxiety disorders and personality disorders are a major concern these days. It is really important to locate them and take the required measures to support them. Typically, you can learn about the mental health of people from their behaviour. This is where our research can help. With the aid of our study, through their actions and personality on social media sites like Twitter, we can try to understand and deduct the mental health of a person. This will make it easier to help the person out.

- **Goal 4: Quality education**

One of the most important things for society and a community is quality education, and the new education system is doing a wonderful job. But by educating a student based on his personality, we can take the level of education a step higher. This helps us to take into account the characteristics of the student and develop a customized education that maximizes the most of the skill set the student has and brings out the best in them. Our study helps to improve the learning experience by identifying the students' personality through our research. Hence contributing to the target "4.4: Increase the number of people with relevant skills for financial success".

- **Goal 8: Economic growth and decent work**

Every organization would love to see full productivity in the workplace and what better way to do this than to make every employee efficient. We may recognize the traits of an individual and their strengths and weaknesses using our analysis. Using this, we can ensure that they are provided with jobs where they can optimally leverage their skills to have optimum performance. Thus contributing to the target "8.2: Diversify, innovate and upgrade for economic productivity".

4 Modelling

As said before, the contribution of this paper is given by the comparison of different techniques to classify the personality traits. In particular, two approaches have been applied. The first one, already explored from previous studies, composes the profile to classify with the totality of its activities on social media. The idea is to create a single object representing the users' profiles and classify it with respect to the MBTI personality type. Differently, the second approach we tried consisted in classifying singularly each activity and then merge together the results for every user in order to obtain the general personality traits.

Both approaches are developed through four different binary classifiers, one for each personality trait, on top of the same dataset, introduced in Section 3.1. However, different sets of features have been used. These are briefly described in the following sections.

The two techniques were later evaluated and compared on the same metrics.

4.1 Profiles classification

Starting from the initial dataset, we decided to extract fifteen features: number of followers, number of following, number of listed, follower ratio, location, number of words per tweet, number of sentences per tweet, number of characters per word, number of character per sentence, hashtags, mentions, URLs and media usage.

We did not use text representation's features such as bag of words, N-grams or tf-idf. Indeed, while many feature extraction techniques have already been tested by other studies with the clear goal of improving the classification's performance, our research's goal was to compare different techniques in organizing the data and observe potential benefits. So, we decided to limit it to some basic features for both the approaches.

For this first approach, we used six classification algorithms: decision tree, random forest, SVM, gradient boosting, K-nearest neighbours (KNN) and Adaboost. For each algorithm and for each personality trait, a number of different models were trained and validated through parameter tuning and cross-validation.

Each best model was then tested on the test set with respect to four metrics: accuracy, precision, recall and F1-score. The dimension of the test set and the results obtained are reported in Section 5.

4.2 Activities classification

For this second approach, we used fifteen feature. Still, while before we could use information about the social profile, such as the number of followers, following, and listed, this would not be significant for the single activities. So, the two features mentioned before were replaced with the number of likes and retweets the tweet received, and the tweet type. This last variable was added to differentiate between different kinds of tweets: standard, retweet, retweet with a comment, and reply. Again, for the reasons introduces previously, we only used a set of basic features.

Before training the classifiers, we assumed that, for each user, all its features share the same MBTI personality type. So, we took the label of the user and appended it to all the activities belonging to its timeline. Then, we trained different models' configurations for three algorithms: decision tree, random forest, and K-nearest neighbours.

We then tested each best model on the test set to see how well we were able to predict the single activities. However, we are not interested in the results for the single activities but in the ones regarding the users. So, we needed to merge together the single MBTI labels at a profile-level.

We decided to implement a standard algorithm based on majority-vote. So, for each trait, we assigned to the user the class that was more frequent among its activities. In the extreme case, the two labels had the same number of occurrences, one was chosen arbitrarily. Note that this case never happened for the 490 profiles in our dataset. We implemented only this version of the merging algorithms even though many others could be applied. More complex alternatives could weight the vote of every activity taking into consideration the length of the same, the timestamp, or other variables.

5 Results

5.1 Expected results

We expect that implementing models with different setups with respect to how data are organized and classified will result in interesting insights that could help to reach the best performances in terms of result's accuracy.

Also, as introduced in Section 1.2 our research adds a new layer of flexibility to the results. Indeed, organizing each profile in various ways, instead of treating it as a single unique document composed by all its activities, lets us implement different filters in order to include in the evaluation only relevant activities with respect to parameters such as the number of characters, activity's content, multimedia usage, etc.

5.2 Results of profiles classification

Since we are using the MBTI model, which is a categorical model. The problem has been decomposed into four classification problems, one with each personality type. Each classifier is responsible for one of the MBTI pairs. We ran the same classification algorithms on all classifiers. The measures adopted to evaluate the classifiers are the accuracy per class(percentage of correct classification per class), Precision (also called positive predictive values), Recall (sensitivity), F1 score (F).

All the models were trained implemented using the scikit-learn library for Python. The initial dataset has been split into a train set and a test one where the latter contains 30% of the total activities.

Table 2 summarizes the results of the Twitter attributes' evaluation for E/I Classification. The F1 score has a value of more than 80 % for all the models we worked on. Decision Tree Classification, SVM, KNN achieved high accuracy. Among these models, Decision Tree classification has better performance, mainly when we compare the Precision value.

Algorithm	Accuracy	Precision	Recall	F1 score
Decision Tree classification	77.57	88.75	100.0	87.36
Random Forest Classification	76.63	77.88	97.59	86.63
SVM	77.57	77.57	100.0	87.37
XGBOOST	74.77	88.0	93.97	85.24
K-nearest Neighbours	77.57	77.57	100.0	87.36
ADABOOST	73.83	89.1	94.7	84.78

Table 2: Accuracy, Precision, Recall, F1 score for E/I with 15 features

Table 3 summarizes the results of the Twitter attributes' evaluation for S/N Classification. Decision Tree Classification, SVM, K-nearest Neighbours achieved high accuracy. Decision Tree classification has obtained the best F1 score.

Algorithm	Accuracy	Precision	Recall	F1 score
Decision Tree classification	88.78	88.78	100.0	94.06
Random Forest Classification	87.85	89.42	97.89	93.46
SVM	88.78	88.78	100.0	94.05
XGBOOST	82.24	88.0	92.63	90.25
K-nearest Neighbours	88.78	88.78	100.0	94.05
ADABOOST	85.04	89.10	94.73	91.83

Table 3: Accuracy, Precision, Recall, F1 score for S/N with 15 features

Table 4 summarizes the results of the Twitter attributes' evaluation for T/F Classification. Decision Tree classification has achieved the highest accuracy of 63.55 % . SVM also had a good performance, mainly when we compared F1 scores

Algorithm	Accuracy	Precision	Recall	F1 score
Decision Tree classification	63.55	72.0	59.016	64.86
Random Forest Classification	56.07	61.66	60.65	61.15
SVM	56.07	56.60	98.36	71.85
XGBOOST	55.14	61.81	55.73	58.62
K-nearest Neighbours	62.61	69.09	62.29	65.51
ADABOOST	58.87	64.40	62.29	63.33

Table 4: Accuracy, Precision, Recall, F1 score for T/F with 15 features

Table 5 summarizes the results of the Twitter attributes' evaluation for J/P classification. We have obtained the highest accuracy of 56% when we used SVM. SVM has also shown good performance when comparing the F1 scores.

Algorithm	Accuracy	Precision	Recall	F1 score
Decision Tree classification	48.59	53.06	44.82	48.59
Random Forest Classification	55.14	58.92	56.89	57.89
SVM	56.07	55.44	96.55	70.44
XGBOOST	47.66	51.78	50.0	50.87
K-nearest Neighbours	52.33	59.45	37.93	46.31
ADABOOST	52.33	59.45	37.93	46.31

Table 5: Accuracy, Precision, Recall, F1 score for J/P with 15 features

The accuracies obtained for the classification of T/F and J/P are low. Given these bad results, these traits cannot be predicted with good reliability and for this reason we will be trying different features such as n-gram, bags of words,tf-idf.

5.3 Results of activities classification

Regarding the classification of the activities, we followed the same process applied before. We also used the same performance metrics. The initial dataset was split as follows. Firstly, we portioned the users into two sets; a train one and a test one that we would have used to estimate the performance of the final merged results. Then, we used the activities in the train set to train and evaluate the models with respect to the classification of the tweets. The number of activities included in this set was 325788. Again, we applied a 30% split in order to validate the models on an independent test set. After choosing the best configurations, we use them to retrain the models on the totality of the activities in the train set. We then used these models to predict the labels for the activities in the user test set which were then merged into a single result for every user. Finally, we compared the obtained results with the actual labels.

Next, the obtained results with respect to the classification of single activities.

Algorithm	Accuracy	Precision	Recall	F1 score
Decision Tree classification	70.02	78.998	82.48	80.70
Random Forest Classification	75.01	78.898	91.65	75.02
K-nearest Neighbours	71.44	76.08	89.81	82.38

Table 6: Accuracy, Precision, Recall, F1 score on activities for E/I with 15 features

Algorithm	Accuracy	Precision	Recall	F1 score
Decision Tree classification	78.80	86.78	88.90	87.81
Random Forest Classification	84.10	86.568	96.48	91.25
K-nearest Neighbours	82.54	84.94	96.45	90.33

Table 7: Accuracy, Precision, Recall, F1 score on activities for S/N with 15 features

Algorithm	Accuracy	Precision	Recall	F1 score
Decision Tree classification	54.84	53.45	56.34	53.40
Random Forest Classification	58.45	55.018	57.16	59.81
K-nearest Neighbours	53.24	56.27	57.14	56.70

Table 8: Accuracy, Precision, Recall, F1 score on activities for T/F with 15 features

Algorithm	Accuracy	Precision	Recall	F1 score
Decision Tree classification	52.84	53.22	54.73	51.08
Random Forest Classification	56.29	54.598	55.61	56.98
K-nearest Neighbours	52.80	54.01	55.44	54.71

Table 9: Accuracy, Precision, Recall, F1 score on activities for J/P with 15 features

The obtained results are similar to the ones obtained before. Once again, we can observe that the dichotomies Extrovert/Introvert and Sensation/iNtuition can be predicted with high reliability. On the other hand, Thinking/Feeling and Judging/Perceiving offers really poor performances. In general, random forests perform the best over-performing both the decision trees and the K-nearest neighbours.

As said before, what we are really interested in is the results we are able to obtain after merging together the single activities through the majority-vote algorithm. Again, we estimated the performance for each of the three algorithms.

Table 10 displays the results for the merged profiles with respect to the trait E/I. Surprisingly, we can immediately observe that the best general performance is obtained using decision trees. It is interesting to observe, in table 6, that Decision trees generally performed worse than Random forests on this trait. This could suggest that F1 score represents a significant metric on the single activities.

Algorithm	Accuracy	Precision	Recall	F1 score
Decision Tree classification	80.38	81.63	96.39	88.40
Random Forest Classification	76.63	77.368	98.80	86.77
K-nearest Neighbours	77.57	77.57	100	87.37

Table 10: Accuracy, Precision, Recall, F1 score on merged activities for E/I with 15 features

Table 11 regards the trait S/N. It is not very insightful since all the models performed exactly the same. This is a consequence of the highly unbalance regarding this trait, as pointed out in section 3.1. Indeed, the perfect value. for recall suggests that every profile is classified as *iNtuitive*. Because of that, this result is not significant and more tests should be run or different metrics observed. For example, a possible future improvement could be balancing the trait to see how the performances are affected.

Regarding the traits T/F and J/P, respectively in Table 12 and Table 13, KNN is the best performer. For the former, KNN clearly over-performs the other two models for what regards recall and F1 score. The latter is instead a close call between KNN and random forest. There is not a clear pattern between the results on the single activities and the merged profiles. Indeed, KNN is not the best model when tested on the single tweets for these two traits. So, there is not an evident correlation between the performance of classifiers of the single activities and that of the combined profile.

Algorithm	Accuracy	Precision	Recall	F1 score
Decision Tree classification	88.79	88.79	100	94.06
Random Forest Classification	88.79	88.79	100	94.06
K-nearest Neighbours	88.79	88.79	100	94.06

Table 11: Accuracy, Precision, Recall, F1 score on merged activities for S/N with 15 features

Algorithm	Accuracy	Precision	Recall	F1 score
Decision Tree classification	55.14	65.12	45.90	53.85
Random Forest Classification	57.94	60.00	78.69	68.09
K-nearest Neighbours	61.68	61.36	88.52	72.48

Table 12: Accuracy, Precision, Recall, F1 score on merged activities for T/F with 15 features

Algorithm	Accuracy	Precision	Recall	F1 score
Decision Tree classification	47.66	54.54	20.69	30.00
Random Forest Classification	57.01	59.09	67.24	62.90
K-nearest Neighbours	56.07	57.33	74.14	64.66

Table 13: Accuracy, Precision, Recall, F1 score on merged activities for J/P with 15 features

5.4 Comparison and discussion

To perform the last comparison between the two techniques, we decided to choose, for every trait, the models with the best performance overall. Table 14 shows how this changed going from the first approach to the second one. We can observe that, in general, the variation fluctuates. For some combinations, an improvement can be observed while for some others the performance gets worse. As said before, there is not a clear pattern that suggests us in which cases one approach is better than the other one.

Algorithm	Accuracy	Precision	Recall	F1 score
E/I	+3.81	-7.12	-3.61	+1.04
S/N	+0.00	+0.00	+0.00	+0.00
T/F	+5.61	+4.76	-9.84	+2.04
J/P	+0.00	+1.89	-22.41	-5.78

Table 14: Accuracy, Precision, Recall, F1 variation for each trait

In general, besides a couple of cases, the performances either improved or decreased. So, for some traits, we observed the kind of change we were expecting. However, the benefits given by the different approach we proposed are not limited to an increase in the results' reliability.

The solution we proposed adds an extra flexibility layer to the process of personality extraction. Indeed, classifying the single activities would allow choosing which tweets should be included in the general results and which not. So, we would be able to filter the activities regarding different parameters (e.g. length, argument discussed, date) in order to extract the personality traits considering only a subset of the user's timeline.

Then, it also is more GDPR compliant since it would help meet some principles of the European regulation such as data minimisation, data accuracy, and storage limitation. Indeed, once an activity is classified and the result stored together with any possible filter information, its content can be deleted.

Then, the solution we proposed allows implementing an incremental system where new activities on social media can be downloaded, classified and then merged together with the previous ones. In this way, the system does not need to classify the whole profile every time we desire to update the results for a specific user.

6 Conclusions

To sum up, the report answers the initial research question. To do it, we used the MBTI personality model and we applied an approach based on the singular activities. So, instead of classifying the whole profile, we focused on labelling the tweets, for each trait of the model. An annotated dataset composed by Twitter's user was used. It provided six different languages, we limited our study to the Italian one.

After using the official Twitter API to retrieve the profiles' content we applied the two approaches. For the reasons explained before, only a limited number of features was used. Regarding the activities-based approach, we implemented a majority-vote algorithm to merge together the results. Finally, we compared the two solutions through some performance metrics.

To conclude, we think our research can represent a clear starting point for future studies that want to explore alternative solutions for the extraction of personality traits from social media.

6.1 Future work

Finally, there are some aspects that can be improved and some others that can be added. Some of them have already been mentioned in the paper. First of all, new algorithms that take into account weights should be evaluated for what regards merging the single activities' results together. Then, the contribution of this research could be applied together with previous studies in order to develop more complex and accurate models. These would include more compound features related, in particular, to the representation of tweets' text. Also, we decided to limit this research to only one social network, Twitter, adding new platforms could reveal interesting insights. Indeed, other social media such as Facebook or Instagram are built on different principles and they could provide new data in order to improve our models. Then, multi-language support can be implemented. We already mentioned, in Section 3.1, the other five languages we have data for. This could be used to train and evaluate new classifiers.

References

[1] J. S. Adelstein, Z. Shehzad, M. Mennes, C. G. DeYoung, X.-N. Zuo, C. Kelly, D. S. Margulies, A. Bloomfield, J. R. Gray, F. X. Castellanos *et al.*, "Personality is reflected in the brain's instrinsic functional architecture," *PloS one*, 2011.

[2] J. Block, *Personality as an affect-processing system: Toward an integrative theory.* Psychology Press, 2002.

[3] A. C. E. Lima and L. N. de Castro, "Predicting temperament from twitter data," in *2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI).* IEEE, 2016, pp. 599–604.

[4] C. Nass and K. M. Lee, "Does computer-generated speech manifest personality? an experimental test of similarity-attraction," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 2000, pp. 329–336.

[5] Y. Bachrach, M. Kosinski, T. Graepel, P. Kohli, and D. Stillwell, "Personality and patterns of facebook usage," in *Proceedings of the 4th annual ACM web science conference*, 2012, pp. 24–32.

[6] G. J. Boyle, "Myers-briggs type indicator (mbti): some psychometric limitations," *Australian Psychologist*, vol. 30, no. 1, pp. 71–74, 1995.

[7] J. E. Barbuto Jr, "A critique of the myers-briggs type indicator and its operationalization of carl jung's psychological types," *Psychological Reports*, vol. 80, no. 2, pp. 611–625, 1997.

[8] A. Furnham, "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*, vol. 21, no. 2, pp. 303–307, 1996.

[9] C. G. Jung, 1875-1961, "Personality types," *The portable Jung*, pp. 178–272, 1971.

[10] I. B. Myers and P. B. Myers, *Gifts differing: Understanding personality type.* Nicholas Brealey, 2010.

[11] B. De Raad, *The Big Five Personality Factors: The psycholexical approach to personality.* Hogrefe & Huber Publishers, 01 2000. ISBN 0-88937-236-5

[12] M. Kosinski, D. Stillwell, and T. Graepel, "Private traits and attributes are predictable from digital records of human behavior," *Proceedings of the national academy of sciences*, vol. 110, no. 15, pp. 5802–5805, 2013.

[13] C. Sumner, A. Byers, R. Boochever, and G. J. Park, "Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets," in *2012 11th International Conference on Machine Learning and Applications*, vol. 2. IEEE, 2012, pp. 386–393.

[14] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore, "Using linguistic cues for the automatic recognition of personality in conversation and text," *Journal of artificial intelligence research*, vol. 30, pp. 457–500, 2007.

[15] M. L. Kern, G. Park, J. C. Eichstaedt, H. A. Schwartz, M. Sap, L. K. Smith, and L. H. Ungar, "Gaining insights from social media language: Methodologies and challenges." *Psychological methods*, vol. 21, no. 4, p. 507, 2016.

[16] S. C. Guntuku, W. Lin, J. Carpenter, W. K. Ng, L. H. Ungar, and D. Preoţiuc-Pietro, "Studying personality through the content of posted and liked images on twitter," in *Proceedings of the 2017 ACM on web science conference*, 2017, pp. 223–227.

[17] X. Ruan, S. Wilson, and R. Mihalcea, "Finding optimists and pessimists on twitter," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2016, pp. 320–325.

[18] D. Quercia, M. Kosinski, D. Stillwell, and J. Crowcroft, "Our twitter profiles, our selves: Predicting personality with twitter," in *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*. IEEE, 2011, pp. 180–185.

[19] T. Dickinson, M. Fernandez, L. A. Thomas, P. Mulholland, P. Briggs, and H. Alani, "Identifying prominent life events on twitter," in *Proceedings of the 8th International Conference on Knowledge Capture*, 2015, pp. 1–8.

[20] J. W. Pennebaker and L. A. King, "Linguistic styles: Language use as an individual difference." *Journal of personality and social psychology*, vol. 77, no. 6, 1999.

[21] G. Farnadi, S. Sushmita, G. Sitaraman, N. Ton, M. De Cock, and S. Davalos, "A multivariate regression approach to personality impression recognition of vloggers," in *Proceedings of the 2014 ACM Multi Media on Workshop on Computational Personality Recognition*, 2014, pp. 1–6.

[22] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: Liwc and computerized text analysis methods," *Journal of language and social psychology*, vol. 29, no. 1, pp. 24–54, 2010.

[23] M. Wilson, "Mrc psycholinguistic database: Machine-usable dictionary, version 2.00," *Behavior research methods, instruments, & computers*, vol. 20, no. 1, pp. 6–10, 1988.

[24] R. Tsarfaty, "The natural language programming (nlpro) project: Turning text into executable code." in *REFSQ Workshops*, 2018.

[25] A. C. E. Lima and L. N. de Castro, "Tecla: A temperament and psychological type prediction framework from twitter data," *Plos one*, vol. 14, no. 3, 2019.

[26] B. Verhoeven, W. Daelemans, and B. Plank, "Twisty: a multilingual twitter stylometry corpus for gender and personality profiling," in *Proceedings of the 10th Annual Conference on Language Resources and Evaluation (LREC 2016)/Calzolari, Nicoletta [edit.]; et al.*, 2016, pp. 1–6.