

Original scientific paper

UDC:
077.5-057.875:159.923.075(64)

Received: September 24, 2023.

Revised: October 30, 2023.

Accepted: November 08, 2023.

 [10.23947/2334-8496-2023-11-3-389-400](https://doi.org/10.23947/2334-8496-2023-11-3-389-400)



Using Students' Digital Written Text in Moroccan Dialect For The Detection of Student Personality Factors

Nisserine El Bahri¹, Zakaria Itahriouan², Anouar Abtoy¹, Samir Brahim Belhaouari³

¹Abdelmalek Essadi University, Tetouan, Morocco, e-mail: nisserine.elbahri@etu.uae.ac.ma, anouar.abtoy@uae.ma

²Moulay Ismail University, Meknes, Morocco; Private University of Fez, Morocco, e-mail: itahriouan@upf.ac.ma

³Hamad Bin Khalifa University, Doha, Qatar, e-mail: sbelhaouari@hbku.edu.qa

Abstract: In the contemporary digital era, social media platforms have a big influence on students' lives. They use these platforms for self-expression, opinion sharing, and experience reporting (writing or sharing videos or photos about personal experiences) in addition to social interaction. Education professionals and academics may get valuable insights into students' thoughts, sentiments, interests, academic success, and even personalities by studying their writing on social media. We can improve our teaching, enhance students' social and emotional development, and create a more engaging learning environment if we have a better knowledge of the student. The purpose of this study is to ascertain whether or not students interact with classmates and other participants in learning platforms in a way that accurately represents their personalities. Data from a sample of students at Abdelmalek Essaadi University of Tetouan were collected from various social media learning environments for the experimental investigation presented in this work, and Symanto AI-based personality tool was used to assess the data. The Big Five Questionnaire was then utilized to assess the personalities of the same students, and the findings were compared to the personality traits discovered by the AI-based approach. The study has shown that the AI based tool has correctly predicted the personality traits of 7 students out of 10 with a correlation of about 0,9 which means that social media-based learning environments can be used by institutions to understand the personality of the student. This paper also gives recommendations about data for obtaining good quality in personality prediction.

Keywords: FFM personalities, social media learning environment, Moroccan dialect text.

Introduction

Student written text plays a special and dynamic role in learning environments that uses social media. Actually, students use social media platforms like (Instagram, WhatsApp and Twitter) more and more, and there are many advantages of using an online learning environment for educational reasons (Christine Greenhow et al., 2019). Those platforms enable informal and spontaneous conversation between students. Therefore, they frequently post variety of content, and they receive in return quick feedback and interactions with other students and teachers. Among the advantage of this trend, these interactions encourages participation and teamwork in students, which can improve their learning process (Josué et al., 2023). Moreover, unlike standard academic writing, students may present ideas, queries, or observations in a less structured manner. As a result, this informal setting may foster individualism, creativity, and the exchange of different perspectives (Eysenck, 1994).

Students' writing in social media often reflects their ideas, feelings, interests, and ways of interacting, which can provide interesting insights about them. Thus, based on these writings, it is possible to determine their personalities on social media using a variety of methods, including: content and interests, language and communication style, frequency and consistency, tone and emotions, interaction patterns, etc... (Rahman et al., 2019).

To understand individual's personality many approaches and theories are used. The following list presents the most well-known models:

- The Myers-Briggs Type Indicator (MBTI) (Pittenger, 1993; Tlili et al., 2016): categorizes persons

¹Corresponding author: nisserine.elbahri@etu.uae.ac.ma



into 16 personality types based on four dichotomies which are extraversion/introversion, sensing/intuition, thinking/feeling, and judging/perceiving. Every combination of these characteristics results in a distinct personality type, such as INFP (Introverted, Intuitive, Feeling, Perceiving) or ESTJ (Extraverted, Sensing, Thinking, Judging).

- The Big Five (BF) Personality Traits (Caprara et al., 1993; Eysenck, 1994): Also known as the Five Factor Model (FFM). According to this Model, there are five different dimensions of personality: Agreeableness, Conscientiousness, Extraversion, Openness, and Neuroticism (Utami, Maharani and Atastina, 2021). It is frequently employed in organizational and behavioral studies as well as psychology research, offers a thorough framework for evaluating and characterizing personality traits. It is considered that these characteristics sum up the fundamental elements of human behavior and personality.

- Freudian Personality Structure (Bronfenbrenner, 1951; Zhang, 2020): The famous Austrian psychotherapist Sigmund Freud developed a theory of personality structure known as the Freudian Personality Structure. According to Freud, there are three primary parts of the human mind (Id, Ego and SuperEgo).

- Hans Eysenck's model (Eysenck, 1991, 1981): The prominent psychologist Hans Eysenck developed a widely known three-dimensional model of personality: Neuroticism/Emotional Stability, Extraversion/Introversion, and Psychoticism. The field of personality psychology has been greatly impacted by Eysenck's work, and his model has been widely applied in studies and personality evaluation.

- DISC (Dominance, influence, Steadiness and Conscientiousness) personality Model (Sugerman, 2009; Utami et al., 2022): Another psychological theory for understanding and classifying human behavior in diverse contexts is the DISC Personality Model. It categorizes people into four main personality traits, denoted by a different letter in the acronym DISC. This model is frequently applied in work environments and interpersonal interactions in order to foster better understanding, cooperation, and communication among people with diverse personality types.

This work considers the Big Five model as a method of analyzing the personality of students out of all the models previously provided since it is the most widely used model and particularly because the AI algorithm for personality detection from text is built on it.

Through the Automated Text-Based Personality Assessment (ATBPA) (Gjurković, Vukojević and Šnajder, 2022), artificial intelligence (AI) may predict personality from text by using well-established psychological models. These latter can determine a person's personality traits from written content through analyzing the writing styles, linguistic patterns, word choices, etc.... (Christian et al., 2021). The AI models are trained using machine learning algorithms including text classification and natural language processing. The training of the model uses mainly the annotated data from a dataset. Therefore, the model acquires the ability to identify patterns and connections between linguistic features and personality traits (Gjurković, Vukojević and Šnajder, 2022).

The main goal of this paper is to use one of these ATBPA tools to identify students' personalities based on their writing in the Moroccan dialect in social media learning environments. For this study, we have chosen Symanto APIs as a tool. To achieve our goal, we have gathered data from students in various social learning environments (Instagram, Twitter, WhatsApp and Google chat). This data has been preprocessed by removing irrelevant information and then translating it into English. Subsequently, it has been processed by the AI-based personality algorithm to predict students' personality traits. Finally, the students' personality predictions obtained by the algorithm were compared to the Big Five Questionnaire results that were gathered from the same students.

The following section presents a summary of the literature on the use of social media learning environment. In section 3, we explain the methodology and the applied data processing approach. Subsequently, the experiment's findings are presented in Part 4 followed by an analysis of the results and a discussion in Section 5. Finally, the paper ends with a conclusion that summarizes the work and presents the implications for further research.

Social media and education

The term "social media" refers to a modern phenomenon that includes both mobile interaction and web-based communication with internet users via web applications (Wickramanayake and Muhammad Jika, 2018). Thanks to how convenient it is to access these applications, the majority of people utilize social media for a variety of purposes, including recounting experiences, communicating, and sharing stories from their everyday lives. In the case of students, the development of Web 2.0 and the emergence of Web 3.0 have enabled students to produce content, exchange ideas, and share knowledge. This development is definitely igniting a revolution in the world of education (Namaziandost and Nasri, 2019).

There are now numerous social media learning environments which are frequently utilized by our

students. Among these, the most well-liked platforms include WhatsApp, Instagram, Facebook, Wiki, Skype, YouTube, LinkedIn, Blogs, Twitter and Google Chat,... (Swaminathan, Harish and Cherian, 2013). These platforms, according to Lim et al., can be grouped into seven categories: media sharing, text-based, social networking, virtual world and games, synchronous communications, conferencing applications and mashups, and mobile-based application (See Yin Lim et al., 2014). The top 17 social media learning environments in terms of monthly active users are listed in Table 1.

Table 1
The top 17 social media learning platforms (Barrot, 2022).

Social media	Initial release	Monthly active users	Owner	Country of origin
Facebook	2004	2603000000	Facebook, Inc.	United States
WhatsApp	2009	2000000000	Facebook, Inc.	United States
YouTube	2005	2000000000	Google	United States
WeChat	2011	1203000000	Tencent	China
Instagram	2010	1082000000	Facebook, Inc.	United States
Tik Tok	2016	800000000	ByteDance	China
QQ	1999	694000000	Tencent	China
LinkedIn	2002	660000000	Microsoft	United States
Sina Weibo	2009	550000000	Pan Weibo	China
Reddit	2005	430000000	Advance Publications	United States
Kuaishou	2011	400000000	Bijing Kuaishoun Technology Co., Ltd.	China
Tumblr	2007	400000000	Automattic	United States
Snapchat	2011	397000000	Snap, Inc.	United States
Pinterest	2009	367000000	Pinterest, Inc.	United States
Twitter	2006	326000000	Twitter, Inc.	United States
Skype	2003	300000000	Skype Technolgies	United States
MySpace	2003	506000000	Meredith Corporation	United States

Materials and Methods

This study focuses on how interactions in learning contexts reveal students' personalities. In the previous section, we discussed the most common social media environments that are frequently visited by students and how they use them to interact about learning issues. On the other hand, there is a set of AI-based tools that can be used to detect the personality of people. These tools can be used either using Application Programming Interfaces (APIs) or Graphical User Interfaces (GUIs). In this context, we have gathered data from students in the classroom and we have used 'Symanto' (<https://www.symanto.com/>) as one of these AI-powered tools to evaluate their personalities.

Detecting the personality based on the tool may not be enough to confirm that it is really the personality of the learner. Therefore, to assess the accuracy of these latter, we asked the same students who participated in the experiment to answer a Big Five Questionnaire test. The purpose is to compare test results to personality traits predicted by AI. As a final goal, we will then be able to investigate how learning environments may be used to understand students' personalities. The general steps of the experiment are shown in Figure 1.

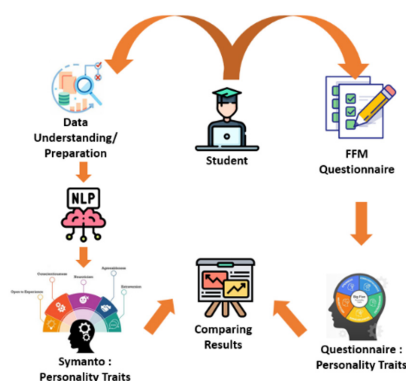


Figure 1. Summary of the student personality comparison and detection technique

Data collection

This research involved ten students of Computer Science Engineering who were enrolled at the National School of Applied Sciences of Tetouan at the Abdelmalek Essaadi University. We gathered their text expressions in different contexts and from multiple social media platforms. Mainly text captured from comments on publications (courses, labs, exercises solutions...), discussions and publications posted by students themselves. The text was integrated and translated before being analyzed by the Machine Learning model. Meanwhile, the same students were asked to respond to the Big Five Questionnaire test.

From numerous social media platforms, information on all target students was collected. The targeted platforms were chosen depending on the data that is readily available for each student. For example, concerning the first student (see Table 2) we gathered 51 samples from Whatsapp, 41 samples from Instagram, 6 samples from Twitter and 5 samples from Google Chat. The quantity of data samples that were collected for each student in each social media platform is displayed in the Table 2.

Table 2

Sample of data for each student per each platform included in the study

Student	Instagram	Twitter	WhatsApp	Google Chat	total by student
Student 1	41	6	51	5	103
Student 2	52	22	77	11	162
Student 3	38	15	42	8	103
Student 4	62	3	33	3	101
Student 5	23	2	68	15	108
Student 6	30	23	41	9	103
Student 7	13	4	56	5	78
Student 8	18	29	39	12	98
Student 9	78	12	74	22	186
Student 10	33	21	52	13	119
Total	388	137	533	103	1161
Average	38,8	13,7	53,3	10,3	116,1

As shown in Table 2, the number of collected data sample is not equivalent comparing different platforms to each other's. This is due to the fact that some platforms are more often used by students than others (WhatsApp and Instagram for example). The total number of data gathered is 1161, with an average of 116,1 samples for each student. The data was stored in CSV files, with each line containing the student's text samples arranged by their originating environment. To make sure that the content of the students' discussions was obvious and comprehensible and that the process of organizing the data was completed without errors, all data was reviewed as well as we ensured the samples belonged to the right students.

Data selection/preprocessing

In this step, we have considered data of all students taking part in the study as an adequate number of samples was collected for each participant. In order to prevent utilizing unrepresentative samples, they had previously been chosen using a variety of estimated characteristics.

For data cleaning, we removed some insignificant data which made up less than 2.8% of the overall data. The data has been transformed before being processed due to Moroccan students' use of their native dialect ("darija") in their writings. The data transformation procedure was a very challenging step. The text could not be used in its original form since the NLP (Natural Language Processing) Model does not support the Moroccan dialect. To address this issue, we translated the content into English as it is a language that the model can comprehend. Therefore, we have carried out a full English translation of the content (see examples in Table 3). The next stage was to incorporate all the student samples, regardless of the environment type. Subsequently, all student data were collected in one entry of the Model because we needed to predict each student's personality separately.

Table 3
Text language standardization to English

Sentence in Moroccan dialect	English translation
9alkom lprof nhar l7ad a5ir ajal dyal les projets rendu	the teacher informs you that this Sunday is the deadline to send your projects
Lcours d8ada rah annulé	tomorrow's class is canceled
B8it les solutions dyal tepeyat	I need practical works solutions
Darni rassi 8adi nmxi	I have a headache; I'm going to leave
5oya 3afak xof m3aya l'erreur fin kayna	Bro, help me find the mistake
Walo dak logiciel mab8ach yt2anstala leya	unfortunately, the software is not installed
Nari majibtx mzyan 9afarta	alas, I didn't get a good mark

Symanto: personality prediction tool

To identify and analyze personalities from the text, there are several AI-based solutions. Many of them are a paid service for business use and can be accessed only for evaluation purpose. However, there are some other applications that are for free use in research domain. In this work we use the free evaluation API of Symanto which is a tool that provides companies with insights from customer data. Symanto is also provided as an API on RapidAPI (<https://rapidapi.com/>).

In this experiment, we used the RapidAPI web interface to send student text, which resulted in an HTTP POST request to the API. The web service response contains predictions made by the AI Model presented in the .json data format. This last contains the probability of each personality trait based on the Big Five, as well as multiple subclasses of each personality trait. Only predicted values of the first level of the Big Five personality traits were considered in this study.

Results

The preliminary results of this work are the outcomes originated from examining student writing using Symanto AI-based tool to detect personality. The results include the probability of each personality trait according to the Big Five Model which is displayed in Table 4.

Table 4
Students' personality detected by Symanto

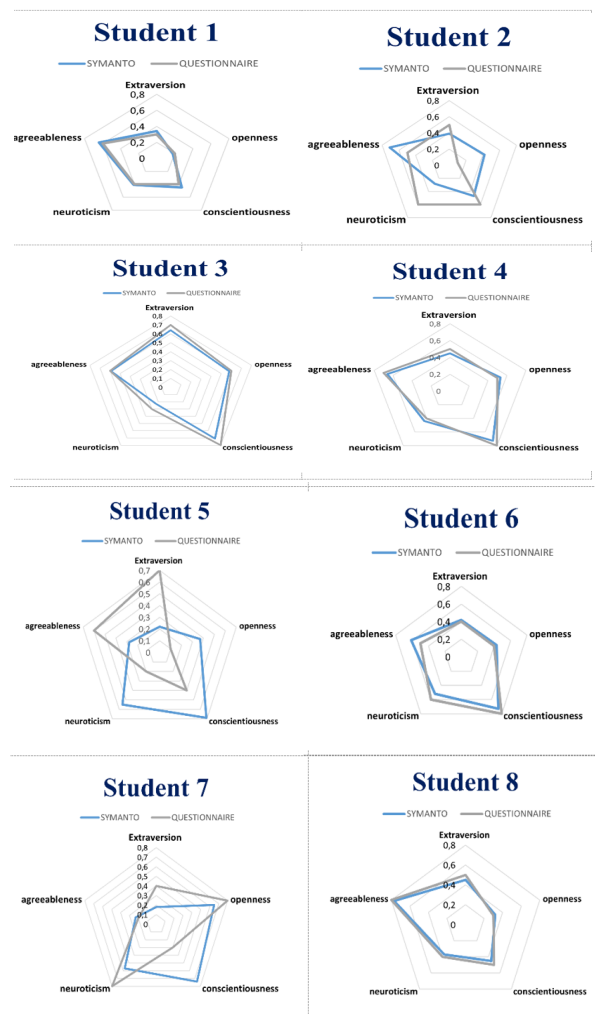
	Extraversion	Openness	Conscientiousness	Neuroticism	Agreeableness
Student 1	0,34	0,18	0,45	0,41	0,64
Student 2	0,39	0,42	0,47	0,28	0,71
Student 3	0,64	0,58	0,71	0,23	0,59
Student 4	0,45	0,53	0,73	0,44	0,66
Student 5	0,22	0,37	0,69	0,55	0,28
Student 6	0,42	0,43	0,73	0,52	0,61
Student 7	0,18	0,65	0,74	0,57	0,23
Student 8	0,45	0,32	0,45	0,37	0,77
Student 9	0,23	0,46	0,4	0,29	0,73
Student 10	0,54	0,66	0,73	0,36	0,71

The same students were also requested to complete a personality questionnaire based on the Big Five Model, as was previously mentioned. The results of this personality test are shown in Table 5.

Table 5
Students' personality detected based on the BF Questionnaire.

	Extraversion	Openness	Conscientiousness	Neuroticism	Agreeableness
Student 1	0,3	0,2	0,4	0,4	0,6
Student 2	0,5	0,1	0,6	0,6	0,5
Student 3	0,7	0,6	0,8	0,3	0,6
Student 4	0,5	0,5	0,8	0,4	0,7
Student 5	0,7	0,1	0,4	0,2	0,6
Student 6	0,4	0,4	0,8	0,6	0,5
Student 7	0,4	0,8	0,3	0,8	0,2
Student 8	0,5	0,3	0,5	0,4	0,8
Student 9	0,3	0,4	0,3	0,3	0,7
Student	0,6	0,7	0,6	0,3	0,8

Representing personality traits detected by the AI based tool and those assumed by the Questionnaire in a radar chart for each student can assist in comparing their results. The following figures illustrate the radar charts but based on the comparison between the results of the tables presented previously (see Figure 2).



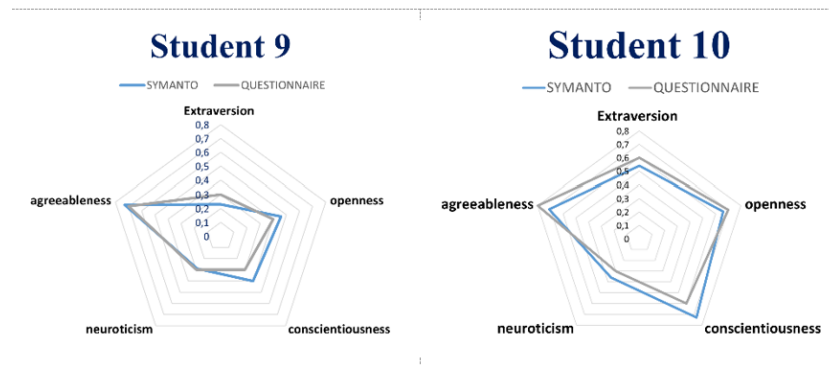


Figure 2. Radar charts of personality predictions for each student

The above-mentioned data demonstrates that diverse personalities predicted by the AI based tool do not always produce similar results based on the Questionnaire. Some students' predictions (students 1, 3, 4, and 8) are quite accurate. However, there is a huge difference between predicted personality of three students compared to their personality detected by the BF Questionnaire (Student 2, Student 5 and Student 7).

Therefore, in order to evaluate results more precisely, we have decided to use metrics that can give more insights from the results. Bellow we present all the metrics used:

- HPT (Highest personality traits): is when a student's highest personality trait value from the AI-based tool and his highest personality trait value from the Questionnaire are same, this setting is set to true. This metric will give us a rate that describes how much the model predicts the most dominant personality trait.
- LPT (Lowest personality traits): is when a student's lowest personality trait value from the AI-based tool and his lowest personality trait value from the Questionnaire are the same, this setting is set to true. This metric will give us a rate that describes how much the model predicts the weakest personality trait in a person.
- ME (Mean error): is a measurement of the average difference between the value of a personality characteristic predicted by an AI-based tool and the value supplied by a Questionnaire for that same personality trait. When this number is small, predictions are reliable and accurate.
- SD (Standard deviation): gauges the variation in the mean error of the difference between the expected and actual values. When this score is low, there are few discrepancies in predictions for the five personality characteristics.

In order to obtain a more accurate result, we have assumed that unifying the precision of the obtained results can provide a more exact metrics' values since the precision of the values produced by the AI-based tool (2 digits after the decimal point) and those returned by the Questionnaire (1 digit after the decimal point) are not the same. Hence, the mean error and standard deviation were recalculated after unifying the values.

Table 6 summarizes the results of different evaluation metrics that compares personality prediction of Symanto AI based solution to that of the Big Five Questionnaire.

Table 6
Evaluation metrics results of the comparison between Symanto APIs and the Big Five Questionnaire

	HPT	LPT	ME	SD	ME (unified precision)	SD (unified precision)
Student 1	True	True	0,0320	0,0164	0,0200	0,0447
Student 2	False	False	0,2180	0,1003	0,2000	0,1000
Student 3	True	True	0,0500	0,0339	0,0600	0,0548
Student 4	True	True	0,0460	0,0152	0,0200	0,0447
Student 5	False	False	0,3420	0,0829	0,3600	0,0894
Student 6	True	True	0,0620	0,0370	0,0600	0,0548
Student 7	False	False	0,2140	0,1494	0,1800	0,1483
Student 8	True	True	0,0360	0,0134	0,0000	0,0000
Student 9	True	True	0,0540	0,0351	0,0600	0,0548
Student 10	True	True	0,0760	0,0351	0,0800	0,0447
All students	7/10	7/10	0,1130	0,0519	0,1040	0,0636

The Symanto APIs predict correctly the highest and lowest personality traits in seven students from a total of ten students. For all students, the mean difference is roughly 0.11, while the standard deviation is about 0.05. The model predicts the Big Five personality characteristics well based on initial measurements or unified precision metrics.

The objective of evaluating the correlation between predictions from the Symanto AI-based model and those from the Questionnaire is to figure out the accuracy of Symanto APIs for each personality trait. Therefore, instead of comparing the values themselves, the purpose is to compare the variety of the predictions. This indicates that if the correlation is close to 1 the values predicted by the AI based tool and those of the BF Questionnaire are very dependent. In other words, the change in the personality from one student to another occurs in exactly the same way even if the predicted values are not completely identical to those identified by the BF Questionnaire. This also means that for each personality trait, there is a very strong correlation between the predictions made by the AI-based model and the outcomes of the Questionnaire. The correlation results are shown in Table 7. It was calculated both separately for each personality trait as well as overall for all personality traits.

Table 7
Correlation between Symanto predictions and Questionnaire results

	Correlation	Correlation (unified precision)
Extraversion	0,4731	0,4572
Openness	0,8402	0,8486
Conscientiousness	0,4691	0,4839
Neuroticism	0,3573	0,3164
Agreeableness	0,7114	0,7578
All traits	0,6143	0,6168

As shown in the previous table, the correlation values of personality traits in the Symanto model range between 0.31 and 0.84. The correlation coefficient for all traits is approximately 0.61 based on the initial values and on the unified precision values. In general, we cannot deny the relationship between the model's predictions and those calculated by the Questionnaire. However, the correlation is not very strong, balanced around 0.6.

Even the correlation between results indicated by the Questionnaire and those predicted by the AI-based model is significant, we cannot consider it substantial. Moreover, we had identified three students (students 2, 5 and 7) whose findings were considerably different when comparing Symanto AI-based model predictions to the Questionnaire ones (see radar charts in Fig. 3). Consequently, presuming that it

contains biased data (see discussion section), we decided to recalculate the correlation after eliminating these three students' data (see Table 8). The correlation increased substantially as a result of this modification approaching 0.9.

Table 8
Correlation between Symanto predictions and Questionnaire results after elimination of biased data

	Correlation	Correlation (unified precision)
Extraversion	0,9591	0,9341
Openness	0,9787	0,9758
Conscientiousness	0,9131	0,9303
Neuroticism	0,8702	0,7727
Agreeableness	0,8526	0,8775
All traits	0,9453	0,9309

We were able to assert the absurdity of the predictions made about the three students by reevaluating the data while taking into consideration the samples which show the personality of the student. Table 9 summarizes the results of this reevaluation while table 10 shows some examples of low and good quality samples.

Table 9
Number of samples classified by quality after the revision.

Student	Number of good-quality Samples	Number of low-quality Samples	Total by student
Student 1	92	11	103
Student 2	33	129	162
Student 3	87	16	103
Student 4	89	12	101
Student 5	21	87	108
Student 6	93	10	103
Student 7	21	57	78
Student 8	90	8	98
Student 9	163	23	186
Student 10	104	15	119
Total	793	368	1161

Table 10
Examples of good and low quality samples after the revision.

Examples of good-quality Samples	Examples of low-quality Samples
I suggest collaborating with you on this project.	The teacher has already shared the correction.
I must complete the exercise on time.	The update is complete.
I think I can help you debug this code.	I will not be present in the next session.

Discussion

Concerning the evaluation of Symanto as a personality prediction tool, we can confirm from the beginning based on the radar charts representation that Symanto AI-based solution was successful in predicting personality traits based on the Big Five Model. Predicted personality of seven students from ten was very close to this obtained by the questionnaire.

The evaluation metrics defined in this research revealed more important information about the evaluated model. For the same seven students, the most and least dominant personality traits were correctly predicted. The low values of the Mean Error and the Standard deviation confirm that the predictions are very accurate.

Unifying the precision of the values obtained was supposed to give a more precise comparison between the results. However, the findings of the metrics recalculation following the adjustment of values with a uniform precision did not significantly affect the initial results. But in any case, we can consider that this consequently confirms the initial results.

The correlation between the two compared results is a very significant metric that calculates numerically if the values of personality vary in the same manner in both compared approaches. A more important insight that correlation gave us is the variation concerning each personality trait separately. The calculation of this metric showed significant correlation based on the results of all students for all personality traits. Regarding the assessment of the correlation of each personality trait apart, the correlation was raised for the Openness and the Agreeableness, on the other hand it was not very significant for the rest of the traits.

Although overall the results obtained clearly show that the AI-based tool can detect the personality of students. We looked for an explanation for the failure of this process for the three students concerned. The most likely hypothesis based on how the AI model worked was the quality of the text gathered for these students. The text written by a person may in certain cases not reflect its true personality. This means that the personality detection algorithm may not be responsible for the error of a detected personality from a text that does not contain expressions that really show this personality. The Symanto team also recommends using text that specially expresses the person point of view because through this the person shows its personality. According to this, we decided to revise (human revision) all the data to check the quality of the text taking in order to classify it according to this criterion. Table 9 represents the number of good-quality and low-quality samples for each student after the revision of the data. Examples of good and low quality text are shown in table 10. Therefore, based on the results of the new revision we decided to exclude the data of the three students and recalculate the same evaluation metrics based on the remaining seven students. We considered that the new results are fairer to evaluate the model because they are not affected by low quality text.

The re-evaluation of the model based on the seven students who wrote texts that show more of their personality yielded more accurate results. The correlation is really very strong between the personality detected by the model and that of the questionnaire. The latter has spread 0.9 for all personality traits while it was between 0.85 and 0.97 for separated personality traits. This confirms substantially the accuracy of the predictions of the personality.

Generally, we can assume that this AI-based tool that employs related techniques from NLP to identify personality from language expression predicts personality traits exceptionally well. This fact is strongly conditioned by having sufficient data collected about each evaluated student, as well as by the quality of the data that should include expressions that reveal the student's thoughts and behavior.

Ultimately, as understanding students' personalities is very important in the teaching and learning context. The present study shows that using automated text-based personality assessment combined with good quality text of student collected from social media-based environment can help in detecting student personality instead of using the traditional Big Five Questionnaire. and therefore, thanks to these tools we can develop specific solutions that automatically detect the learner's personality in social media-based learning environments.

Conclusions

In the current study, we collected data from social networks where students discuss about learning. We specifically gathered data from Twitter, Google Chat, Instagram, and WhatsApp groups. Data was processed using the AI-based Symanto APIs in order to identify students' personalities. The outcomes of this latter were analyzed to determine if the combination of student data from learning environments and this tool can help in identifying student's personality. The results from the Symanto AI-based tool were compared to results from the personality test Questionnaire.

According to the study's outcomes, Symanto APIs can be really helpful in determining a student's personality based on their interactions in learning environments. Most of the student's personality traits were implied in their written text. Moreover, text quality is the most important factor in determining a student's personality.

The evaluated tool accurately identified the dominant personality characteristic and almost every other trait with a very low error value. Furthermore, the correlation between projected personality from text and those found by the Questionnaire was very significant (around 0,9) after omitting three students who were considered as outlier samples. Either for personality in general or for individual personality traits, the correlation value is considerably strong.

This experiment also shows that not all of the students' personalities can accurately be identified from text. Out of 10 students, three had personalities that the AI-based tool was unable to identify. This failure can be explained by the text itself more often than by the tool's inefficiency. The data of these students does not contain sufficient expressions that show explicitly students' personality traits. Therefore, it is strongly recommended to check the quality of the text before processing it by the model to detect the personality.

In future work, we will evaluate other AI personality detection tools and test them on other data to be collected. We will also study other personality detection techniques that are not text-based to ultimately build a multimodal solution that detects personality from several data sources.

Acknowledgements

This work was supported by "Centre National pour la Recherche Scientifique et Technique" (Grant agreement number SHSE-2021/49).

Conflict of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Contributions

Data Curation, N. El Bahri, A. Abtoy; methodology, Z. Itahriouan, N. El Bahri; Formal analysis, N. El Bahri, Z. Itahriouan, S. Brahim Belhaouari; Validation, Z. Itahriouan, A. Abtoy; writing—original draft preparation, N. El Bahri; writing—review and editing, Z. Itahriouan, A. Abtoy, S. Brahim Belhaouari; Funding acquisition, Z. Itahriouan. All authors have read and agreed to the published version of the manuscript.

References

- Barrot, J. S. (2022). Social media as a language learning environment: a systematic review of the literature (2008-2019). *Computer assisted language learning*, 35(9), 2534-2562. <https://doi.org/10.1080/09588221.2021.188673>
- Bronfenbrenner, U., (1951). Toward an integrated theory of personality, in: *Perception: An Approach to Personality*. Ronald Press Company, New York, NY, US, pp. 206–257. <https://doi.org/10.1037/11505-008>
- Caprara, G. V., Barbaranelli, C., Borgogni, L., & Perugini, M. (1993). The "Big Five Questionnaire": A new questionnaire to assess the five factor model. *Personality and Individual Differences*, 15(3), 281-288. [https://doi.org/10.1016/0191-8869\(93\)90218-R](https://doi.org/10.1016/0191-8869(93)90218-R)
- Christian, H., Suhartono, D., Chowanda, A., Zamli, K.Z. (2021). Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging. *Journal of Big Data* 8, 68. <https://doi.org/10.1186/s40537-021-00459-1>
- Christine Greenhow, Sarah M. Galvin, K. Bret Staudt Willet, (2019). What Should Be the Role of Social Media in Education? [WWW Document]. Retrieved from <https://journals.sagepub.com/doi/abs/10.1177/2372732219865290> (accessed 9.5.2023).
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual review of psychology*, 41(1), 417-440. <https://doi.org/10.1146/annurev.ps.41.020190.002221>
- Hans, C., Suhartono, D., Andry, C., & Zamli, K. Z. (2021). Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging. *Journal of Big Data*, 8(68). <https://doi.org/10.1186/s40537-021-00459-1>
- Eysenck, H. J. (1981). General features of the model. In *A model for personality* (pp. 1-37). Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-67783-0>
- Eysenck, H. J. (1991). Dimensions of personality: The biosocial approach to personality. In *Explorations in temperament: International perspectives on theory and measurement* (pp. 87-103). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4899-0643-4_7
- Eysenck, H. J. (1994). *The "Big Five" or "Giant 3"? Criteria for a paradigm. The developing structure of temperament and personality from infancy to adulthood*. Hillsdale, NJ: Erlbaum.
- Gjurković, M., Vukojević, I., & Šnajder, J. (2022). SIMPA: statement-to-item matching personality assessment from text. *Future Generation Computer Systems*, 130, 114-127. <https://doi.org/10.1016/j.future.2021.12.014>

- Greenhow, C., Galvin, S. M., & Staudt Willet, K. B. (2019). What should be the role of social media in education?. *Policy Insights from the Behavioral and Brain Sciences*, 6(2), 178-185. <https://doi.org/10.1177/2372732219865290>
- Josué, A., Bedoya-Flores, M. C., Mosquera-Quiñonez, E. F., Mesías-Simisterra, A. E., & Bautista-Sánchez, J. V. (2023). Educational Platforms: Digital Tools for the teaching-learning process in Education. *Ibero-American Journal of Education & Society Research*, 3(1), 259-263. <https://doi.org/10.56183/iberoeds.v3i1.626>
- Namaziandost, E., & Nasri, M. (2019). The impact of social media on EFL learners' speaking skill: a survey study involving EFL teachers and students. *Journal of Applied Linguistics and Language Research*, 6(3), 199-215. <https://www.jallr.com/index.php/JALLR/article/view/1031>
- Pittenger, D. J. (1993). Measuring the MBTI... and coming up short. *Journal of Career Planning and Employment*, 54(1), 48-52. <https://img3.reoveme.com/m/614576efb2b91676.pdf>
- Rahman, M. A., Al Faisal, A., Khanam, T., Amjad, M., & Siddik, M. S. (2019, May). Personality detection from text using convolutional neural network. In *2019 1st international conference on advances in science, engineering and robotics technology (ICASERT)* (pp. 1-6). IEEE. <https://doi.org/10.1109/ICASERT.2019.8934548>
- See Yin Lim, J., Agostinho, S., Harper, B., & Chicharo, J. (2014). The engagement of social media technologies by undergraduate informatics students for academic purpose in Malaysia. *Journal of Information, Communication and Ethics in Society*, 12(3), 177-194. <https://doi.org/10.1108/JICES-03-2014-0016>
- Sugerman, J. (2009). Using the DiSC® model to improve communication effectiveness. *Industrial and Commercial Training*, 41(3), 151-154. <https://doi.org/10.1108/00197850910950952>
- Swaminathan, T. N., Harish, A., & Cherian, B. (2013). Effect of social media outreach engagement in institutions of higher learning in India. *Asia-Pacific Journal of Management Research and Innovation*, 9(4), 349-357. <https://doi.org/10.1177/2319510X14523101>
- Tlili, A., Essalmi, F., Jemni, M., & Chen, N. S. (2016). Role of personality in computer based learning. *Computers in Human Behavior*, 64, 805-813. <https://doi.org/10.1016/j.chb.2016.07.043>
- Utami, E., Hartanto, A. D., Adi, S., Oyong, I., & Raharjo, S. (2022). Profiling analysis of DISC personality traits based on Twitter posts in Bahasa Indonesia. *Journal of King Saud University-Computer and Information Sciences*, 34(2), 264-269. <https://doi.org/10.1016/j.jksuci.2019.10.008>
- Utami, N. A., Maharani, W., & Atastina, I. (2021). Personality classification of facebook users according to big five personality using SVM (support vector machine) method. *Procedia Computer Science*, 179, 177-184. <https://doi.org/10.1016/j.procs.2020.12.023>
- Wickramanayake, L., & Muhammad Jika, S. (2018). Social media use by undergraduate students of education in Nigeria: A survey. *The Electronic Library*, 36(1), 21-37. <https://doi.org/10.1108/EL-01-2017-0023>
- Zhang, S. (2020, April). Psychoanalysis: The influence of Freud's theory in personality psychology. In *International Conference on Mental Health and Humanities Education (ICMHHE 2020)* (pp. 229-232). Atlantis Press. <https://doi.org/10.2991/assehr.k.200425.051>