Journal of Computer Science

Original Research Paper

Predicting Smartphone Addiction in Teenagers: An Integrative Model Incorporating Machine Learning and Big Five Personality Traits

Jacobo Osorio, Marko Figueroa and Lenis Wong

Department of Information Systems Engineering Program, Peruvian University of Applied Sciences, Peru

Article history
Received: 14-07-2023
Revised: 20-09-2023
Accepted: 18-10-2023

Corresponding Author:
Lenis Wong
Department of Information Systems Engineering Program, Peruvian University of Applied Sciences, Peru
Email: lwongpuni@gmail.com

Abstract: Smartphone addiction has emerged as a growing concern in society, particularly among teenagers, due to its potential negative impact on physical, emotional social well-being. The excessive use of smartphones has consistently shown associations with negative outcomes, highlighting a strong dependence on these devices, which often leads to detrimental effects on mental health, including heightened levels of anxiety, distress, stress depression. This psychological burden can further result in the neglect of daily activities as individuals become increasingly engrossed in seeking pleasure through their smartphones. The aim of this study is to develop a predictive model utilizing machine learning techniques to identify smartphone addiction based on the "Big Five Personality Traits (BFPT)". The model was developed by following five out of the six phases of the "Cross Industry Standard Process for Data Mining (CRISP-DM)" methodology, namely "business understanding," "data understanding," "data preparation," "modeling," and "evaluation." To construct the database, data was collected from a school using the Big Five Inventory (BFI) and the Smartphone Addiction Scale (SAS) questionnaires. Subsequently, four algorithms (DT, RF, XGB LG) were employed the correlation between the personality traits and addiction was examined. The analysis revealed a relationship between the traits of neuroticism and conscientiousness with smartphone addiction. The results demonstrated that the RF algorithm achieved an accuracy of 89.7%, a precision of 87.3% the highest AUC value on the ROC curve. These findings highlight the effectiveness of the proposed model in accurately predicting smartphone addiction among adolescents.

Keywords: Smartphone Addiction, Machine Learning, Predictive Model, Big Five Personality Traits, Random Forest

Introduction

In contemporary times, it is widely acknowledged that people have become increasingly reliant on smartphones. Despite the availability of studies highlighting this issue, no significant preventive measures or actions have been taken to address this growing problem. It is important to emphasize that smartphone addiction has been linked to physical (Li *et al*., 2021) and emotional effects (Chen *et al*., 2021; Cheng and Meng, 2021; Lei *et al*., 2020; Wickord and Quaiser-Pohl, 2022), contributing to psychological challenges in individuals' lives (Lei *et al*., 2020; Rho *et al*., 2019).

"Addictions are physical and psycho-emotional diseases that create a dependency on or need for a substance, activity or relationship; they are a set of signs and symptoms that are influenced by biological, genetic, psychological and social factors" (Castillo-Viera *et al*., 2022). Considering this definition, smartphone addiction can be regarded as a psychological addiction due to its repetitive involvement of pleasurable behaviors, leading to a loss of control that hinders individuals in their daily activities (Minsa, 2021). It has been observed that 83% of Peru's urban population uses smartphones, with over 50% using them for entertainment purposes (Ipsos, 2021). Although smartphones were initially designed for the primary purpose of communication, they have become widely used for entertainment purposes. This includes activities such as using social networks, streaming series or movies even online shopping, which reinforce factors that contribute to addiction (Nida, 2022).

Besides multiple factors leading to addictive behaviors, it is important to underscore the association between addiction and the neuroticism trait (Lei *et al*., 2020; Müller *et al*., 2021). Individuals with a high score in this personality trait are more susceptible to distractibility and engaging in obsessive thoughts associated with addiction, anxiety, or stress. This predisposition can lead to the development of addictive behaviors and a strong dependence on their devices (Li *et al*., 2022), causing them to neglect their daily activities in pursuit of solitude (Abu-Taieh *et al*., 2022; Chen *et al*., 2021), having side effects such as anxiety (Li *et al*., 2022; Müller *et al*., 2021), distress (Chen *et al*., 2021; Lei *et al*., 2020) and stress (Müller *et al*., 2021). The effects extend to the individual's mental, physical emotional well-being. It is essential to note that these psychological issues are also recognized as contributing factors to addiction (Cheng and Meng, 2021; Lei *et al*., 2020).

To address this issue, various studies have emerged emphasizing the necessity of implementing strategies to mitigate smartphone addiction. It is highly recommended to implement programs aimed at raising awareness and promoting responsible mobile device usage, both within educational settings and within families. However, despite the current research on the correlation between smartphone addiction and personality traits, the utilization of machine learning algorithms as a solution remains largely unexplored. Thus, there is a pressing need to develop a novel approach that leverages new technologies, enabling more accurate prediction of smartphone addiction among adolescents.

Therefore, this article presents a smartphone addiction prediction model that combines machine learning techniques and the big five personality traits. The random forest algorithm is employed to analyze the relationship between the five personality traits and smartphone addiction, utilizing the "Big Five Inventory (BFI)" and "Smartphone Addiction Scale (SAS)" questionnaires.

Age Groups

It is important to consider age as it plays a significant factor in the relationship between individuals and their smartphones (Eichenberg *et al*., 2021).

On the one hand, considering age as a relevant factor, teenagers (Abu-Taieh *et al*., 2022; Duan *et al*., 2021) are more susceptible to present smartphone addiction due to increased exposure to technology, which begins at an early age when there is a lack of self-regulation. As a result, teenagers may have a higher likelihood of being exposed to smartphones and being drawn towards engaging with satisfying stimuli, thereby increasing their vulnerability to smartphone addiction.

On the other hand, adults are prone to experience emotional dependence on the device, where they feel the need to be constantly connected due to work and personal responsibilities. Because of this, adults face difficulties in balancing the time they dedicate to it despite what was previously described, adults have greater self-control when using a smartphone, which shows an inversely proportional relationship between age and smartphone addiction (Marengo *et al*., 2020).

Personality Traits

The relationship between the Big Five Personality Traits (BFPT) and smartphone addiction was analyzed to identify the primary characteristics associated with this addiction based on individuals' personality traits. Subsequently, the relation between each trait and addiction was examined, encompassing the following five traits from the BFPT: Neuroticism, conscientiousness, extraversion, agreeableness openness.

Regarding the influence of personality traits, it was found that extraversion is related to smartphone addiction (Toyama and Hayashi, 2022). It has been identified that people with high levels have the need to constantly interact and seek out satisfactory social experiences, leading to the device serving to fulfill their socialization needs (Eichenberg *et al*., 2021; Peltonen *et al*., 2020).

Moreover, neuroticism is considered the most influential trait (Müller *et al*., 2021). High levels are associated with a tendency to experience negative emotions (Müller *et al*., 2021) and difficulty in managing stress or anxiety, making individuals more susceptible to developing addiction. For this reason, using the device provides a sense of security and control, which leads to an increase in its constant use (Lei *et al*., 2020; Müller *et al*., 2021; Zeighami *et al*., 2021).

Conscientiousness has an inverse relationship with addiction, whereby individuals with lower levels tend to be less aware of the amount of time they spend using their devices. They may exhibit disorganized tendencies and often prioritize less important activities over more significant ones (Marengo *et al*., 2020; Müller *et al*., 2021).

On the other hand, the traits of openness and agreeableness have not shown a direct relationship with smartphone addiction (Eichenberg *et al*., 2021; Erdem and Uzun, 2022; Müller *et al*., 2021; Peltonen *et al*., 2020). Although these traits influence other aspects of human behavior, they are not directly linked to smartphone addiction.

In conclusion, there is a direct relationship between the trait of neuroticism and an inverse relationship with the trait of conscientiousness (Peterka-Bonetta *et al*., 2019; Toyama and Hayashi, 2022) and, despite the existence of evidence based on the analysis carried out, it is important to highlight that these statements are only hypotheses since addiction is a complex phenomenon that is influenced by multiple factors such as the social environment, behavior patterns other psychological aspects.

Machine Learning Algorithms

Machine Learning algorithms are employed to analyze the collected data and establish a correlation between personality traits and smartphone usage patterns, as depicted in Table 1. The aim of this comparison is to identify the algorithm that best suits the research objectives and facilitates the development of an accurate detection model to enhance our understanding of smartphone addiction. Among the algorithms utilized for predicting behaviors and psychological disorders, we consider Decision Tree (DT), Random Forest (RF), XG Boost (XBG) Logistic Regression (LR).

The DT algorithm is known for its interpretability (Lee and Kim, 2021), enabling us to identify direct relationships between personality traits and addiction. It delivers outstanding results in terms of precision and specificity when predicting behaviors (Makino *et al*., 2021). However, while the DT algorithm is effective for behavior prediction, the RF algorithm outperforms it in terms of performance (Razavi *et al*., 2020).

By leveraging an ensemble of multiple decision trees, the Random Forest algorithm allows for more comprehensive study and analysis (Abu-Taieh *et al*., 2022).

The RF algorithm demonstrated superior performance in detecting mental health crises (Garriga *et al*., 2022; Xia *et al*., 2022) and has demonstrated superior performance in most cases (Lee and Kim, 2021) due to its ability to identify complex patterns in the data.

Similarly, the LR algorithm is employed in the prediction of behaviors. Unlike the others, it is a binary classification algorithm that estimates the probabilities of belonging to a specific class, which makes it particularly useful when attempting to understand the influence of each trait or variable in predicting behavior.

Materials and Methods

Cross Industry Standard Process for Data Mining (CRISP-DM) is an approach that is structured for the development of data mining projects and the creation of prediction models (Blasi and Alsuwaiket, 2020). It consists of six main phases that can be seen in Fig. 1. The phases are (1) Business understanding, (2) Data understanding, (3) Data preparation, (4) Modeling, (5) Evaluation (6) Deployment. For the development of the research, the first five phases were carried out.

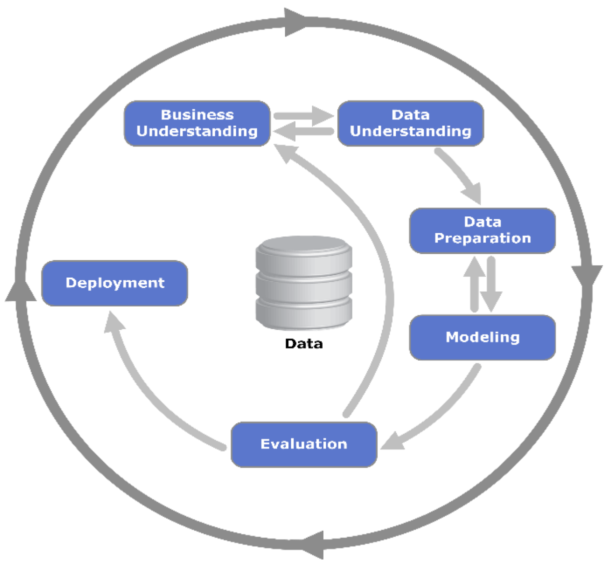

Fig. 1: CRIPS-DM methodology (Peralta, 2014)

Table 1: Comparison of algorithms for behavior prediction

Reference	Algorithm	Purpose	Complexity	Robustness
Chen *et al*. (2022); Duan *et al*. (2021); Garriga *et al*. (2022); Haque *et al*. (2021); Kim *et al*. (2021); Lee and Kim (2021); Makino *et al*. (2021); Xia *et al*. (2022)	Decision Tree (DT)	Perform classifications and predictions	Low complexity, with an easy-to-understand structure	Moderate robustness against noise and outlier data, as it can be influenced by the presence of extreme values
Abu-Taieh *et al*. (2022); Chen *et al*. (2022); Haque *et al*. (2021); Kim *et al*. (2021); Lee and Kim (2021); Peltonen *et al*. (2020); Razavi *et al*. (2020); Xia *et al*. (2022)	Random Forest (RF)	Perform classifications and predictions	Medium complexity, with manageable interpretability	Robustness against noisy and outlier data due to its combination of multiple decision trees
Chen *et al*. (2022); Garriga *et al*. (2022); Haque *et al*. (2021); Lee and Kim (2021); Xia *et al*. (2022)	XG Boost (XGB)	Enhance the performance and accuracy of machine learning models	Highly complex, requires more advanced knowledge	Robustness against noise and outlier data thanks to its ability to handle errors and learn from them
Chen *et al*. (2022); Garriga *et al*. (2022); Razavi *et al*. (2020); Xia *et al*. (2022)	Logistic Regression (LR)	Classification involves estimating the probabilities of belonging to a class or not	Low complexity, easy to understand and implement	Moderate robustness against noisy and outlier data, thereby influenced by the presence of extreme values

Business Understanding

With the objective of identifying the personality traits that are related to smartphone addiction in adolescents between 12 and 17, a review of the literature and the social context has been carried out to analyze the personality traits related to smartphone addiction.

Data Understanding

The variables used for the prediction model are based on studies that analyze the big five personality traits (Eichenberg *et al*., 2021; Erdem and Uzun, 2022; Peltonen *et al*., 2020; Wickord and Quaiser-Pohl, 2022) are shown as a direct relationship with smartphone addiction this is observed in Table 2. After establishing the variables and understanding their relationship with the addiction, a connection with avoidance is identified. On the one hand, people who present high levels of neuroticism tend to use their smartphones as a way of distraction from their worries because of social anxiety; this could be a reason for their preference to keep in touch through social networks, which leads to an increase in smartphone use. On the other hand, the trait that is inversely related to addiction is conscientiousness, since people with lower levels are less prone to set limits regarding the use of the smartphone, this results in non-existent self-regulation leading to addiction behaviors.

Regarding the other traits, despite not having a direct relationship with addiction, they do influence people's behavior. For example, a high degree of extraversion implies the need to be in contact with others. Furthermore, this trait along with a low degree of agreeableness is related to video games (Peltonen *et al*., 2020). Finally, the openness trait is not related to smartphone addiction (Wickord and Quaiser-Pohl, 2022).

Data Preparation

This phase consists of data selection, data cleaning, data construction, data integration data formatting. To collect essential data, the Big Five Inventory (BFI) survey was used, which is designed to evaluate the personality traits of the participants (Cha and Seo, 2018). In addition to assessing personality traits, our objective was to determine the level of smartphone addiction among each teenager. Hence, following the BFI questions, we included the Smartphone Addiction Scale-Short Version (SAS-SV) survey.

A study was conducted at a private school in San Miguel, Lima, Peru, where 118 anonymous surveys were collected from students aged 12-16 years. Each survey consisted of 54 closed-ended questions regarding personality traits and smartphone usage. The responses were rated on a Likert scale ranging from 1-5 for the Big Five Inventory (BFI) which has 5 dimensions and is characterized by its acceptable internal consistency (Cronbach's α = 0.75 for neuroticism, 0.65 for agreeableness, 0.71 for conscientiousness, 0.86 for extraversion 0.69 for openness). Additionally, the responses were rated from 1-6 for the Smartphone Addiction Scale-Short Version (SAS-SV) where each scale shows a high level of internal consistency (Cronbach's α = 0.89). The data collection process lasted for two weeks, aiming to obtain a representative sample from the student population of the school.

Table 2: Description of each personality trait

Reference	Personality trait	Description	Low trait	High trait
Eichenberg *et al*. (2021); Lei *et al*. (2020); Müller *et al*. (2021); Peltonen *et al*. (2020); Peterka-Bonetta *et al*. (2019); Toyama and Hayashi (2022); Wickord and Quaiser-Pohl (2022)	Neuroticism	Index of emotional stability and impulse control	Security and trust	Sensitivity and nervousness
Eichenberg *et al*. (2021); Erdem and Uzun (2022); Müller *et al*. (2021); Peltonen *et al*. (2020); Peterka-Bonetta *et al*. (2019); Toyama and Hayashi (2022); Wickord and Quaiser-Pohl (2022)	Conscientiousness	It involves self-discipline Organization and personal responsibility	Carelessness lack of planning	Efficiency and organization
Eichenberg *et al*. (2021); Erdem and Uzun (2022); Peltonen *et al*. (2020); Toyama and Hayashi (2022); Wickord and Quaiser-Pohl (2022)	Extraversion	Characterized by energy and taste for social interaction	Reserve and preference for solitude	Expressiveness and search for social interactions courtesy and cooperation
Eichenberg *et al*. (2021); Erdem and Uzun (2022); Peltonen *et al*. (2020); Peterka-Bonetta *et al*. (2019); Wickord and Quaiser-Pohl (2022)	Agreeableness	It reflects a willingness to be kind, compassionate cooperative	Defiance and estrangement	
Eichenberg *et al*. (2021); Erdem and Uzun (2022); Peltonen *et al*. (2020); Wickord and Quaiser-Pohl (2022)	Openness	Intellectual curiosity, creativity is a preference for novelty	Caution and conventional thinking	Imagination and curiosity

During the data cleaning process, each survey was carefully reviewed to ensure that there were no blank questions. Additionally, reverse questions were included to ensure response validity. If a respondent answered affirmatively to a question and then answered affirmatively to its reverse question, it indicated inconsistent responses those surveys were discarded. Within this stage, each survey was thoroughly checked any surveys with outlier values were removed. Finally, after the data cleaning process, a clean sample of 96 questionnaires was obtained from the initial total of 118 surveys, for further analysis.

To select the relevant characteristics, statistical methods were employed with the main objective of establishing correlations between the variables and understanding their relationships. A total of 7 variables were considered, as outlined in Table 3, to comprehend the sample and examine the association between each personality trait and the smartphone addiction intensity.

Utilizing the collected data and established variables, the Pearson correlation coefficient was applied to analyze the influence of these variables on smartphone addiction. This analysis was conducted using the IBM SPSS Statistics software, which generated the correlations depicted in (Fig. 2) and the descriptive statistics presented in Table 4. Considering the level of smartphone addiction or "F06" as the primary variable, a strong relationship can be observed with the neuroticism variable or "F05" with a coefficient of 0.342. This indicates a positive association between the two variables, suggesting that as neuroticism increases, the likelihood of having a higher level of smartphone addiction also increases.

On the other hand, the conscientiousness variable, or "F02" shows an inverse relationship with smartphone addiction, with a coefficient of -0.415. This implies a negative association between the two variables, meaning that as conscientiousness increases, the probability of having a smartphone addiction decreases. Similar findings have been reported in previous research (Müller *et al*., 2021), where neuroticism and conscientiousness traits have influence coefficients of 0.379 and -0.404, respectively.

However, variables such as openness (F01), extraversion (F03) agreeableness (F04) are not closely related to smartphone addiction; they present coefficients of -0.082, and -0.079 0.039 respectively. These values indicate a weak or insignificant influence on smartphone addiction based on the sample. As in (Müller *et al*., 2021), these traits influence people's behavior, but they are not relevant to smartphone addiction.

An analysis of the data was made and the following statements, which are supported by the results in Table 4, were drawn:

- The sample of students has a greater than average degree of openness to experience and agreeableness with an average of 34.99 and 32.88 respectively a standard deviation of 6.204 and 5.155 respectively. As a result, students will have a favorable attitude towards trying new experiences and will show attitudes of compassion and cooperation
- The sample of students has a slightly higher degree of conscientiousness than the average, having a mean value of 29.31 and a standard deviation of 5.935. So, students will tend to be more organized and responsible in their behaviors as well as their decisions
- The sample of students has an average degree of neuroticism and extraversion, having a mean of 23.68 and a standard deviation of 6.036 for neuroticism a mean of 26.20 with a standard deviation of 7.106 for extraversion, which means that the students do not stand out in any of those traits
- The sample has a level of addiction below the average. Of the total, only 25% present smartphone addiction. The mean value was calculated and 24.82 was obtained with a standard deviation of 10.154, which means that there is significant variability in the results, with some students presenting values considerably higher or lower than the average, showing different levels of addiction. Thus, it can be concluded that the level of addiction of the students is low

Table 3: Variables used for the predictive model

ID	Variable	Description
F01	Openness	Indicates the respondent's level of openness to the experience
F02	Conscientiousness	Indicates the level of awareness of the respondent
F03	Extraversion	Indicates the level of extraversion of the respondent
F04	Agreeableness	Indicates the level of friendliness of the respondent
F05	Neuroticism	Indicates the level of neuroticism or emotional instability of the respondent
F06	Smartphone Addiction Scale (SAS)	Indicates the respondent's level of openness to the experience
F07	Age	Indicates the respondent's age in years

Table 4: Statistical data of the variables, obtained from IBM SPSS Statistics

ID	Median	Description	No.
F01	34.99	6,204	96
F02	29.31	5,935	96
F03	26.20	7,106	96
F04	32.88	5,155	96
F05	23.68	6,036	96
F06	24.82	10,154	96
F07	14.01	1,244	96

Correlaciones

		F01	F02	F03	F04	F05	F06	F07
F01	Correlación de Pearson	1	.351**	.275**	.279**	.002	-.082	.070
	Sig. (bilateral)		<.001	.007	.006	.981	.429	.501
	N	96	96	96	96	96	96	96
F02	Correlación de Pearson	.351**	1	.214*	.234*	-.245*	-.415**	.002
	Sig. (bilateral)	<.001		.036	.022	.016	<.001	.981
	N	96	96	96	96	96	96	96
F03	Correlación de Pearson	.275**	.214*	1	.222*	-.511**	-.079	.069
	Sig. (bilateral)	.007	.036		.030	<.001	.445	.505
	N	96	96	96	96	96	96	96
F04	Correlación de Pearson	.279**	.234*	.222*	1	-.214*	.039	.113
	Sig. (bilateral)	.006	.022	.030		.037	.705	.271
	N	96	96	96	96	96	96	96
F05	Correlación de Pearson	.002	-.245*	-.511**	-.214*	1	.342**	-.050
	Sig. (bilateral)	.981	.016	<.001	.037		<.001	.628
	N	96	96	96	96	96	96	96
F06	Correlación de Pearson	-.082	-.415**	-.079	.039	.342**	1	-.242*
	Sig. (bilateral)	.429	<.001	.445	.705	<.001		.018
	N	96	96	96	96	96	96	96
F07	Correlación de Pearson	.070	.002	.069	.113	-.050	-.242*	1
	Sig. (bilateral)	.501	.981	.505	.271	.628	.018	
	N	96	96	96	96	96	96	96

Fig. 2: Correlation between the variables and the SAS-SV

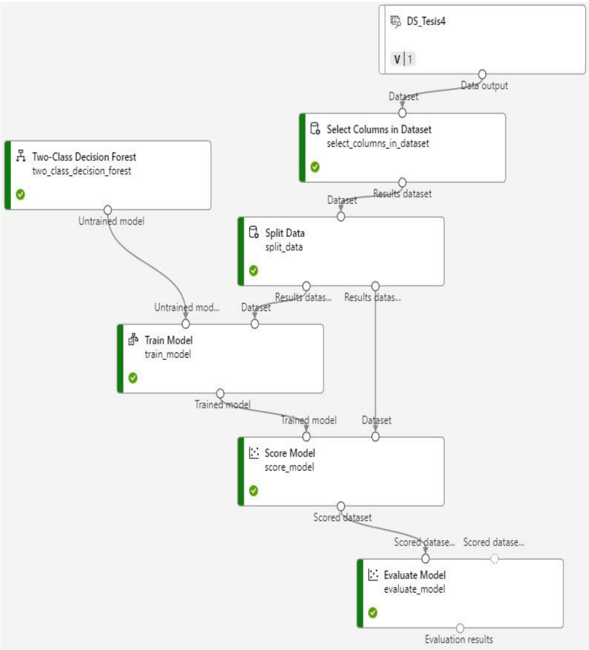

Fig. 3: Components used for random forest training

Modeling

This phase describes the classification techniques that will be used in this study. The modeling development was conducted using the Azure Machine Learning Studio platform. Throughout the process, a compute instance was employed to execute the Python code effectively. Furthermore, a cluster specifically designed for machine learning tasks was created as an add-on. These instances had the objective of training the algorithms using nodes from the platform as well as performing automated training for machine learning, which facilitated the training process and optimized the performance of the models.

The collected data was uploaded to the platform and the "select columns in dataset" component was used to define the relevant columns for the algorithm training, where age and the results of each personality trait in the SAS questionnaire were considered. The individual answers to each question were not considered to simplify algorithm training and to reduce the number of required parameters, which gave us greater interpretability.

Subsequently, a "split data" component was used to establish that 70% of the data was used for training and the remaining 30% for testing, in other words, 67 surveys were used for training and 29 surveys for testing. These samples were used to discover if it was possible to determine the respondents' addiction based on their results from the questionnaires.

Figure 3 illustrates the components utilized in the model. The "train model" component was employed for training purposes, while the "score model" component was used for conducting tests and making predictions.

Evaluation

Finally, the "evaluate model" component was utilized to analyze the algorithm's results. The same configuration was applied to both the DT and LR algorithms. In contrast to the three previous algorithms, in the case of the XGB algorithm, it was decided to use the "automated ML" procedure because the platform did not have the corresponding nodes for training.

The performance of the algorithms was evaluated using specific metrics (Jiménez *et al*., 2023), each with its own corresponding (Eqs. 1-5):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (3)$$

$$Specificity = \frac{TN}{TN + FP} \quad (4)$$

$$Fall\ Out = \frac{FP}{FP + TN} \quad (5)$$

Results and Discussion

After the training, the Confusion Matrix that evaluated the performance of each one of the algorithms was obtained. This can be seen in Fig. 4, generally giving us favorable and similar results for all the algorithms, highlighting the RF algorithm (Fig. 4a), which obtained more correct predictions than incorrect ones. The interpretation of the variables used for the metrics can be seen in Table 5.

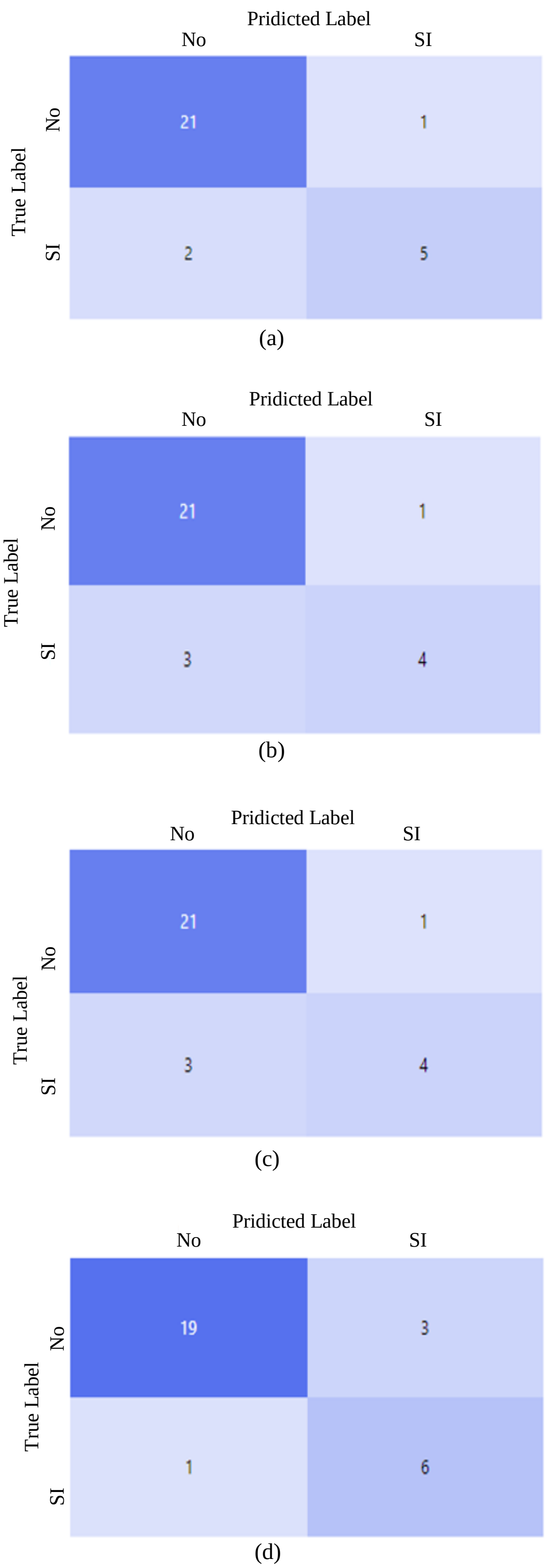

Fig. 4: Confusion matrix of trained algorithms

The RF algorithm (Fig. 4a), XGB algorithm (Fig. 4b) DT algorithm (Fig. 4c) demonstrated a higher number of correct predictions for true negatives, indicating their proficiency in identifying cases where addiction is not present. On the other hand, the LR algorithm (Fig. 4d) yielded more true positives. However, it also exhibited a higher number of false positives, suggesting a tendency to misclassify certain cases.

The results presented in Table 6 emphasize the performance of each algorithm in accurately predicting and minimizing errors.

Using the Azure Machine Learning Studio platform, the algorithms' performance was evaluated by calculating various metrics, providing a comprehensive perspective of the models' performance. The results, as shown in Table 7, offer insights into the algorithms' performance. It is worth mentioning that the results obtained through automated calculations may have slight variations compared to manually computed values from the confusion matrix.

The tree-based algorithms (RF, XGB DT) demonstrated higher accuracy, ranging from 83-87%, compared to LR. However, LR achieved better results in the sensitivity metric, surpassing 70%. Despite the similarity in results, the RF algorithm outperformed the others in this study, with a precision of 87.3% and an accuracy of 89.7%. These metrics indicate its lower likelihood of false positives and its overall ability to accurately classify smartphone addiction based on personality traits.

Table 5: Interpretation of the variables used for the confusion matrix

Code	Variable	Description
TP	True Positive	Correct predictions of adolescents with smartphone addiction based on their personality traits
TN	True Negative	Correct predictions of adolescents without smartphone addiction based on their personality traits
FP	False Positive	Incorrect predictions of adolescents with smartphone addiction based on their personality traits
FN	False Negative	Incorrect predictions of adolescents without smartphone addiction based on their personality traits

Table 6: Results of the number of correct and incorrect predictions according to the algorithm

Algorithm	Result	Correct predictions	Incorrect predictions
Random forest	With addiction	5	1
	Without addiction	21	2
XG boost	With addiction	4	1
	Without addiction	21	3
Decision tree	With addiction	4	1
	Without addiction	21	3
Logistic regression	With addiction	4	1
	Without addiction	21	3

Table 7: Comparative results of performance metrics of trained algorithms

Algorithm	Accuracy	Precision	Sensitivity	Fall Out	Specificity	AUC
Random forest	0.897	0.873	0.669	0.045	0.955	0.987
XG boost	0.862	0.838	0.526	0.045	0.955	0.948
Decision tree	0.862	0.857	0.526	0.045	0.955	0.938
Logistic regression	0.862	0.808	0.720	0.136	0.864	0.942

DOI: 10.3844/jcssp.2024.181.190

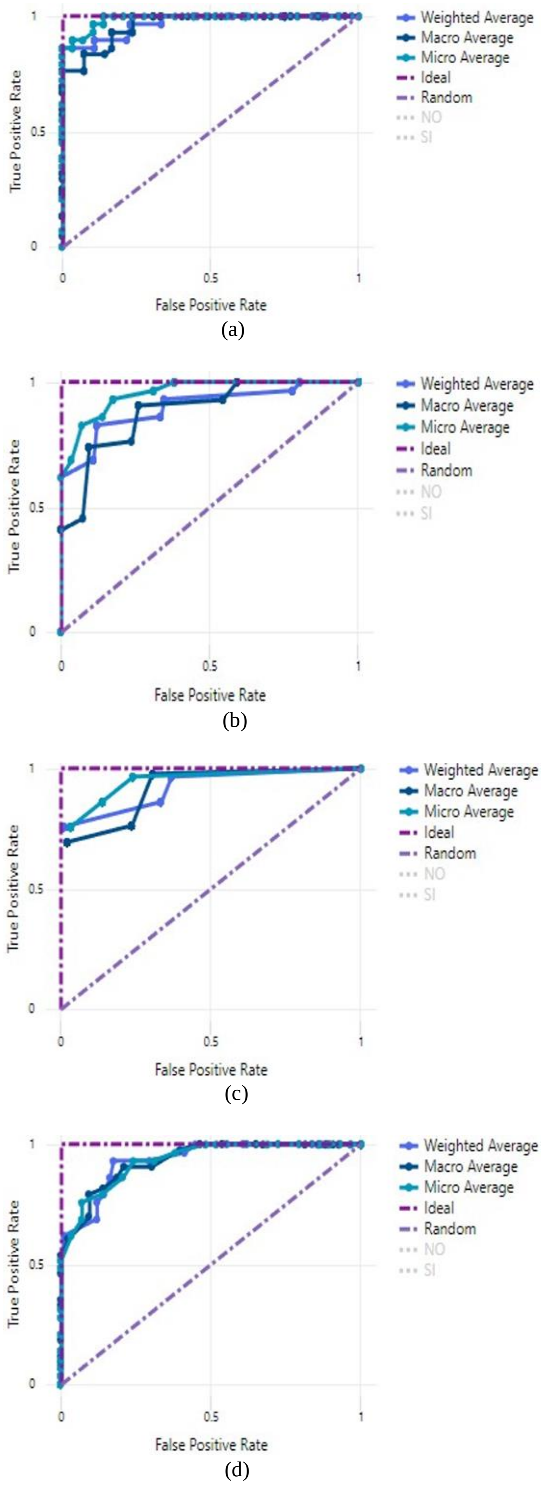

Fig. 5: ROC curve of the trained algorithms

To assess the performance of the classification models, the Area Under the Curve (AUC) metric was utilized, which offered insights into the relationship between "Fall out" and sensitivity. Figure 5 illustrates the ROC curves for each algorithm, all exhibiting values exceeding 90%. Notably, the ROC curve of the RF algorithm (Fig. 5a) demonstrates a closer proximity to 1 on the Y-axis, indicating a higher overall accuracy in distinguishing between positive and negative instances. It also demonstrates superior performance with a high ROC curve, signifying a greater probability of accurately classifying positive instances. This is supported by Table 7, which shows the highest AUC value of 98.7%.

When considering the influence of personality traits, the RF algorithm also exhibits the highest predictive performance for addiction. It is remarkable precision and sensitivity values validate its ability to effectively identify the relationship between personality traits and smartphone addiction. In contrast, the XGB (Fig. 5b), DT (Fig. 5c) LR (Fig. 5d) algorithms demonstrate lower precision and recall, indicating comparatively inferior performance.

The complexity of the RF algorithm, which arises from the ensemble nature of decision trees, was effectively managed through parameter tuning and feature selection techniques. This allowed us to strike a balance between model complexity and interpretability. We found that by carefully selecting hyperparameters and leveraging feature importance scores, we could maintain a reasonable level of model transparency while still preserving the algorithm's exceptional predictive capability.

Conclusion

Based on the proposed model using the RF algorithm, there is evidence of a relationship between personality traits and smartphone addiction in teenagers, highlighting the traits of neuroticism and conscientiousness. It is worth mentioning that, as of now no study has aimed to predict smartphone addiction based on personality traits.

Based on Pearson's correlation analysis, it was identified that the variables showing the strongest correlation with addiction are "Conscientiousness" (F02), "Neuroticism" (F05) "Age" (F07). Additionally, an inverse relationship was found between "Neuroticism" (F05) and all other personality traits, except for "Openness to experience" (F01). This finding substantiates previous research results, which have consistently shown similar outcomes for neuroticism and conscientiousness traits in relation to addiction (Müller *et al*., 2021).

Among the trained algorithms, the RF algorithm demonstrated superior suitability with an accuracy of 89.7% and a precision of 87.3%. It showed accurate predictions and high overall performance, achieving the highest AUC value in the ROC curve, which indicates better classification probability for positive instances.

In future work, it is recommended to incorporate new data to enhance the performance of the algorithms used. It is also important to evaluate the algorithms' performance with external data for cross-validation to ensure the robustness of the results. Finally, exploring other relevant variables that influence addiction prediction, such as the environment, social conditions, or economic resources could be very beneficial.

Acknowledgment

We are deeply grateful to the university, the school the students who conscientiously participated in our research by completing the surveys and thereby contributing to our database. Our appreciation also extends to the psychologist who played a crucial role in developing and validating the survey.

Funding Information

This research was financed by the research Department of the Universidad Peruana de Ciencias Aplicadas.

Author's Contributions

Jacobo Osorio and Marko Figueroa: Literature analysis, data collection, model and implementation of models, experimentation analysis of results. Manuscript written.

Lenis Wong: Study supervision, result analysis, manuscript reviewed discussion.

Ethics

The article is authentic and contains unpublished material. The corresponding author affirms that no ethical concerns exist all authors have read and endorsed the article.

References

Abu-Taieh, E. M., AlHadid, I., Kaabneh, K., Alkhawaldeh, R. S., Khwaldeh, S., Masa'deh, R. E., & Alrowwad, A. A. (2022). Predictors of Smartphone Addiction and Social Isolation among Jordanian Children and Adolescents Using SEM and ML. *Big Data and Cognitive Computing*, *6*(3), 92. https://doi.org/10.3390/BDCC6030092

Blasi, A. H., & Alsuwaiket, M. (2020). Analysis of students' misconducts in higher education using decision tree and ann algorithms. *Engineering, Technology & Applied Science Research*, *10*(6), 6510-6514. https://doi.org/10.48084/etasr.3927

Castillo-Viera, E., Gago-Valiente, F.-J., Giménez-Fuentes-Guerra, F.-J., Abad-Robles, M.-T., & Moreno-Sánchez, E. (2022). Physical Activity Programmes in the Treatment of Addictions: A Systematic Review. *Applied Sciences*, 12(18), 9117. https://doi.org/10.3390/app12189117

Cha, S. S., & Seo, B. K. (2018). Smartphone use and smartphone addiction in middle school students in Korea: Prevalence, social networking service and game use. *Health Psychology Open*, *5*(1), 2055102918755046. https://doi.org/10.1177/2055102918755046

Chen, W. Y., Yan, L., Yuan, Y. R., Zhu, X. W., Zhang, Y. H., & Lian, S. L. (2021). Preference for solitude and Mobile phone addiction among Chinese college students: the mediating role of psychological distress and moderating role of mindfulness. *Frontiers in Psychology*, *12*, 750511. https://doi.org/10.3389/fpsyg.2021.750511

Chen, Y., Zhang, X., Lu, L., Wang, Y., Liu, J., Qin, L., ... & Chen, M. C. (2022). Machine learning methods to identify predictors of psychological distress. *Processes*, *10*(5), 1030. https://doi.org/10.3390/PR10051030

Cheng, Y., & Meng, J. (2021). The association between depression and problematic smartphone behaviors through smartphone use in a clinical sample. *Human Behavior and Emerging Technologies*, *3*(3), 441-453. https://doi.org/10.1002/HBE2.258

Duan, L., He, J., Li, M., Dai, J., Zhou, Y., Lai, F., & Zhu, G. (2021). Based on a decision tree model for exploring the risk factors of smartphone addiction among children and adolescents in China during the COVID-19 pandemic. *Frontiers in Psychiatry*, *12*, 652356. https://doi.org/10.3389/fpsyt.2021.652356

Eichenberg, C., Schott, M., & Schroiff, A. (2021). Problematic smartphone use-comparison of students with and without problematic smartphone use in light of personality. *Frontiers in Psychiatry*, *11*, 599241. https://doi.org/10.3389/fpsyt.2020.599241

Erdem, C., & Uzun, A. M. (2022). Smartphone addiction among undergraduates: Roles of personality traits and demographic factors. *Technology, Knowledge and Learning*, *27*(2), 579-597. https://doi.org/10.1007/s10758-020-09467-1

Garriga, R., Mas, J., Abraha, S., Nolan, J., Harrison, O., Tadros, G., & Matic, A. (2022). Machine learning model to predict mental health crises from electronic health records. *Nature Medicine*, *28*(6), 1240-1248. https://doi.org/10.1038/s41591-022-01811-5

Haque, U. M., Kabir, E., & Khanam, R. (2021). Detection of child depression using machine learning methods. *PLoS One*, *16*(12), e0261131. https://doi.org/10.1371/JOURNAL.PONE.0261131

Ipsos. (2021). El Internet de Cada Día: Hábitos y Actitudes Hacia El Internet. https://www.ipsos.com/es-pe/el-internet-de-cada-dia-habitos-y-actitudes-hacia-el-internet-2021

Jiménez, O., Jesús, A., & Wong, L. (2023, May). Model for the Prediction of Dropout in Higher Education in Peru applying Machine Learning Algorithms: Random Forest, Decision Tree, Neural Network and Support Vector Machine. In *2023 33rd Conference of Open Innovations Association (FRUCT)* (pp. 116-124). IEEE. https://doi.org/10.23919/FRUCT58615.2023.10143068

Kim, S., Lee, H. K., & Lee, K. (2021). Screening of mood symptoms using MMPI-2-RF scales: An application of machine learning techniques. *Journal of Personalized Medicine*, *11*(8), 812. https://doi.org/10.3390/JPM11080812

Lee, J., & Kim, W. (2021). Prediction of problematic smartphone use: A machine learning approach. *International Journal of Environmental Research and Public Health*, *18*(12), 6458. https://doi.org/10.3390/ijerph18126458

Lei, L. Y. C., Ismail, M. A. A., Mohammad, J. A. M., & Yusoff, M. S. B. (2020). The relationship of smartphone addiction with psychological distress and neuroticism among university medical students. *BMC Psychology*, *8*, 1-9. https://doi.org/10.1186/S40359-020-00466-6

Li, X., Fu, S., Fu, Q., & Zhong, B. (2021). Youths' habitual use of smartphones alters sleep quality and memory: Insights from a national sample of Chinese students. *International Journal of Environmental Research and Public Health*, *18*(5), 2254. https://doi.org/10.3390/ijerph18052254

Li, Y., Wang, Z., You, W., & Liu, X. (2022). Core self-evaluation, mental health and mobile phone dependence in Chinese high school students: Why should we care. *Italian Journal of Pediatrics*, *48*(1), 28. https://doi.org/10.1186/S13052-022-01217-6

Makino, K., Lee, S., Bae, S., Chiba, I., Harada, K., Katayama, O., ... & Shimada, H. (2021). Development and validation of new screening tool for predicting dementia risk in community-dwelling older Japanese adults. *Journal of Translational Medicine*, *19*, 1-9. https://doi.org/10.1186/S12967-021-03121-9

Marengo, D., Sindermann, C., Häckel, D., Settanni, M., Elhai, J. D., & Montag, C. (2020). The association between the Big Five personality traits and smartphone use disorder: A meta-analysis. *Journal of Behavioral Addictions*, *9*(3), 534-550. https://doi.org/10.1556/2006.2020.00069

Minsa. (2021). La Pandemia COVID-19 Ha Incrementado La Adicción a Los Videojuegos En Niños y Adolescentes. https://www.gob.pe/institucion/minsa/noticias/348005-minsa-la-pandemia-covid-19-ha-incrementado-la-adiccion-a-los-videojuegos-en-ninos-y-adolescentes

Müller, M., Sindermann, C., Rozgonjuk, D., & Montag, C. (2021). Mind-wandering mediates the associations between neuroticism and conscientiousness, and tendencies towards smartphone use disorder. *Frontiers in Psychology*, *12*, 661541. https://doi.org/10.3389/fpsyg.2021.661541

Nida. (2022). La Adicción y La Salud. https://nida.nih.gov/es/publicaciones/las-drogas-el-cerebro-y-la-conducta-la-ciencia-de-la-adiccion/la-adiccion-y-la-salud

Peltonen, E., Sharmila, P., Asare, K. O., Visuri, A., Lagerspetz, E., & Ferreira, D. (2020). When phones get personal: Predicting Big Five personality traits from application usage. *Pervasive and Mobile Computing*, *69*, 101269. https://doi.org/10.1016/J.PMCJ.2020.101269

Peralta, F. C. (2014, October). Fases del modelo de proceso de la metodología CRISP-DM. https://www.researchgate.net/figure/Fases-del-modelo-de-proceso-de-la-metodologia-CRISP-DM-35_fig2_284215308

Peterka-Bonetta, J., Sindermann, C., Elhai, J. D., & Montag, C. (2019). Personality associations with smartphone and internet use disorder: A comparison study including links to impulsivity and social anxiety. *Frontiers in Public Health*, *7*, 127. https://doi.org/10.3389/FPUBH.2019.00127

Razavi, R., Gharipour, A., & Gharipour, M. (2020). Depression screening using mobile phone usage metadata: A machine learning approach. *Journal of the American Medical Informatics Association*, *27*(4), 522-530. https://doi.org/10.1093/jamia/OCZ221

Rho, M. J., Park, J., Na, E., Jeong, J. E., Kim, J. K., Kim, D. J., & Choi, I. Y. (2019). Types of problematic smartphone use based on psychiatric symptoms. *Psychiatry Research*, *275*, 46-52. https://doi.org/10.1016/J.PSYCHRES.2019.02.071

Toyama, M., & Hayashi, Y. (2022). Links of personality traits to media multitasking: Conscientiousness predicts mobile phone use in the college classroom. *Current Psychology*, *41*(12), 8660-8667. https://doi.org/10.1007/S12144-020-01258-2

Wickord, L. C., & Quaiser-Pohl, C. (2022). Psychopathological symptoms and personality traits as predictors of problematic smartphone use in different age groups. *Behavioral Sciences*, *12*(2), 20. https://doi.org/10.3390/bs12020020

Xia, F., Li, Q., Luo, X., & Wu, J. (2022). Machine learning model for depression based on heavy metals among aging people: A study with National Health and Nutrition Examination Survey 2017-2018. *Frontiers in Public Health*, *10*, 939758. https://doi.org/10.3389/FPUBH.2022.939758

Zeighami, R., Hosseini, S. M., Mohammadi, N., & Shahsavari, S. (2021). Predicting addiction susceptibility based on big five personality traits. *Iranian Journal of Psychiatry and Behavioral Sciences*, *15*(4). https://doi.org/10.5812/IJPBS.109103