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

Jacobos Osorio, Marko Figueroa and Lenis Wong

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

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 to identify correlations between personality traits and addiction. 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% and 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) 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 a strongly 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 (XGB) 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: CRISP-DM methodology (Peralta, 2014)](Image)

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

<table>
<thead>
<tr>
<th>Reference</th>
<th>Personality trait</th>
<th>Description</th>
<th>Low trait</th>
<th>High trait</th>
</tr>
</thead>
<tbody>
<tr>
<td>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)</td>
<td>Neuroticism</td>
<td>Index of emotional stability and impulse control</td>
<td>Security and trust</td>
<td>Sensitivity and nervousness</td>
</tr>
<tr>
<td>Eichenberg et al. (2021); Erdem and Uzun (2022); Peltonen et al. (2020); Toyama and Hayashi (2022); Wickord and Quaiser-Pohl (2022)</td>
<td>Conscientiousness</td>
<td>It involves self-discipline Organization and personal responsibility</td>
<td>Carelessness lack of planning</td>
<td>Efficiency and organization</td>
</tr>
<tr>
<td>Eichenberg et al. (2021); Erdem and Uzun (2022); Peltonen et al. (2020); Toyama and Hayashi (2022); Wickord and Quaiser-Pohl (2022)</td>
<td>Extraversion</td>
<td>Characterized by energy and taste for social interaction</td>
<td>Reserve and preference for solitude</td>
<td>Expressiveness and search for social interactions courtesy and cooperation</td>
</tr>
<tr>
<td>Eichenberg et al. (2021); Erdem and Uzun (2022); Peltonen et al. (2020); Peterka-Bonetta et al. (2019); Wickord and Quaiser-Pohl (2022)</td>
<td>Agreeableness</td>
<td>It reflects a willingness to be kind, compassionate cooperative</td>
<td>Defiance and estrangement</td>
<td></td>
</tr>
<tr>
<td>Eichenberg et al. (2021); Erdem and Uzun (2022); Peltonen et al. (2020); Wickord and Quaiser-Pohl (2022)</td>
<td>Openness</td>
<td>Intellectual curiosity, creativity is a preference for novelty</td>
<td>Caution and conventional thinking</td>
<td>Imagination and curiosity</td>
</tr>
</tbody>
</table>
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 and 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>
<thead>
<tr>
<th>ID</th>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>F01</td>
<td>Openness</td>
<td>Indicates the respondent's level of openness to the experience</td>
</tr>
<tr>
<td>F02</td>
<td>Conscientiousness</td>
<td>Indicates the level of awareness of the respondent</td>
</tr>
<tr>
<td>F03</td>
<td>Extraversion</td>
<td>Indicates the level of extraversion of the respondent</td>
</tr>
<tr>
<td>F04</td>
<td>Agreeableness</td>
<td>Indicates the level of friendliness of the respondent</td>
</tr>
<tr>
<td>F05</td>
<td>Neuroticism</td>
<td>Indicates the level of neuroticism or emotional instability of the respondent</td>
</tr>
<tr>
<td>F06</td>
<td>Smartphone Addiction</td>
<td>Indicates the respondent's level of openness to the experience</td>
</tr>
<tr>
<td>F07</td>
<td>Age</td>
<td>Indicates the respondent's age in years</td>
</tr>
</tbody>
</table>

Table 3: Variables used for the predictive model

<table>
<thead>
<tr>
<th>ID</th>
<th>Median</th>
<th>Description</th>
<th>No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>F01</td>
<td>34.99</td>
<td>6,204</td>
<td>96</td>
</tr>
<tr>
<td>F02</td>
<td>29.31</td>
<td>5,935</td>
<td>96</td>
</tr>
<tr>
<td>F03</td>
<td>26.20</td>
<td>7,106</td>
<td>96</td>
</tr>
<tr>
<td>F04</td>
<td>32.88</td>
<td>5,155</td>
<td>96</td>
</tr>
<tr>
<td>F05</td>
<td>23.68</td>
<td>6,036</td>
<td>96</td>
</tr>
<tr>
<td>F06</td>
<td>24.82</td>
<td>10,154</td>
<td>96</td>
</tr>
<tr>
<td>F07</td>
<td>14.01</td>
<td>1,244</td>
<td>96</td>
</tr>
</tbody>
</table>

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

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}
\]

\[
\text{Precision} = \frac{TP}{TP + FP} \tag{2}
\]

\[
\text{Sensitivity} = \frac{TP}{TP + FN} \tag{3}
\]

\[
\text{Specificity} = \frac{TN}{TN + FP} \tag{4}
\]

\[
\text{Fall Out} = \frac{FP}{FP + TN} \tag{5}
\]

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.
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

<table>
<thead>
<tr>
<th>Code</th>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>True Positive</td>
<td>Correct predictions of adolescents with smartphone addiction based on their personality traits</td>
</tr>
<tr>
<td>TN</td>
<td>True Negative</td>
<td>Correct predictions of adolescents without smartphone addiction based on their personality traits</td>
</tr>
<tr>
<td>FP</td>
<td>False Positive</td>
<td>Incorrect predictions of adolescents with smartphone addiction based on their personality traits</td>
</tr>
<tr>
<td>FN</td>
<td>False Negative</td>
<td>Incorrect predictions of adolescents without smartphone addiction based on their personality traits</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Result</th>
<th>Correct predictions</th>
<th>Incorrect predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random forest</td>
<td>With addiction</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Without addiction</td>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td>XG boost</td>
<td>With addiction</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Without addiction</td>
<td>21</td>
<td>3</td>
</tr>
<tr>
<td>Decision tree</td>
<td>With addiction</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Without addiction</td>
<td>21</td>
<td>3</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>With addiction</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Without addiction</td>
<td>21</td>
<td>3</td>
</tr>
</tbody>
</table>

**Table 7:** Comparative results of performance metrics of trained algorithms

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Sensitivity</th>
<th>FallOut</th>
<th>Specificity</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random forest</td>
<td>0.917</td>
<td>0.937</td>
<td>0.669</td>
<td>0.045</td>
<td>0.955</td>
<td>0.987</td>
</tr>
<tr>
<td>XG boost</td>
<td>0.862</td>
<td>0.838</td>
<td>0.526</td>
<td>0.045</td>
<td>0.955</td>
<td>0.987</td>
</tr>
<tr>
<td>Decision tree</td>
<td>0.862</td>
<td>0.857</td>
<td>0.526</td>
<td>0.045</td>
<td>0.955</td>
<td>0.987</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>0.862</td>
<td>0.808</td>
<td>0.720</td>
<td>0.136</td>
<td>0.864</td>
<td>0.942</td>
</tr>
</tbody>
</table>
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

Jacob 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


