POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

EXECUTIVE SUMMARY OF THE THESIS

Understanding cybervictimization through smartphone addiction, online gambling and routine activities: an integrated approach using PLS-SEM and Machine Learning

TESI MAGISTRALE IN MANAGEMENT ENGINEERING - INGEGNERIA GESTIONALE

Author: **Federico Trevisan**

POLIMI Advisor: **Sergio Terzi**

UPM Advisor: **Alberto Urueña López**

Academic Year: **2025-2026**

Abstract

The rapid expansion of digital technologies and the intensification of online activities have significantly increased cybervictimization risks, highlighting the need to understand the behavioural and situational mechanisms underlying vulnerability in digital environments. This thesis investigates the role of Smartphone Addiction, online gambling behaviour, and digital routines in explaining cybervictimization risk, integrating the Lifestyle-Routine Activity Theory (L-RAT) within a unified empirical framework. The study relies on a large national dataset and adopts an integrated methodological approach combining Partial Least Squares Structural Equation Modelling (PLS-SEM) and Machine Learning techniques. L-RAT is modelled as a second-order formative-formative construct capturing Exposure, Proximity, Target Suitability, and Capable Guardianship. PLS-SEM is employed to test the structural relationships among constructs and assess mediation mechanisms, while a Random Forest model evaluates predictive performance and variable importance.

The results show that Smartphone Addiction significantly increases cybervictimization risk both directly and indirectly through L-RAT mechanisms, with a substantial portion of its total effect transmitted via routine activity structures. In particular, Proximity to motivated offenders emerges as a central explanatory dimension. Online gambling behaviour amplifies routine-based exposure processes, whereas Social Support plays a protective role. The comparison between structural modelling and predictive modelling reveals convergence between theoretical relevance and predictive importance of key variables, while demonstrating meaningful discriminative capacity in classification performance.

The study contributes to the literature in two main ways. First, it reinforces the applicability of Lifestyle-Routine Activity Theory in contemporary digital contexts by integrating behavioural addiction mechanisms into the criminological framework. Second, it demonstrates the value of combining explanatory structural modelling with machine learning algorithms to jointly address theory testing and predictive risk assessment in the analysis of online behavioural vulnerability.

Key-words: Cybervictimization; Smartphone Addiction; Online Gambling; Lifestyle-Routine Activity Theory; PLS-SEM; Machine Learning; Random Forest; Digital Behaviour.

1 Introduction

The progressive digitalisation of everyday life has profoundly transformed social interaction, economic exchange, and leisure activities. Smartphones have become central access devices to digital environments, enabling continuous connectivity, social networking, financial transactions, and online gambling. While these developments have expanded opportunities for communication and participation, they have also increased exposure to cyber risks, including fraud, phishing, identity theft, and other forms of cybervictimization.

Cybervictimization is not randomly distributed across users. Contemporary research increasingly suggests that vulnerability in digital environments depends on behavioural patterns, situational exposure, and individual psychosocial characteristics, reflecting structured opportunity configurations rather than isolated incidents. In this context, two parallel strands of literature remain only partially integrated. On the one hand, criminological approaches, particularly Lifestyle-Routine Activity Theory (L-RAT), argue that victimization risk increases when motivated offenders encounter suitable targets in the absence of capable guardians. In digital contexts, routine online activities, exposure to risky environments, and insufficient protective behaviours may replicate the structural conditions described by the theory.

On the other hand, behavioural addiction research highlights the role of problematic or addictive smartphone use as a driver of maladaptive online behaviours. Smartphone Addiction has been associated with impulsivity, sensation-seeking, reduced self-regulation, and intensified digital engagement. These characteristics may indirectly increase cyber risk by amplifying exposure and weakening protective routines. However, empirical models integrating addiction mechanisms with criminological frameworks remain limited.

Online gambling represents a particularly relevant behavioural domain in this regard. Gambling applications are highly accessible through smartphones, often embedded within social and interactive environments. Individuals engaged in online gambling may display intensified digital routines, frequent financial transactions, and repeated interactions in environments where motivated offenders operate. Yet, the extent to which gambling behaviours amplify routine-based exposure mechanisms leading to cybervictimization has not been systematically modelled within a unified structural framework.

Within this context, the present thesis is developed within the research projects "Smartphone addiction, loss of psychosocial well-being and online victimization of online gamblers (*CIBERADICJUEGO2024*), Reference SUBV24/00012" and "The commodification of human behaviour: Data sharing and privacy threats among online gamblers, Reference SUBV25/00014", funded by the Spanish Ministry of Social Rights, Consumption and Agenda 2030, DG Gambling. These projects investigate the relationships between smartphone addiction, psychosocial well-being, data sharing practices, and online victimization in digital gambling environments, providing the empirical foundation for the analytical strategy adopted in this thesis.

The study is guided by the following research questions:

RQ1. *To what extent do Lifestyle-Routine Activity patterns directly explain cybervictimization risk in digital environments? (Refers to H1)*

RQ2. *To what extent do Smartphone Addiction and Online Gambling behaviours directly contribute to cybervictimization risk? (Refers to H2 and H4)*

RQ3. *Through which mediating mechanisms do Smartphone Addiction and Online Gambling behaviours influence cybervictimization risk? (Refers to H3, H5 and H6)*

RQ4. *How do psychosocial factors, such as Social Digital Pressure and Social Support, shape Smartphone Addiction and indirectly structure cybervictimization risk? (Refers to H7 and H8)*

By addressing these questions, the thesis aims to develop a theoretically grounded and empirically validated framework capable of explaining cybervictimization risk in contemporary digital contexts. The study explicitly bridges criminological theory and behavioural addiction research while integrating explanatory and predictive modelling approaches.

2 Theoretical Framework

2.1 Smartphone Addiction

Smartphone Addiction is defined as a behavioural addiction characterized by compulsive use, withdrawal symptoms, and functional impairment in daily life. Consistent empirical evidence indicates a direct association between smartphone addiction and cybervictimization. In particular, Herrero et al. [1] demonstrated that individuals with higher levels of smartphone addiction are significantly more likely to experience cybervictimization, even when controlling for other psychological and behavioural factors.

In this study, SA is not merely treated as an isolated psychological trait but as a primary driver that reshapes digital opportunity structures. Beyond its direct impact, emerging evidence suggests that smartphone addiction is associated with other risky digital behaviours, including online gambling. Smartphones provide constant and immediate access to gambling platforms, lowering situational barriers to participation and facilitating impulsive engagement. Studies indicate that individuals with higher levels of smartphone addiction are more likely to engage in online gambling behaviours and display related risk profiles, pointing to an additional indirect pathway through which smartphone addiction may contribute to cybervictimization [1], [2]. Consequently, SA acts as a central node in the model, both as a direct predictor of risk and as a catalyst for further problematic online activities.

2.2 Online Gambling Behaviour

Online gambling has evolved into a widespread digital leisure activity, facilitated by the normalization of online financial transactions and constant device connectivity. Unlike traditional gambling, the online environment removes spatial and temporal constraints, increasing the frequency and intensity of engagement. In this study, gambling is defined as the wagering of value on probabilistic outcomes, encompassing activities such as sports betting and digital casinos. This transformation has significantly altered user risk profiles, as online platforms lower entry barriers and provide immediate rewards, often fostering impulsive engagement and patterns comparable to behavioural addictions [3], [4].

Beyond its psychological impact, online gambling is a critical factor in cybervictimization due to the specific behavioural patterns it mandates. Individuals engaging in online gambling spend extended periods online, perform frequent financial transactions, and regularly disclose sensitive personal data [5], [6]. These practices place users in digital environments with heightened exposure to malicious actors. Consequently, gambling may exert a direct effect on victimization risk by fostering behaviours associated with reduced caution. Furthermore, gambling participation reshapes routine activities by requiring repeated interactions with complex digital ecosystems. These routines increase exposure to offenders and amplify "target suitability" through constant financial activity, while potentially weakening "capable guardianship" by fostering habituation to digital risk [7], [2]. Thus, online gambling is hypothesized to indirectly elevate risk by acting as a behavioural catalyst that modifies the L-RAT framework.

2.3 Psychosocial antecedents: Social Digital Pressure and Social Support

The transition toward a "constantly online" society has redefined social expectations, introducing Social Digital Pressure (SDP) as a critical environmental driver of behaviour. Unlike internal psychological traits, SDP emerges from contextual demands for availability and responsiveness, structurally reinforced by platform features such as read receipts and activity indicators [8], [9]. Individuals experiencing high SDP perceive a social obligation to maintain persistent visibility, transforming optional engagement into a requirement. Empirical evidence indicates that SDP is a significant predictor of SA, as users increase device usage to comply with external social norms rather than intrinsic motivation [1].

In contrast, Social Support (SS), defined as the perception of being valued and embedded in a meaningful network, serves as a primary protective factor [10]. Within the digital domain, Social Support influences behaviour through a compensatory mechanism: individuals with lower perceived offline support often seek validation and emotional regulation through digital feedback, such as likes and instant messaging [11], [12]. This reliance on smartphones to fulfil unmet social needs facilitates the development of addictive patterns. Consistent with this, research shows that Social Support is negatively associated with Smartphone Addiction, such that higher levels of support correspond to lower levels of addictive use, while also suggesting that high digital pressure may erode these protective social resources over time [13], [14].

2.4 Lifestyle-Routine Activity Theory

Traditionally, criminological research focused on offender motivations. However, Lifestyle-Routine Activity Theory (L-RAT) shifted the focus toward how victims' daily routines create "situational opportunities" for crime. Originally developed by Cohen and Felson and by Hindelang et al. [15], [16] to explain physical property crimes, the theory posits that victimization occurs when a motivated offender and a suitable target converge in space and time in the absence of capable guardianship. In this thesis, L-RAT is adapted to the digital ecosystem, where physical co-presence is replaced by situational convergence within digital platforms and transactional contexts. Cybervictimization is thus treated not as a random event or a purely technical failure, but as a systematic outcome of the "human factor", the routine behaviours and digital lifestyles of users.

To operationalize this framework, L-RAT is modelled as a second-order formative-formative construct [17], [18] comprising four interrelated dimensions:

Exposure to motivated offenders: the intensity of online presence. High exposure in unregulated digital environments increases the probability of encountering threats as a byproduct of routine behaviour.

Proximity to motivated offenders: reflects interactional density. It represents how closely users come into contact with offenders through communication channels (e.g.: social networks, instant messaging), regardless of geographical distance.

Target suitability: the attractiveness or vulnerability of a user, signalled through digital traces and the disclosure of sensitive personal or financial information.

Capable guardianship: the combination of technical safeguards (passive) and cautious behavioural strategies (active). Insufficient guardianship, often due to overreliance on technology or lack of awareness, significantly increases risk.

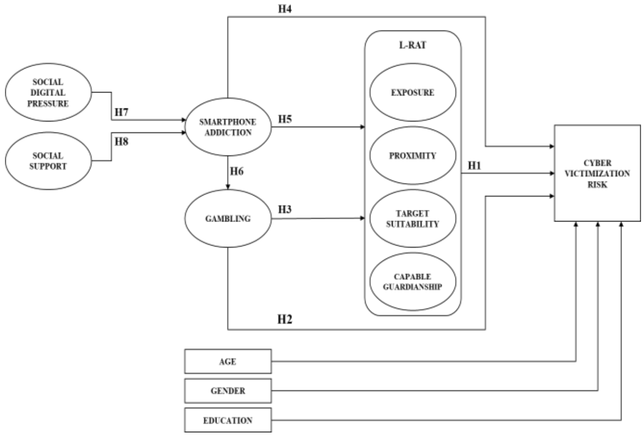
By integrating these dimensions, this study tests the hypothesis that higher levels of exposure, proximity, and suitability, coupled with lower guardianship, directly increase cybervictimization risk. Furthermore, L-RAT serves as the primary mediating mechanism through which behavioural drivers, such as Smartphone Addiction and Online Gambling, exert their indirect influence on risk.

2.5 The proposed research framework

The present study integrates behavioural addiction mechanisms with criminological opportunity structures into a unified explanatory system, as illustrated in Figure 1. The core of this integrated architecture is the conceptualization of Smartphone Addiction and Online Gambling not just as isolated behaviours, but as structural antecedents that actively reshape the L-RAT dimensions.

This integrated architecture advances the literature in three key respects. First, it bridges the gap between behavioural psychology and criminology by explicitly incorporating addiction-related constructs (Smartphone Addiction and Online Gambling) as primary drivers that actively restructure digital routine exposure. Second, it adopts a rigorous methodological approach by modelling L-RAT as a second-order formative-formative construct, preserving the theoretical nuances of opportunity dimensions while assessing their distinct empirical contributions. Third, it shifts the focus toward a systemic view, where cybervictimization risk emerges from the complex interaction between psychosocial antecedents and situational digital routines. By treating L-RAT as a mediating mechanism, the framework enables a precise evaluation of how individual predispositions are transformed into systematic digital vulnerability.

Figure 1: *Structural model*



3 Methodology

The methodological framework of this study adopts a dual analytical approach, integrating theory-driven explanatory modelling with data-driven predictive analytics. By combining Partial Least Squares Structural Equation Modelling (PLS-SEM) and Machine Learning (Random Forest), the research aims to both test complex structural relationships and evaluate the practical predictability of cybervictimization risk [19].

3.1 Data source and sample characteristics

The empirical analysis is based on data from the National Survey on Cybersecurity and Trust in Spanish Households, conducted by the Ministry of Economy and Digital Transformation between January 2019 and June 2021. This large-scale survey provides a comprehensive snapshot of digital habits and security incidents across the Spanish population.

To ensure data integrity, a rigorous pre-processing pipeline is implemented. First, observations are restricted to the 2019-2021 timeframe to prevent duplication of observations across waves. Second, incomplete responses are removed to avoid non-random missing data bias, which is critical for stable latent construct estimation. Finally, demographic variables are re-categorised (e.g.: educational levels and age groups) to enhance model parsimony and computational efficiency. After data cleaning and preprocessing, the final analytical sample consists of N = 7,064 respondents, representative of Spanish internet users. The sample shows a higher proportion of individuals with university education (48.7%), consistent with national patterns of technology adoption.

3.2 PLS-SEM: measurement and structural modelling

The first phase employs PLS-SEM to test the theoretical hypotheses (H1-H8) and explore mediation pathways. PLS-SEM is used since this method is particularly appropriate when the research objective involves complex structural models including higher-order constructs and formative specifications [19], [18]. Unlike covariance-based SEM, PLS-SEM prioritizes variance explanation in endogenous constructs rather than reproducing the observed covariance matrix [20].

Measurement model evaluation: Before testing the hypotheses, the quality of the constructs is assessed. SA, SDP, and SS are specified as reflective constructs. In reflective measurement models, indicators are assumed to be manifestations of an underlying latent construct. Indicator loadings are therefore expected to be high and interchangeable [21], [18]. For reflective constructs (SA, SDP, SS), measurement evaluation follows criteria shown in Table 1. The L-RAT framework is modelled as a second-order formative construct composed of four first-order formative dimensions: Exposure, Proximity, Target Suitability, and Capable Guardianship. In formative measurement models, indicators define and constitute the construct rather than reflect it. Indicators are not interchangeable, and internal consistency metrics are not appropriate [21], [19]. Evaluation focuses on criteria presented in Table 2.

Table 1**:** *Evaluation of reflective measurement models*

Evaluation	Method / Indicator	Threshold
Indicator reliability	Outer loadings	≥ 0.708
Internal consistency reliability	Cronbach's Alpha	≥ 0.70
	rhoA	≥ 0.70
	Composite Reliability (rhoC)	0.70 - 0.95
Convergent validity	Average Variance Extracted (AVE)	≥ 0.50
Discriminant validity	HTMT ratio	< 0.85 (strict) < 0.90 (lenient)
	Fornell-Larcker criterion	√AVE > inter-construct correlations

Table 2: Evaluation of formative measurement models

Evaluation	Method / Indicator	Threshold
Convergent validity	Redundancy analysis	Path coefficient ≥ 0.708
Indicator collinearity	VIF	< 5 (preferably < 3)
Indicator relevance	Indicator weights (bootstrap CI)	CI does not include 0
Indicator importance	Indicator loadings	Examined when weights are non-significant

The L-RAT framework is operationalised as a formative-formative Higher-Order Construct (HOC) using a disjoint two-stage approach [22]. In the first stage, the latent variable scores for the four lower-order dimensions (Exposure, Proximity, Suitability, and Guardianship) are calculated within a model where they are linked directly to the endogenous variables. In the second stage, these scores are used as manifest indicators to define the L-RAT HOC. This method ensures that the measurement of the higher-order abstraction is not biased by the structural relationships of the overall model. The validation of this construct focuses on two pillars: collinearity, ensured by Variance Inflation Factor (VIF) values consistently below the threshold of 3, and indicator relevance, confirmed through bootstrapping (10,000 subsamples) to verify that the outer weights of all four dimensions are statistically significant and contribute meaningfully to the composite construct.

Structural model evaluation: The structural model specifies directional relationships among latent constructs based on the theoretical framework. In PLS-SEM, structural paths are estimated to maximize explained variance (R^2) of endogenous constructs [20]. Structural model evaluation is resumed in Table 3.

Table 3: Evaluation of the structural model

Evaluation	Method / Indicator	Threshold
Collinearity	VIF (latent variables)	< 5 (preferably < 3)
Path significance	Path coefficients (bootstrapping)	$t > 1.96$ ($\alpha = .05$)
Explanatory power	Coefficient of determination (R^2)	0.75 (substantial) 0.50 (moderate), 0.25 (weak)
	Effect size (f^2)	0.02 (small) 0.15 (medium) 0.35 (large)

Coefficient of determination (R^2) and Effect size (f^2) metrics provide complementary information regarding explanatory strength and predictor relevance [19], [18].

Mediation analysis: The structural model evaluates complex mediation pathways by calculating indirect effects through bootstrapping. To classify the nature of these mediations, the study follows the decision tree proposed by Zhao et al. [23]. The analysis first establishes the significance of the indirect effect. If significant, it examines the sign and significance of the direct effect: a significant direct effect with the same sign as the indirect one indicates complementary mediation, while an opposite sign suggests competitive mediation. If the direct effect is non-significant, the relationship is classified as indirect-only mediation (full mediation). To quantify the strength of these mechanisms, the Variance Accounted For (VAF) is calculated, representing the ratio of the indirect effect to the total effect, thereby providing a clear measure of how much of the independent variable's impact is captured by the mediator.

3.3 ML: predictive analytics via Random Forest

A Random Forest classifier is employed for predictive modelling [24]. RF is an ensemble learning method based on multiple decision trees constructed through bootstrap aggregation and random feature selection. It is robust to multicollinearity, handles nonlinear relationships, and provides measures of variable importance. For this phase, a specific data transformation is applied: all continuous indicators are rescaled using the Percent of Maximum Possible (POMP) method to a 0-100 range. This facilitates comparability across indicators and enhances interpretability of feature importance. The dataset is divided into training (80%) and test (20%) subsets using stratified sampling to preserve class distribution. Model tuning and validation are conducted within the training set to prevent information leakage [25]. The RF model is trained using a cross-validation approach to prevent overfitting, focusing on two main outputs:

Classification outcome: Measured via Accuracy, AUC, and Confusion matrix (Precision, Recall, F1-score) to determine the model's performance in identifying risk.

Variable importance: Unlike SEM coefficients, RF provides a ranking of "Feature Importance," identifying which specific behaviours or routine factors contribute most to reducing classification error, regardless of linear constraints.

3.4 Integrated analytical strategy

Recent methodological discussions emphasize the complementarity of explanatory and predictive modelling [26]. Structural modelling clarifies causal structures and mediation mechanisms, while predictive modelling evaluates practical risk classification and variable relevance. The novelty of this methodology lies in the synergy between the two techniques. The Integrated analytical strategy follows a sequential logic where PLS-SEM provides the theoretical validation of the paths, and Machine Learning confirms the practical relevance of those paths through predictive strength. This “hybrid” approach ensures that the results are not only statistically significant, but also robust enough to be used for cybersecurity interventions and preventive policies.

4 Results

This chapter presents the main empirical results of the study, structured in two main phases: the explanatory analysis using PLS-SEM to test the structural hypotheses and the predictive analysis using Random Forest to evaluate the classification performance of the model.

4.1 PLS-SEM: structural model and hypothesis testing

Before assessing the structural relationships, the measurement model is validated to ensure reliability and validity. All reflective constructs (Smartphone Addiction, Social Support, and Social Digital Pressure) meet the criteria for internal consistency (Cronbach’s Alpha, rhoC and rhoA > 0.70) and convergent validity (AVE > 0.50). Discriminant validity is also confirmed via both the Fornell-Larcker criterion and the HTMT ratio (< 0.85). For the formative L-RAT dimensions, multicollinearity is ruled out (VIF < 3). Detailed numerical results for the PLS-SEM model are provided in Appendix A.

The validation of L-RAT as a Higher-Order Construct (HOC) is performed using a disjoint two-stage approach. Given the formative nature of the construct, the evaluation focused on collinearity diagnostics and indicator relevance. Variance Inflation Factor (VIF) values for the formative indicators are below the recommended threshold of 3, indicating absence of problematic multicollinearity. Bootstrapped confidence intervals for indicator weights confirmed the statistical relevance of the formative dimensions. Where indicator weights are non-significant but loadings remained substantial, theoretical justification supported their retention, preserving conceptual coverage of the L-RAT framework. The results confirm that the higher-order formative specification of L-RAT is empirically supported and theoretically coherent.

The structural model results (summarised in Table 4) support most of the proposed hypotheses. L-RAT patterns exhibit a strong direct association with cybervictimization (H1: $\beta = 0.321$, $p < 0.001$). Regarding behavioural drivers, Smartphone Addiction (H4: $\beta = 0.151$) and Online Gambling (H2: $\beta = 0.062$) both show significant direct effects on risk. Regarding psychosocial antecedents, SS exhibits a strong and statistically significant negative association with SA (H8: $\beta = -0.400$), confirming its protective role. In contrast, SDP shows a small but statistically significant negative relationship with SA (H7: $\beta = -0.058$), contrary to the hypothesized positive association. Therefore, H7 is not supported. The magnitude of this effect is limited, suggesting that the influence of digital pressure on addictive patterns may be more context-dependent than expected. Finally, the model explains 19.4% of the variance in cybervictimization risk, 17.1% of L-RAT patterns, and 16.8% of Smartphone Addiction.

Overall, the structural results confirm that cybervictimization risk is primarily structured by routine activity mechanisms, with behavioural addiction acting as both a direct and indirect driver of digital vulnerability.

4.2 PLS-SEM: mediation analysis

The mediation analysis, following the framework of Zhao et al. [23], reveals the mechanisms through which behaviours are transformed into risk. As shown in Table 5, the relationship between Smartphone Addiction and cybervictimization is partially mediated by both L-RAT patterns and Online Gambling.

Table 4: *Structural Model results, second-order model*

Endogenous construct	Coefficient of determination (R^2)	Relationship			Total Sample (n = 7064) Inner VIF	Path coef β	BC CI 95%
Smartphone Addiction	**0.168** [0.159 ; 0.176]	SS	→	SA	1.007	-0.400	[-0.422 ; -0.378]
		SDP	→	SA	1.007	-0.058	[-0.079 ; -0.037]
Gambling	**0.019** [0.015 ; 0.022]	SA	→	Gambling	*NA*	0.137	[0.112 ; 0.162]
LRAT	**0.171** [0.161 ; 0.179]	SA	→	LRAT	1.019	0.340	[0.315 ; 0.362]
		Gambling	→	LRAT	1.019	0.193	[0.164 ; 0.221]
Target Variable pc1a_Riesgo	**0.194** [0.183 ; 0.205]	SA	→	pc1a_Riesgo	1.263	0.151	[0.126 ; 0.175]
		Gambling	→	pc1a_Riesgo	1.084	0.062	[0.034 ; 0.091]
		LRAT	→	pc1a_Riesgo	1.215	0.321	[0.291 ; 0.351]
		Age	→	pc1a_Riesgo	1.127	-0.062	[-0.086 ; -0.038]
		Gender	→	pc1a_Riesgo	1.048	-0.009	[-0.031 ; 0.013]
		Education	→	pc1a_Riesgo	1.004	-0.002	[-0.023 ; 0.019]

Table 5: *Mediation analysis results*

Mediator	Total Sample (n = 7064) Total effect		Direct effect		Indirect effect		
LRAT			**SA → pc1a_Riesgo**		**SA → LRAT → pc1a_Riesgo**		
	Coefficient	BC CI 95%	Coefficient	BC CI 95%	Coefficient	BC CI 95%	VAF
	0.260 ***	[0.235 ; 0.284]	0.151 ***	[0.125 ; 0.175]	0.109 ***	[0.097 ; 0.122]	42.062%
			Gambling → pc1a_Riesgo		**Gambling → LRAT → pc1a_Riesgo**		
	Coefficient	BC CI 95%	Coefficient	BC CI 95%	Coefficient	BC CI 95%	VAF
	0.124 ***	[0.093 ; 0.155]	0.062 ***	[0.034 ; 0.090]	0.062 ***	[0.051 ; 0.074]	49.999%
Gambling			**SA → pc1a_Riesgo**		**SA → Gambling → pc1a_Riesgo**		
	Coefficient	BC CI 95%	Coefficient	BC CI 95%	Coefficient	BC CI 95%	VAF
	0.159 ***	[0.134 ; 0.183]	0.151 ***	[0.125 ; 0.175]	0.008 **	[0.005 ; 0.013]	5.342%
Gambling & LRAT			**SA → pc1a_Riesgo**		**SA → Gambling → LRAT → pc1a_Riesgo**		
	Coefficient	BC CI 95%	Coefficient	BC CI 95%	Coefficient	BC CI 95%	VAF
	0.159 ***	[0.134 ; 0.183]	0.151 ***	[0.125 ; 0.175]	0.009 ***	[0.006 ; 0.011]	5.342%

Specifically, the indirect effect of Smartphone Addiction through L-RAT is significant, indicating complementary mediation (VAF ≈ 42%). This suggests that while addiction has a direct "impulsive" impact on risk, a substantial portion of its effect is transmitted through the restructuring of digital routines. A similar pattern emerges for Online Gambling, whose effect on cybervictimization is also largely mediated by L-RAT (VAF ≈ 50%), reinforcing the central role of routine-based opportunity structures. In contrast, the mediation pathway from SA through Gambling remains comparatively limited, indicating that Gambling primarily acts as an amplifier of exposure mechanisms rather than as an independent transmission channel.

4.3 ML: predictive performance and feature importance

Detailed numerical results for the ML model are provided in Appendix B.

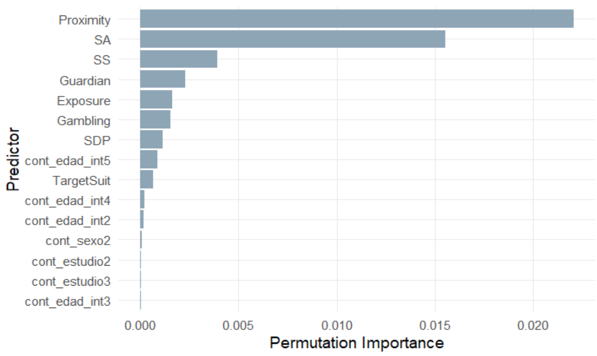
Table 6 reports the main classification metrics. Although overall accuracy is relatively high (0.786), the model exhibits a marked imbalance between sensitivity and specificity at the default probability threshold of 0.5. In particular, sensitivity (0.203) indicates that a substantial proportion of victimized individuals are not correctly identified, whereas specificity (0.974) remains very high. This pattern suggests that the model is considerably more effective in identifying non-victimized cases than in detecting high-risk individuals. As discussed in the thesis, this imbalance should be interpreted in light of the underlying class distribution and the choice of classification threshold. The AUC value (0.769) confirms meaningful discriminative capacity beyond random classification, while the F1-score (0.317) summarizes the trade-off between precision and recall in the presence of potential class imbalance. The confusion matrix (reported in Appendix B) further illustrates this imbalance: while the majority of non-victimized cases are correctly classified, a substantial number of victimized individuals are misclassified as non-victims, reinforcing the asymmetry between sensitivity and specificity.

Table 6: *Classification metrics*

Metric	Estimate
Accuracy	0.786
Sensitivity (Recall)	0.203
Specificity	0.974
Precision (Positive Predictive Value)	0.714
F1-score	0.317
AUC	0.769

The Variable Importance Ranking (Figure 2) highlights that the L-RAT dimensions (specifically Proximity and Exposure) and Smartphone Addiction are the most influential features in reducing classification error. Notably, the Machine Learning results mirror the PLS-SEM findings, confirming that behavioural intensity and situational exposure are the primary determinants of digital vulnerability.

Figure 2: *Variable importance plot*



5 Discussion

5.1 Reinterpreting the research questions

The empirical findings can be meaningfully interpreted through the lens of the four research questions guiding this study.

5.1.1 RQ1 - Direct role of L-RAT patterns

The findings confirm that Lifestyle-Routine Activity patterns constitute the central explanatory mechanism of cybervictimization risk. The positive and statistically significant association between L-RAT and victimization supports H1 and reinforces the applicability of routine activity theory in digital environments. Higher levels of exposure and proximity to motivated offenders, combined with greater target suitability and weaker guardianship, significantly increase vulnerability.

This result is consistent with the theoretical assumption that victimization risk emerges from structured opportunity configurations rather than isolated behavioural traits. The higher-order formative specification of L-RAT further supports its multidimensional nature, demonstrating that digital vulnerability is shaped by the combined effect of distinct situational components. Overall, RQ1 is clearly supported: routine activity mechanisms remain the primary structural driver of cybervictimization risk.

5.1.2 RQ2 - Behavioural drivers of cybervictimization

The structural model confirms that both Smartphone Addiction and Online Gambling behaviours exert significant direct effects on cybervictimization risk. Smartphone Addiction shows a stronger direct effect (H4 supported), indicating that addiction-related dysregulation contributes independently to vulnerability beyond situational exposure mechanisms.

Online Gambling also displays a statistically significant direct association with the output (H2 supported), although with a smaller magnitude. This suggests that gambling does not operate as an autonomous risk determinant but rather as a complementary driver embedded within broader behavioural patterns. These findings indicate that behavioural traits are not merely contextual modifiers but have an intrinsic impact on digital safety.

RQ2 is therefore supported: behavioural drivers, particularly Smartphone Addiction, directly contribute to cybervictimization risk.

5.1.3 RQ3 - Mediating mechanisms linking behaviour to risk

The mediation analysis clarifies how behavioural patterns are translated into victimization risk. Smartphone Addiction significantly increases cybervictimization through L-RAT mechanisms (H5 supported), with approximately 42% of its total effect transmitted via routine exposure structures. This complementary mediation pattern confirms that addiction restructures digital routines, increasing exposure and proximity to offenders.

Online Gambling also indirectly increases victimization risk through L-RAT (H3 supported), reinforcing its role as an exposure amplifier.

Additionally, Smartphone Addiction indirectly increases risk through Gambling (H6 supported), although this pathway is comparatively weaker. Together, these findings indicate that behavioural predispositions and situational opportunity structures interact dynamically.

RQ3 is therefore supported: mediating mechanisms, particularly L-RAT, play a fundamental role in translating behavioural engagement into structured vulnerability.

5.1.4 RQ4 - Psychosocial antecedents of SA

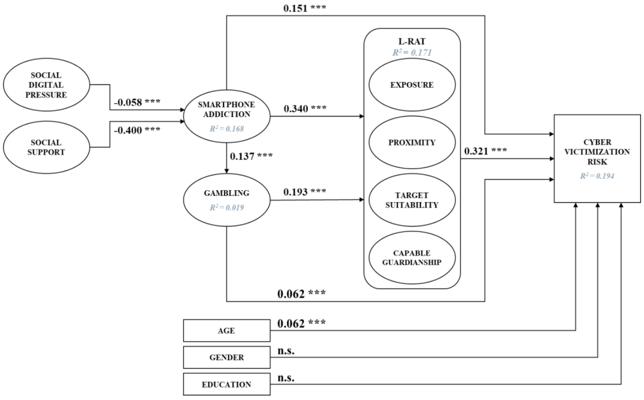
The analysis reveals differentiated roles for psychosocial antecedents. Social Support exhibits a strong negative association with Smartphone Addiction (H8 supported), confirming its protective function. Individuals perceiving higher levels of support demonstrate lower addictive engagement, reducing downstream vulnerability.

In contrast, Social Digital Pressure shows a small but statistically significant negative association with Smartphone Addiction, contrary to the hypothesized positive relationship (H7 not supported). Although literature suggests that digital pressure may intensify problematic use, the empirical evidence indicates that this relationship may be context-dependent. The magnitude of the effect remains limited compared to Social Support, suggesting that protective mechanisms play a more substantial role than digital pressure in shaping addiction within this sample.

RQ4 is partially supported: psychosocial factors significantly influence Smartphone Addiction, but the direction of digital pressure differs from theoretical expectations.

Figure 3 visually summarizes the structural relationships discussed above, highlighting the central role of L-RAT and the mediating pathways

Figure 3: *Structural model results*



linking behavioural patterns to cybervictimization risk.

5.2 Theoretical implications

The results extend Lifestyle-Routine Activity Theory by empirically integrating behavioural addiction mechanisms into its explanatory structure. Rather than treating opportunity structures as fully exogenous, the findings demonstrate that addictive patterns reshape digital routines, thereby increasing exposure and proximity to motivated offenders. Moreover, Smartphone Addiction emerges as a structural antecedent of digital risk rather than solely a psychosocial condition. This perspective bridges criminological opportunity theory and behavioural addiction research, highlighting the interaction between predispositions and situational contexts. Finally, the multidimensional modelling of L-RAT as a second-order formative-formative construct preserves its theoretical ontology and avoids reductionist interpretations of digital vulnerability.

5.3 Methodological implications

The study demonstrates the appropriateness of specifying L-RAT as a higher-order formative-formative construct using a disjoint two-stage approach, ensuring alignment between theoretical structure and measurement specification. Furthermore, the integration of PLS-SEM and Machine Learning provides complementary insights. Structural modelling clarifies causal pathways and mediation mechanisms, while predictive modelling confirms the practical relevance of the identified drivers. The convergence between structural significance and variable importance strengthens the robustness of the empirical evidence.

5.4 Practical and policy implications

The findings suggest that interventions targeting cybervictimization risk should operate at multiple levels. Preventive strategies should not focus exclusively on technical guardianship but also address behavioural addiction mechanisms that intensify exposure. Given the amplifying role of online gambling, regulatory and educational initiatives should incorporate digital safety awareness within gambling environments. Strengthening social support networks may further mitigate addiction-related vulnerability. Overall, effective digital risk prevention requires an integrated approach combining behavioural, situational, and structural interventions.

6 Conclusions

This study examined the structural mechanisms underlying cybervictimization risk by integrating Lifestyle-Routine Activity Theory, Smartphone Addiction, Online Gambling behaviours, and psychosocial antecedents within a unified empirical framework. The results confirm that L-RAT constitutes the central explanatory mechanism of digital vulnerability. Patterns of exposure, proximity to motivated offenders, target suitability, and capable guardianship significantly structure the likelihood of victimization, reinforcing the relevance of routine activity theory in contemporary digital environments.

Smartphone Addiction emerges as both a direct and indirect driver of cybervictimization risk. Beyond its statistically significant direct association with the outcome, approximately 42% of its total effect is transmitted through L-RAT mechanisms, indicating complementary mediation. Addiction-related behavioural dysregulation therefore reshapes digital routines, increasing structured exposure to potential offenders. Online Gambling behaviours also contribute to risk both directly and indirectly through L-RAT, although with comparatively smaller magnitude, functioning primarily as an exposure amplifier within the broader behavioural system.

Psychosocial antecedents display differentiated roles. Social Support confirms its protective function by negatively predicting Smartphone Addiction, thereby indirectly reducing vulnerability. In contrast, Social Digital Pressure exhibits a small but statistically significant negative association with addiction, contrary to the hypothesized direction, suggesting that its influence may be context-dependent within this sample. Finally, the convergence between PLS-SEM and Random Forest analyses strengthens the robustness of the findings, as structurally significant constructs also demonstrate high predictive relevance.

6.1 Theoretical and methodological contributions

From a theoretical perspective, the study advances the application of Lifestyle-Routine Activity Theory by integrating behavioural addiction mechanisms into its explanatory structure. The findings demonstrate that opportunity configurations are not fully exogenous but can be reshaped by individual behavioural predispositions. Smartphone Addiction thus operates as a structural antecedent of digital risk rather than solely as a psychosocial outcome. Moreover, modelling L-RAT as a second-order formative-formative construct preserves its multidimensional ontology and avoids reductionist interpretations of situational exposure.

Methodologically, the use of a disjoint two-stage approach ensures alignment between theoretical specification and measurement modelling. The integration of explanatory (PLS-SEM) and predictive (Random Forest) approaches provides complementary evidence: structural modelling clarifies mediation pathways, while machine learning confirms discriminative capacity. The convergence between theoretical significance and predictive importance enhances the empirical credibility of the framework.

6.2 Limitations

Despite its contributions, the study presents several limitations. First, the cross-sectional design prevents strong causal inference. While structural modelling allows for theoretically grounded directional interpretation, longitudinal or experimental data would be required to establish temporal precedence. Second, all measures are based on self-reported data, which may introduce response bias or social desirability effects. Although measurement validity criteria are satisfied, behavioural log data could provide additional robustness. Third, the dataset reflects a specific national context. Although large and heterogeneous, generalization to other cultural or regulatory environments should be approached with caution. Finally, while the predictive model demonstrates satisfactory overall accuracy, its relatively low sensitivity indicates limited ability to identify high-risk individuals at the default classification threshold. Future research could explore alternative algorithms and cost-sensitive learning strategies.

These limitations do not undermine the core findings but indicate directions for further refinement.

6.3 Future research directions

Future research may extend this framework in several directions. Longitudinal designs would allow examination of the dynamic evolution of Smartphone Addiction, digital routine patterns, and cybervictimization risk over time, enabling stronger causal inference and analysis of potential feedback mechanisms between victimization experiences and behavioural adaptation. Second, the integration of behavioural trace data and objective usage metrics could complement self-reported measures, enhancing measurement precision and reducing common method bias. Third, further research could explore alternative predictive models beyond Random Forest, including gradient boosting or other ensemble techniques, in order to assess whether classification performance, particularly sensitivity in identifying high-risk individuals, can be improved under different modelling specifications. Fourth, multi-group analyses could investigate whether the structural relationships differ between online gamblers and non-gamblers, or across demographic segments, thereby refining the understanding of heterogeneous risk profiles.

Overall, future work should continue to integrate behavioural and situational perspectives to advance a more comprehensive understanding of digital vulnerability.

References

[1] J. Herrero, A. Torres, P. Vivas, A. Hidalgo, F. J. Rodríguez, and A. Urueña, 'Smartphone Addiction and Cybercrime Victimization in the Context of Lifestyles Routine Activities and Self-Control Theories: The User's Dual Vulnerability Model of Cybercrime Victimization', *Int. J. Environ. Res. Public. Health*, vol. 18, no. 7, p. 3763, Apr. 2021, doi: 10.3390/ijerph18073763.

[2] G. Testa *et al.*, 'Cross-jurisdictional factors linked to gambling frequency in adolescents from 28 European countries: a machine learning approach', *Psychiatry Res.*, vol. 351, p. 116602, Sep. 2025, doi: 10.1016/j.psychres.2025.116602.

[3] S. M. Gainsbury, 'Online Gambling Addiction: the Relationship Between Internet Gambling and Disordered Gambling', *Curr. Addict. Rep.*, vol. 2, no. 2, pp. 185–193, Jun. 2015, doi: 10.1007/s40429-015-0057-8.

[4] H. Wardle, G. Reith, E. Langham, and R. D. Rogers, 'Gambling and public health: we need policy action to prevent harm', *BMJ*, vol. 365, p. l1807, May 2019, doi: 10.1136/bmj.l1807.

[5] N. Canale, M. D. Griffiths, A. Vieno, V. Siciliano, and S. Molinaro, 'Impact of Internet gambling on problem gambling among adolescents in Italy: Findings from a large-scale nationally representative survey', *Comput. Hum. Behav.*, vol. 57, pp. 99–106, Apr. 2016, doi: 10.1016/j.chb.2015.12.020.

[6] J.-J. Escario and A. V. Wilkinson, 'Exploring predictors of online gambling in a nationally representative sample of Spanish adolescents', *Comput. Hum. Behav.*, vol. 102, pp. 287–292, Jan. 2020, doi: 10.1016/j.chb.2019.09.002.

[7] Holt and Bossler, *Cybercrime in Progress : Theory and prevention of technology-enabled offenses*. Routledge, 2015. doi: 10.4324/9781315775944.

[8] F. Reer, W. Y. Tang, and T. Quandt, 'Psychosocial well-being and social media engagement: The mediating roles of social comparison orientation and fear of missing out', *New Media Soc.*, vol. 21, no. 7, pp. 1486–1505, Jul. 2019, doi: 10.1177/1461444818823719.

[9] M. Gui and M. Büchi, 'From Use to Overuse: Digital Inequality in the Age of Communication Abundance', *Soc. Sci. Comput. Rev.*, vol. 39, no. 1, pp. 3–19, Feb. 2021, doi: 10.1177/0894439319851163.

[10] S. Cohen and T. A. Wills, 'Stress, social support, and the buffering hypothesis', *Psychol. Bull.*, vol. 98, no. 2, pp. 310–357, 1985, doi: 10.1037/0033-2909.98.2.310.

[11] P. A. Busch and S. McCarthy, 'Antecedents and consequences of problematic smartphone use: A systematic literature review of an emerging research area', *Comput. Hum. Behav.*, vol. 114, p. 106414, Jan. 2021, doi: 10.1016/j.chb.2020.106414.

[12] Z. Zhou and Q. Cheng, 'Effects of online and offline social support on problematic smartphone use among Chinese adolescents: mediating role of needs frustration', *Behav. Inf.*

Technol., pp. 1–13, Nov. 2025, doi: 10.1080/0144929X.2025.2585328.

[13] H. D. Erdem, J. Herrero, and A. Urueña, 'The Bidirectional Relationships between Social Pressure in Digital Contexts, Depression, and Social Support over Time', *Psychosoc. Interv.*, vol. 34, no. 3, pp. 189–200, 2025, doi: 10.5093/pi2025a15.

[14] J. Herrero, A. Torres, P. Vivas, Á. E. Arenas, and A. Urueña, 'Examining the Empirical Links Between Digital Social Pressure, Personality, Psychological Distress, Social Support, Users' Residential Living Conditions, and Smartphone Addiction', *Soc. Sci. Comput. Rev.*, vol. 40, no. 5, pp. 1153–1170, Oct. 2022, doi: 10.1177/0894439321998357.

[15] L. E. Cohen and M. Felson, 'Social Change and Crime Rate Trends: A Routine Activity Approach', *Am. Sociol. Rev.*, vol. 44, no. 4, pp. 588–608, 1979, doi: 10.2307/2094589.

[16] M. J. Hindelang, M. R. Gottfredson, and J. Garofalo, *Victims of personal crime: an empirical foundation for a theory of personal victimization*. Cambridge, Mass. : Ballinger Pub. Co., 1978. Accessed: Jan. 04, 2026. [Online]. Available: http://archive.org/details/victimsofpersona0000hind

[17] A. Diamantopoulos, P. Riefler, and K. P. Roth, 'Advancing formative measurement models', *J. Bus. Res.*, vol. 61, no. 12, pp. 1203–1218, Dec. 2008, doi: 10.1016/j.jbusres.2008.01.009.

[18] J. F. Hair, G. T. M. Hult, C. M. Ringle, M. Sarstedt, N. P. Danks, and S. Ray, *Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R: A Workbook*. in Classroom Companion: Business. Cham: Springer International Publishing, 2021. doi: 10.1007/978-3-030-80519-7.

[19] J. F. Hair, Ed., *A primer on partial least squares structural equation modeling (PLS-SEM)*. Los Angeles, Calif.: Sage, 2014.

[20] M. Tenenhaus, V. E. Vinzi, Y.-M. Chatelin, and C. Lauro, 'PLS path modeling', *Comput. Stat. Data Anal.*, vol. 48, no. 1, pp. 159–205, Jan. 2005, doi: 10.1016/j.csda.2004.03.005.

[21] C. B. Jarvis, S. B. MacKenzie, and P. M. Podsakoff, 'A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research', *J. Consum. Res.*, vol. 30, no. 2, pp. 199–218, Sep. 2003, doi: 10.1086/376806.

[22] M. Sarstedt, J. F. Hair, J.-H. Cheah, J.-M. Becker, and C. M. Ringle, 'How to Specify, Estimate, and Validate Higher-Order Constructs in PLS-SEM', *Australas. Mark. J.*, vol. 27, no. 3, pp. 197–211, Aug. 2019, doi: 10.1016/j.ausmj.2019.05.003.

[23] X. Zhao, J. G. Lynch Jr., and Q. Chen, 'Reconsidering Baron and Kenny: Myths and truths about mediation analysis', *J. Consum. Res.*, vol. 37, no. 2, pp. 197–206, 2010, doi: 10.1086/651257.

[24] L. Breiman, 'Random Forests', *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[25] B. Boehmke and B. Greenwell, *Hands-On Machine Learning with R*, 1st edn. Chapman and Hall/CRC, 2019. doi: 10.1201/9780367816377.

[26] G. Shmueli *et al.*, 'Predictive model assessment in PLS-SEM: guidelines for using PLSpredict', *Eur. J. Mark.*, vol. 53, no. 11, pp. 2322–2347, Jun. 2019, doi: 10.1108/EJM-02-2019-0189.

A. Appendix A - PLS-SEM results

A.1 Measurement model results

Table A.1: *Reflective measurement model results*

	Total Sample (N = 7,064)			
Construct /indicator	Outer loadings (Alpha)	rhoA	rhoC	AVE
SA	**(0.905)**	**0.912**	**0.924**	**0.604**
i6aan_4	0.793			
i6aan_5	0.792			
i6aan_9	0.770			
i6aan_10	0.783			
i6ban_2	0.628			
i6ban_6	0.836			
i6ban_8	0.764			
i6ban_9	0.829			
SDP	**(0.850)**	**0.921**	**0.906**	**0.764**
i14pcan_1	0.879			
i14pcan_2	0.915			
i14pcan_3	0.823			
SS	**(0.673)**	**0.700**	**0.820**	**0.605**
i9pcan_1	0.784			
i9pcan_2	0.680			
i9pcan_3	0.859			

Table A.2: *Reflective measurement model results: discriminant validity*

Total Sample (N = 7,064)

Fornell-Larcker criterion*				HTMT**			
	SA	**SDP**	**SS**		**SA**	**SDP**	**SS**
SA	**0.777**			**SA**			
SDP	-0.092	**0.874**		**SDP**	**0.098** [0.073 ; 0.125]		
SS	-0.405	0.083	**0.778**	**SS**	**0.508** [0.479 ; 0.536]	**0.103** [0.072 ; 0.134]	

* the **emboldened** numbers on the diagonal represent the square root of the AVE for each construct; the remaining numbers are the correlations between the reflective constructs from the latent scores.

** In parentheses, the 95% CI.

Table A.3: *Formative measurement model results*

	Total Sample (N = 7,064)				
Dimension / indicator	Outer VIF	Weight	Weight's BC CI 95%	Loading	Loading's BC CI 95%
Exposure	**1.038**	**0.359 *****	**[0.309 ; 0.407]**		
a1_pcan_6	1.06	0.286 ***	[0.214 ; 0.357]		
a1_pcan_7	1.06	-0.146 ***	[-0.227 ; -0.068]		
a1_pcan_8	1.10	0.670 ***	[0.604 ; 0.733]		
a1_pcan_13	1.11	0.524 ***	[0.450 ; 0.596]		
Guardian	**1.045**	**-0.359 *****	**[-0.403 ; -0.315]**		
g3_pcan_1	1.36	-0.156 ***	[-0.246 ; -0.065]		
g3_pcan_3	1.30	0.527 ***	[0.442 ; 0.610]		
g3_pcan_5	1.36	0.156 ***	[0.067 ; 0.245]		
g3_pcan_7	1.31	0.325 ***	[0.236 ; 0.410]		
g3_pcan_9	1.35	0.467 ***	[0.395 ; 0.538]		
Proximity	**1.032**	**0.606 *****	**[0.563 ; 0.649]**		
d1_pcan_1	1.26	0.257 ***	[0.186 ; 0.327]		
d1_pcan_2	1.22	0.020 n.s.	[-0.048 ; 0.087]	0.342 ***	[0.283 ; 0.400]

d1_pcan_3	1.21	-0.113 ***	[-0.179 ; -0.046]		
d1_pcan_4	1.27	0.327 ***	[0.257 ; 0.397]		
d1_pcan_5	1.21	0.198 ***	[0.126 ; 0.270]		
d1_pcan_6	1.27	0.472 ***	[0.397 ; 0.542]		
d1_pcan_7	1.28	0.462 ***	[0.395 ; 0.527]		
TargetSuit	**1.047**	**0.393 *****	**[0.349 ; 0.437]**		
g2_pcan_1	2.37	0.344 ***	[0.234 ; 0.455]		
g2_pcan_2	2.33	0.310 ***	[0.193 ; 0.421]		
g2_pcan_3	1.86	-0.128 n.s.	[-0.271 ; 0.014]	0.151 ***	[0.062 ; 0.240]
g2_pcan_4	1.89	-0.357 ***	[-0.492 ; -0.217]		
g2_pcan_5	2.36	0.727 ***	[0.644 ; 0.807]		

Note: The following convention values have been adopted for significance levels:

- t-values below 1.96 are considered not significant at the 5% level ($p > 0.05$)
- values between 1.96 and 2.58 are marked with * ($p < 0.05$)
- values between 2.58 and 3.29 are marked with ** ($p < 0.01$)
- values above 3.29 are marked with *** ($p < 0.001$)

A.2 Structural model results

Table A.4: Structural Model results, first-order model

Endogenous construct	Relationship			Total Sample (N = 7,064) Inner VIF	Path coef β	BC CI 95%
Smartphone Addiction	SS	→	SA	1.007	-0.400	[-0.422 ; -0.377]
	SDP	→	SA	1.007	-0.058	[-0.079 ; -0.037]
Gambling	SA	→	Gambling	*NA*	0.137	[0.112 ; 0.162]
Exposure	SA	→	Exposure	1.019	0.102	[0.076 ; 0.126]
	Gambling	→	Exposure	1.019	0.243	[0.214 ; 0.269]
Proximity	SA	→	Proximity	1.019	0.180	[0.156 ; 0.203]
	Gambling	→	Proximity	1.019	0.112	[0.083 ; 0.140]
TargetSuit	SA	→	TargetSuit	1.019	0.252	[0.226 ; 0.274]
	Gambling	→	TargetSuit	1.019	0.076	[0.050 ; 0.101]
Guardian	SA	→	Guardian	1.019	-0.266	[-0.288 ; -0.242]
	Gambling	→	Guardian	1.019	-0.022	[-0.045 ; 0.001]
Target Variable pc1a_Riesgo	SA	→	pc1a_Riesgo	1.296	0.168	[0.144 ; 0.193]
	Gambling	→	pc1a_Riesgo	1.110	0.065	[0.038 ; 0.094]
	Exposure	→	pc1a_Riesgo	1.130	0.087	[0.061 ; 0.113]
	Proximity	→	pc1a_Riesgo	1.084	0.292	[0.262 ; 0.319]
	TargetSuit	→	pc1a_Riesgo	1.100	0.050	[0.026 ; 0.072]
	Guardian	→	pc1a_Riesgo	1.096	-0.067	[-0.088 ; -0.046]
	Age	→	pc1a_Riesgo	1.136	-0.063	[-0.087 ; -0.040]
	Gender	→	pc1a_Riesgo	1.062	-0.011	[-0.033 ; 0.010]
	Education	→	pc1a_Riesgo	1.008	-0.006	[-0.027 ; 0.015]

Table A.5: Structural Model results, second-order model

Endogenous construct	Relationship			Total Sample (N = 7,064) Inner VIF	Path coef β	BC CI 95%
Smartphone Addiction	SS	→	SA	1.007	-0.400	[-0.422 ; -0.378]
	SDP	→	SA	1.007	-0.058	[-0.079 ; -0.037]
Gambling	SA	→	Gambling	*NA*	0.137	[0.112 ; 0.162]
LRAT	SA	→	LRAT	1.019	0.340	[0.315 ; 0.362]
	Gambling	→	LRAT	1.019	0.193	[0.164 ; 0.221]
Target Variable pc1a_Riesgo	SA	→	pc1a_Riesgo	1.263	0.151	[0.126 ; 0.175]
	Gambling	→	pc1a_Riesgo	1.084	0.062	[0.034 ; 0.091]
	LRAT	→	pc1a_Riesgo	1.215	0.321	[0.291 ; 0.351]
	Age	→	pc1a_Riesgo	1.127	-0.062	[-0.086 ; -0.038]
	Gender	→	pc1a_Riesgo	1.048	-0.009	[-0.031 ; 0.013]
	Education	→	pc1a_Riesgo	1.004	-0.002	[-0.023 ; 0.019]

***Table A.6**: Explanatory power (R^2) results, first-order model*

Construct	Total sample (N = 7,064) R^2 value	BC CI 95%
SA	0.168	[0.158 ; 0.176]
Gambling	0.019	[0.015 ; 0.022]
Exposure	0.076	[0.069 ; 0.083]
Proximity	0.050	[0.045 ; 0.056]
TargetSuit	0.074	[0.068 ; 0.081]
Guardian	0.073	[0.066 ; 0.078]
Target Variable pc1a_Riesgo	0.212	[0.200 ; 0.222]

***Table A.7**: Explanatory power (R^2) results, second-order model*

Construct	Total sample (N = 7,064) R^2 value	BC CI 95%
SA	0.168	[0.159 ; 0.176]
Gambling	0.019	[0.015 ; 0.022]
LRAT	0.171	[0.161 ; 0.179]
Target Variable pc1a_Riesgo	0.194	[0.183 ; 0.205]

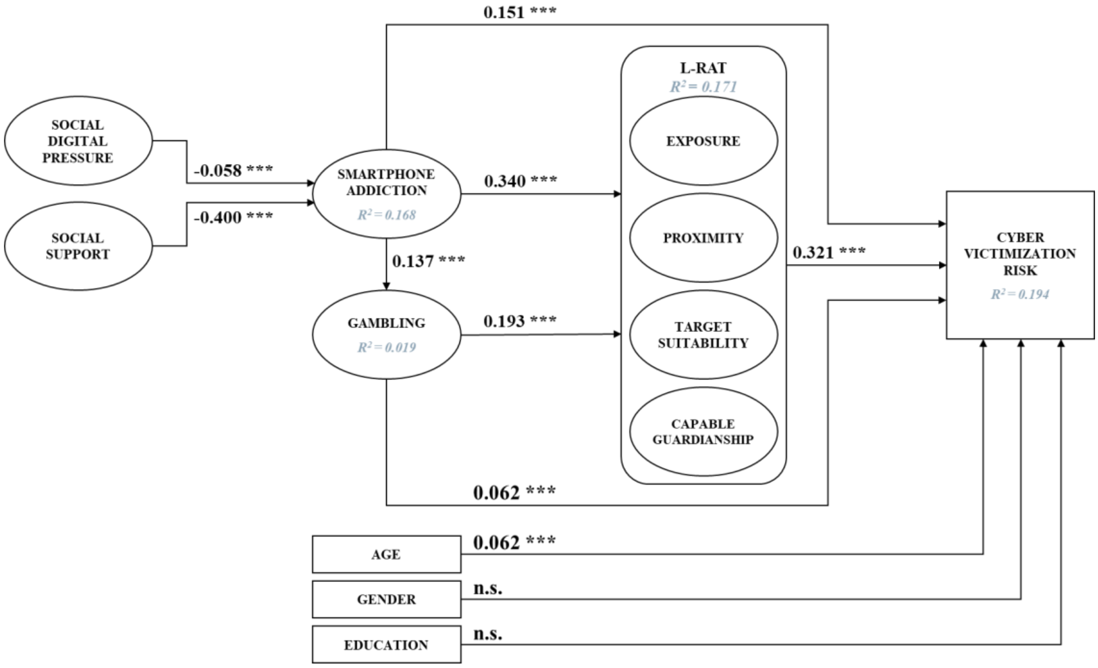
***Table A.8:** Effect size (f^2) results, first-order model*

Endogenous construct	Relationship			Total Sample (N = 7,064) R^2 included	R^2 excluded	f^2 value	Effect size
Smartphone Addiction	SS	→	SA	0.168	0.005	1.950e-01	medium
	SDP	→	SA	0.168	0.162	6.404e-03	negligible
Gambling	SA	→	Gambling	0.019	0.000	1.917e-02	negligible
Exposure	SA	→	Exposure	0.076	0.062	1.488e-02	negligible
	Gambling	→	Exposure	0.076	0.014	6.752e-02	small
Proximity	SA	→	Proximity	0.050	0.015	3.703e-02	small
	Gambling	→	Proximity	0.050	0.035	1.606e-02	negligible
TargetSuit	SA	→	TargetSuit	0.074	0.008	7.116e-02	small
	Gambling	→	TargetSuit	0.074	0.066	9.058e-03	negligible
Guardian	SA	→	Guardian	0.073	0.001	7.694e-02	small
	Gambling	→	Guardian	0.073	0.071	1.356e-03	negligible
Target Variable pc1a_Riesgo	SA	→	pc1a_Riesgo	0.212	0.162	6.303e-02	small
	Gambling	→	pc1a_Riesgo	0.212	0.201	1.346e-02	negligible
	Exposure	→	pc1a_Riesgo	0.212	0.195	2.194e-02	small
	Proximity	→	pc1a_Riesgo	0.212	0.106	1.349e-01	small
	TargetSuit	→	pc1a_Riesgo	0.212	0.205	8.282e-03	negligible
	Guardian	→	pc1a_Riesgo	0.212	0.201	1.377e-02	negligible
	Age	→	pc1a_Riesgo	0.212	0.201	1.328e-02	negligible
	Gender	→	pc1a_Riesgo	0.212	0.212	9.980e-05	negligible
	Education	→	pc1a_Riesgo	0.212	0.212	-9.284e-05	negligible

***Table A.9:** Effect size (f^2) results, second-order model*

Endogenous construct	Relationship			Total Sample (N = 7,064) R^2 included	R^2 excluded	f^2 value	Effect size
Smartphone Addiction	SS	→	SA	0.168	0.005	1.950e-01	medium
	SDP	→	SA	0.168	0.162	6.390e-03	negligible
Gambling	SA	→	Gambling	0.019	0.000	1.914e-02	negligible
LRAT	SA	→	LRAT	0.171	0.046	1.502e-01	medium
	Gambling	→	LRAT	0.171	0.125	5.567e-02	small
Target Variable pc1a_Riesgo	SA	→	pc1a_Riesgo	0.194	0.150	5.514e-02	small
	Gambling	→	pc1a_Riesgo	0.194	0.184	1.261e-02	negligible
	LRAT	→	pc1a_Riesgo	0.194	0.065	1.604e-01	medium
	Age	→	pc1a_Riesgo	0.194	0.184	1.270e-02	negligible
	Gender	→	pc1a_Riesgo	0.194	0.194	8.282e-05	negligible
	Education	→	pc1a_Riesgo	0.194	0.194	-3.548e-05	negligible

Figure A.1: *Structural model results*



A.3 Mediation analysis results

Table A.10: *Mediation analysis results*

	Total Sample (N = 7,064)						
Mediator	Total effect		Direct effect		Indirect effect		
LRAT			**SA → pc1a_Riesgo**		**SA → LRAT → pc1a_Riesgo**		
	Coefficient	BC CI 95%	Coefficient	BC CI 95%	Coefficient	BC CI 95%	VAF
	0.260 ***	[0.235 ; 0.284]	0.151 ***	[0.125 ; 0.175]	0.109 ***	[0.097 ; 0.122]	42.062%
			Gambling → pc1a_Riesgo		**Gambling → LRAT → pc1a_Riesgo**		
	Coefficient	BC CI 95%	Coefficient	BC CI 95%	Coefficient	BC CI 95%	VAF
	0.124 ***	[0.093 ; 0.155]	0.062 ***	[0.034 ; 0.090]	0.062 ***	[0.051 ; 0.074]	49.999%
Gambling			**SA → pc1a_Riesgo**		**SA → Gambling → pc1a_Riesgo**		
	Coefficient	BC CI 95%	Coefficient	BC CI 95%	Coefficient	BC CI 95%	VAF
	0.159 ***	[0.134 ; 0.183]	0.151 ***	[0.125 ; 0.175]	0.008 **	[0.005 ; 0.013]	5.342%
Gambling & LRAT			**SA → pc1a_Riesgo**		**SA → Gambling → LRAT → pc1a_Riesgo**		
	Coefficient	BC CI 95%	Coefficient	BC CI 95%	Coefficient	BC CI 95%	VAF
	0.159 ***	[0.134 ; 0.183]	0.151 ***	[0.125 ; 0.175]	0.009 ***	[0.006 ; 0.011]	5.342%

Note: The following convention values have been adopted for significance levels:

- t-values below 1.96 are considered not significant at the 5% level ($p > 0.05$)
- values between 1.96 and 2.58 are marked with * ($p < 0.05$)
- values between 2.58 and 3.29 are marked with ** ($p < 0.01$)
- values above 3.29 are marked with *** ($p < 0.001$)

B. Appendix B - ML results

B.1 Model specification and hyperparameter tuning

Table B.1: *Optimal Random Forest hyperparameters*

Hyperparameter	Value	Selection method	Description
n_trees	1844	Tuned (CV grid search)	Number of trees in the ensemble
mtry	5	Tuned (CV grid search)	Predictors randomly sampled at each split
max_depth	6	Tuned (CV grid search)	Maximum tree depth
min_n	25	Tuned (CV grid search)	Minimum observations in a terminal node
sample.fraction	0.632	Fixed (Breiman, 2001)	Bootstrap sampling proportion
replace	TRUE	Fixed (Breiman, 2001)	Sampling with replacement

B.2 Model performance evaluation

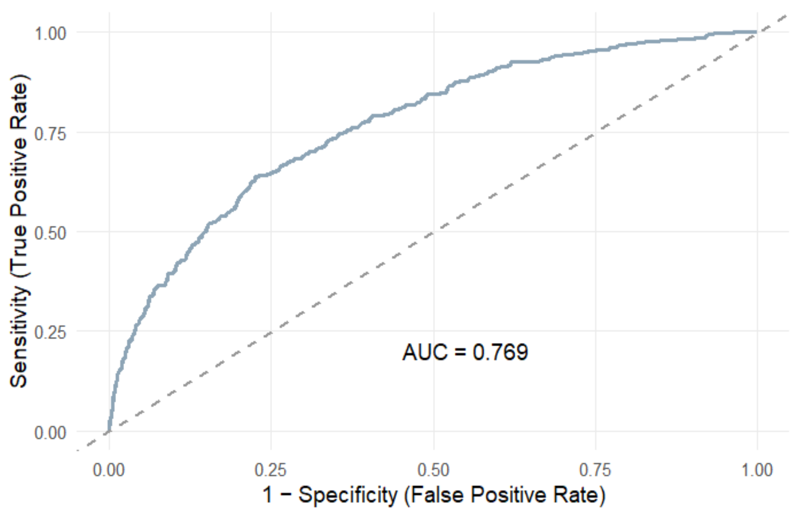
Table B.2: *Confusion matrix*

		Predicted	
		NonVictim	Victim
Actual	NonVictim	1041	28
	Victim	274	70

Table B.3: *Classification metrics*

Metric	Estimator	Estimate
Accuracy	binary	0.786
Sensitivity (Recall)	binary	0.203
Specificity	binary	0.974
Precision (Positive Predictive Value)	binary	0.714
F1-score	binary	0.317
AUC	binary	0.769

Figure B.1: *ROC curve*

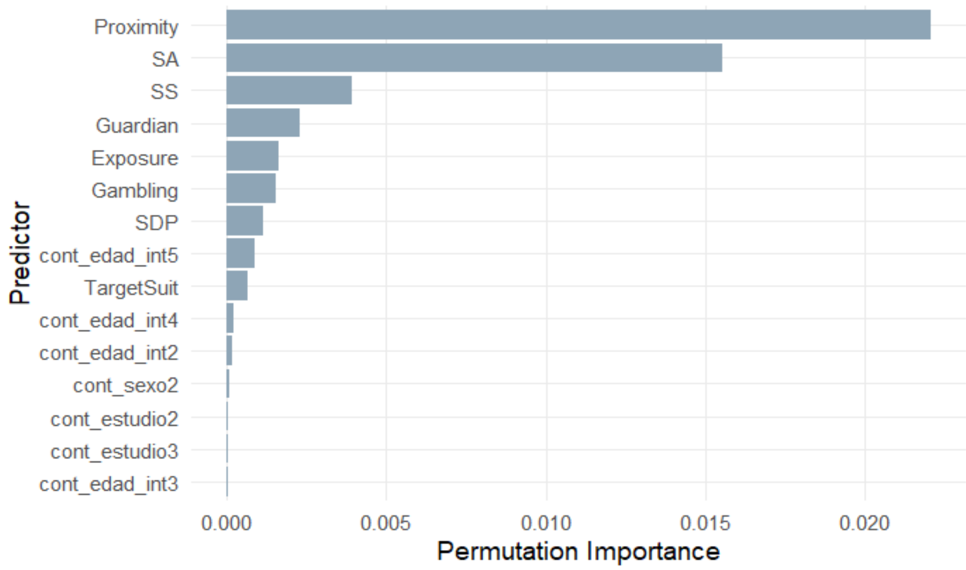


B.3 Variable importance analysis

Table B.4: *Variable importance*

Rank	Variable	Importance
1	Proximity	2.207e-02
2	SA	1.552e-02
3	SS	3.925e-03
4	Guardian	2.3027e-03
5	Exposure	1.644e-03
6	Gambling	1.567e-03
7	SDP	1.162e-03
8	cont_edad_int5	8.837e-04
9	TargetSuit	6.643e-04
10	cont_edad_int4	2.205e-04
11	cont_edad_int2	1.677e-04
12	cont_sexo2	8.040e-05
13	cont_estudio2	6.639e-05
14	cont_estudio3	6.159e-05
15	cont_edad_int3	3.371e-05

Figure B.2: *Variable importance plot*



C. Appendix C - R Code

```
"TFM Trevisan Federico
Date: 2026-02-03
Title: Understanding cybervictimization through smartphone addiction, online
gambling and routine activities: an integrated approach using PLS-SEM and ML"

# ==============================================================================
# 1. CLEANING, DIRECTORY AND LIBRARIES

rm(list = ls())

current_path <- dirname(rstudioapi::getActiveDocumentContext()$path)
setwd(current_path)

library(haven)
library(tidyverse)
library(dplyr)
library(psych)
library(purrr)
library(tibble)
library(seminr)
library(boot)
library(tidymodels)
library(yardstick)
library(car)
library(readxl)
library(caret)
library(vip)
library(ranger)

set.seed(123)

# ==============================================================================
# 2. DATA IMPORT

# Importing "df_agg", starting database
load("dati_grande.RData")

# Convert "haven_labelled" in "numeric"
df_agg <- df_agg %>%
  mutate(across(where(~ inherits(.x, "haven_labelled")), as.numeric))

# RECODING SCALES
# Target Suitability (g2_pcan_1:5) -> inverted: ↑ = ↑Suitability
df_agg <- df_agg %>%
  mutate(across(starts_with("g2_pcan_"), ~ 6 - .x))

# Social Digital Pressure (i14pcan_1:3) -> inverted: ↑ = ↑SDP
df_agg <- df_agg %>%
  mutate(across(starts_with("i14pcan_"), ~ 6 - .x))

# Social Support (i9pcan_1:3) -> inverted: ↑ = ↑SS
df_agg <- df_agg %>%
  mutate(across(starts_with("i9pcan_"), ~ 6 - .x))

# ==============================================================================
# 3. DATABASE - PLS-SEM

# First-order Formative Constructs (FOC L-RAT)
exposure_vars  <- paste0("a1_pcan_", c(6,7,8,13))
proximity_vars <- paste0("d1_pcan_", 1:7)
target_vars    <- paste0("g2_pcan_", 1:5)
guardian_vars  <- paste0("g3_pcan_", c(1,3,5,7,9))

# Smartphone Addiction (reflective)
add_vars <- c("i6aan_4", "i6aan_5", "i6aan_9", "i6aan_10",
              "i6ban_2", "i6ban_6", "i6ban_8", "i6ban_9")

# Social Digital Pressure (reflective)
sdp_vars <- c("i14pcan_1", "i14pcan_2", "i14pcan_3")

# 3.4 Social Support (reflective)
ss_vars  <- c("i9pcan_1", "i9pcan_2", "i9pcan_3")
```

```r
# Gambling
gambling <- "jugador"

# Dependent variable pc1a_Riesgo (0-6 continuous)
df_agg$pc1a_Riesgo <- rowSums(df_agg[, paste0("pc1a_pcan_", c(1,2,4,5,6,7))],
                              na.rm = TRUE)
output <- "pc1a_Riesgo"

# DATABASE - PLS-SEM (ex: TFM_mod)
DB_PLSSEM <- df_agg %>%
  select(all_of(c(exposure_vars, proximity_vars, target_vars, guardian_vars,
                  add_vars, sdp_vars, ss_vars, gambling, output,
                  "cont_sexo", "cont_edad_int", "cont_estudio"))) %>%
  drop_na()

# ============================================================================
# 4. DATABASE - ML

# POMP SUM for FORMATIVE CONSTRUCTS
pomp_sum <- function(df, items, min_item, max_item){
  k         <- length(items)
  raw_sum   <- rowSums(df[,items])
  min_total <- k*min_item
  max_total <- k*max_item
  ((raw_sum-min_total)/(max_total-min_total)) * 100}

# POMP MEAN for REFLECTIVE CONSTRUCTS
pomp_mean <- function(df, items, min_item, max_item){
  m <- rowMeans(df[, items])
  ((m - min_item)/(max_item - min_item))*100}

DB_ML_PROV <- DB_PLSSEM %>%
  mutate(
    # Formative constructs
    Exposure    = pomp_sum(., exposure_vars,  min_item = 0, max_item = 1),
    Proximity   = pomp_sum(., proximity_vars, min_item = 0, max_item = 1),
    TargetSuit  = pomp_sum(., target_vars,    min_item = 1, max_item = 5),
    Guardian    = pomp_sum(., guardian_vars,  min_item = 1, max_item = 5),

    # Reflective constructs
    SA          = pomp_mean(., add_vars,      min_item = 1, max_item = 5),
    SDP         = pomp_mean(., sdp_vars,      min_item = 1, max_item = 5),
    SS          = pomp_mean(., ss_vars,       min_item = 1, max_item = 5),

    # Target rescaled to 0-100
    pc1a_Riesgo = round((pc1a_Riesgo / 6) * 100, 2),

    # Gambling rescaled to 0-100 (0=non-gambler, 100=gambler)
    Gambling = as.numeric(jugador) * 100
  ) %>%
  transmute(
    Exposure, Proximity, TargetSuit, Guardian, SA, SDP, SS, Gambling,
    cont_sexo, cont_edad_int, cont_estudio, pc1a_Riesgo
  )

# Ensure categorical controls are factors (for one-hot encoding)
DB_ML_PROV <- DB_ML_PROV %>%
  mutate(
    cont_sexo     = factor(cont_sexo),
    cont_estudio  = factor(cont_estudio),
    cont_edad_int = factor(cont_edad_int)
  )

# One-hot encoding ONLY for sociodemographic controls
dummies <- model.matrix(
  ~ cont_sexo + cont_estudio + cont_edad_int,
  data = DB_ML_PROV
)[, -1]
dummies <- as_tibble(dummies)

# Final ML dataset: scaled numeric predictors + target + dummies
DB_ML <- bind_cols(
  DB_ML_PROV %>%
    select(Exposure, Proximity, TargetSuit, Guardian, SA, SDP, SS, Gambling, pc1a_Riesgo) %>%
    mutate(across(where(is.numeric), ~ round(.x, 2))),
```

```
  dummies
)

# ==============================================================================
# 5. PLS-SEM: FIRST-ORDER MODEL

# 5.1 - MEASUREMENT MODEL (first-order model)
mod_mm_base <- constructs(
  # FORMATIVES
  composite("Exposure",   exposure_vars,  weights = mode_B),
  composite("Proximity",  proximity_vars, weights = mode_B),
  composite("TargetSuit", target_vars,    weights = mode_B),
  composite("Guardian",   guardian_vars,  weights = mode_B),
  # REFLECTIVES
  composite("SA",  add_vars, weights = mode_A),
  composite("SDP", sdp_vars, weights = mode_A),
  composite("SS",  ss_vars,  weights = mode_A),
  # SINGLE-ITEM
  composite("Gambling",  single_item("jugador")),
  composite("Age",       single_item("cont_edad_int")),
  composite("Gender",    single_item("cont_sexo")),
  composite("Education", single_item("cont_estudio")),
  composite("Output",    single_item("pc1a_Riesgo"))
)

# 5.2 - STRUCTURAL MODEL (first-order model)
mod_sm_base <- relationships(
  paths(from = c("SS","SDP"), to = "SA"),
  paths(from = "SA", to = "Gambling"),
  paths(from = c("SA","Gambling"),
        to = c("Exposure","Proximity","TargetSuit","Guardian")),
  paths(from = c("SA","Gambling","Exposure","Proximity","TargetSuit","Guardian",
                 "Age","Gender","Education"),
        to = "Output")
)

# 5.3 - ESTIMATION OF PLS-SEM (first-order model)
mod_base <- estimate_pls(data = DB_PLSSEM,
                         measurement_model = mod_mm_base,
                         structural_model  = mod_sm_base)

# 5.4 - Bootstrap of the model to obtain significance levels (first-order model)
boot_base         <- bootstrap_model(mod_base, nboot = 10000)
summary(boot_base)

# 5.5 - Reliability measures for reflexive constructs (first-order model)
reliability_base <- summary(mod_base)$reliability
reliability_base

# 5.6 - Fornell-Larcker criterion (√AVE > inter-construct correlations)
cat("\nbase - sqrt(AVE):\n")
print(sqrt(reliability_base[c("SA","SDP","SS"), "AVE"]))

# Correlations between reflective constructs from latent scores
FL_lv_scores <- as.data.frame(mod_base$construct_scores)
cat("\nCorrelations between reflective constructs (SA, SDP, SS):\n")
print(round(cor(FL_lv_scores[, c("SA","SDP","SS")]), 3))

# 5.7 - Bias-corrected bootstrap Confidence interval (first-order model)
# (following indications of Streukens and Leroi-Werelds, 2016)
bc_excel_ci <- function(t_star, t0, conf = 0.95,
                        index_round = c("round","ceiling_floor")){
  # --- Step 1: select t0, t_star ---
  index_round <- match.arg(index_round)
  t_star      <- as.numeric(t_star)
  t_star      <- t_star[is.finite(t_star)]
  R           <- length(t_star)

  # --- Step 2: acceleration coefficient a (MD^3 / MD^2) ---
  m       <- mean(t_star)
  md3_sum <- sum((t_star - m)^3)
  md2_sum <- sum((t_star - m)^2)
  denom   <- 6 * (md2_sum^(3/2))
  a       <- if (denom == 0) 0 else (md3_sum / denom)
  # intermediate denominator D5 = 6 * ( (sum MD^2)^(3/2) )
```

```
  # --- Step 3.1: z0 from bootstrap proportion ---
  p_less  <- mean(t_star < t0)
  eps     <- 0.5 / R
  p_less  <- min(max(p_less, eps), 1 - eps)
  z0      <- qnorm(p_less)
  alpha   <- (1 - conf) / 2
  zL      <- qnorm(alpha)
  zU      <- qnorm(1 - alpha)
  # Z-lower e Z-upper (95% -> +/-1.96; 90% -> +/-1.645)

  # --- Step 3.2: z' (prime) ---
  zL_prime <- z0 + (z0 + zL) / (1 - a * (z0 + zL))
  zU_prime <- z0 + (z0 + zU) / (1 - a * (z0 + zU))

  # --- Step 3.3: p-values associated to z' ---
  pL_prime <- pnorm(zL_prime)
  pU_prime <- pnorm(zU_prime)

  # --- Step 4.1: observation numbers = p' * R ---
  kL_raw   <- pL_prime * R
  kU_raw   <- pU_prime * R
  if (index_round == "round") {
    kL <- as.integer(round(kL_raw))
    kU <- as.integer(round(kU_raw))
  }else{
    kL <- as.integer(ceiling(kL_raw))
    kU <- as.integer(floor(kU_raw))
  }
  # conservative choice: lower = ceiling, upper = floor
  kL <- max(1L, min(R, kL))
  kU <- max(1L, min(R, kU))  # clamp among 1 and R

  # --- Step 4.2: bounds through statistical order (SMALL) ---
  s     <- sort(t_star)
  lower <- s[kL]
  upper <- s[kU]
  list(lower = lower, upper = upper, t0 = t0, mean_boot = m, R = R, a = a,
       z0 = z0, p_less = p_less, zL = zL, zU = zU, zL_prime = zL_prime,
       zU_prime = zU_prime, pL_prime = pL_prime, pU_prime = pU_prime,
       kL_raw = kL_raw, kU_raw = kU_raw, kL = kL, kU = kU)
}

bc_excel_ci_all_paths <- function(boot_obj, conf = 0.95, only_nonzero = TRUE,
                                  index_round = c("round", "ceiling_floor")){
  B0    <- boot_obj$path_coef
  BP    <- boot_obj$boot_paths
  preds <- rownames(B0)
  targs <- colnames(B0)  # automatic path selection
  idx   <- which(!is.na(B0), arr.ind = TRUE)
  if (only_nonzero) idx <- idx[B0[idx] != 0, , drop = FALSE]
  R     <- dim(BP)[3]
  out   <- vector("list", nrow(idx))
  for (k in seq_len(nrow(idx))) {
    i      <- idx[k, 1]
    j      <- idx[k, 2]
    name   <- paste0(preds[i], " -> ", targs[j])
    t0     <- B0[i, j]
    t_star <- BP[i, j, ]
    ci     <- bc_excel_ci(t_star = t_star,t0 = t0, conf = conf,
                      index_round = match.arg(index_round))
    out[[k]] <- data.frame(path = name, beta = ci$t0, bc_l = ci$lower,
                           bc_u = ci$upper, a = ci$a, z0 = ci$z0, kL = ci$kL,
                           kU = ci$kU, R = ci$R, row.names = NULL)
  }
  do.call(rbind, out)
}

tab_base <- bc_excel_ci_all_paths(boot_base, conf = 0.95, only_nonzero = TRUE,
                                  index_round = "ceiling_floor")
tab_base

# 5.8 - R2 bias-corrected bootstrap CI (first-order model)
bootstrap_R2_from_paths_and_lvcor <- function(boot_obj, target, use_abs = TRUE){
  B0     <- boot_obj$path_coef # Beta original matrix (predictor x target)
  if (!(target %in% colnames(B0))) stop("target not a column in path_coef.")
  betas0 <- B0[, target]
```

```
  preds  <- names(betas0)[!is.na(betas0) & betas0 != 0]
  # predictors as incoming path != 0
  if (length(preds) == 0) stop("No incoming predictor for this target variable")
  C0     <- cor(boot_obj$construct_scores, use = "pairwise.complete.obs")
  # LV correlations from original model
  r_xy   <- C0[preds, target] # correlations pred->target vector
  if (use_abs) r_xy <- abs(r_xy)
  BP     <- boot_obj$boot_paths
  R      <- dim(BP)[3]
  # bootstrap path coefficients (array pred x target x R)
  # for each pred i, estract beta bootstrap towards the target: matrix R x nPred
  beta_star <- sapply(preds, function(p) BP[p, target, ])
  beta_star <- as.matrix(beta_star)
  if (use_abs) beta_star <- abs(beta_star)
  R2_star   <- as.numeric(beta_star %*% r_xy) # Eq.6: line product sum (replica)
  # original estimate, coherent with the formula
  beta0_vec <- betas0[preds]
  if (use_abs) beta0_vec <- abs(beta0_vec);
  R2_0      <- sum(beta0_vec * r_xy, na.rm = TRUE)
  list(target= target, preds= preds, lv_cor= r_xy, R2_0= R2_0, R2_star= R2_star)
}

all_endogenous <- function(B) colnames(B)[colSums(abs(B), na.rm = TRUE) > 0]

endo <- all_endogenous(boot_base$path_coef) # for each endogenous var

r2_results <- lapply(endo, function(y){
  tmp <- bootstrap_R2_from_paths_and_lvcor(boot_base,target = y, use_abs = TRUE)
  ci  <- bc_excel_ci(tmp$R2_star, tmp$R2_0, conf = 0.95)
  data.frame(target = y, R2_0 = tmp$R2_0, bc_l = ci$lower, bc_u = ci$upper,
             a = ci$a, z0 = ci$z0, R = ci$R, row.names = NULL)
  })

r2_table <- do.call(rbind, r2_results)
r2_table

# 5.9 - Effect size (first-order model)
compute_f2_from_r2_df <- function(boot_obj, r2_df, endo_col = "target",
                                  r2_col = "R2_0"){
  B0  <- boot_obj$path_coef
  C0  <- cor(boot_obj$construct_scores, use = "pairwise.complete.obs")
  out <- list()
  k   <- 0
  for (i in seq_len(nrow(r2_df))){
    Y <- as.character(r2_df[[endo_col]][i])
    R2_incl   <- as.numeric(r2_df[[r2_col]][i])
    if (!is.finite(R2_incl)) next
    if (!(Y %in% colnames(B0))) next
    betas     <- B0[, Y]
    preds     <- names(betas)[!is.na(betas) & betas != 0]
    for (X in preds) {beta_j <- betas[X]
    cor_j     <- C0[X, Y]
    contrib_j <- beta_j * cor_j
    R2_excl   <- R2_incl-contrib_j
    denom     <- (1 - R2_incl)
    f2        <- if (denom <= 0) NA_real_ else (R2_incl - R2_excl) / denom
    k         <- k + 1; out[[k]] <- data.frame(from = X, to = Y,
                                          R2_incl = R2_incl, R2_excl = R2_excl,
                                          f2 = f2, row.names = NULL)
    }
  }
  res <- do.call(rbind, out)
  res$size <- cut(res$f2, breaks = c(-Inf, 0.02, 0.15, 0.35, Inf),
                  # size effect label (Cohen)
                  labels = c("negligible", "small", "medium", "large"))
  res
}

f2_results <- compute_f2_from_r2_df(boot_obj = boot_base, r2_df = r2_table,
                                    endo_col = "target", r2_col = "R2_0")
f2_results

# 5.10 - VIF first-order formative items
calc_vif <- function(data, indicators){
  purrr::map(indicators, function(v){
    others <- setdiff(indicators, v)
```

```
    if(length(others) == 0) return(NA)
    m <- lm(as.formula(paste(v, "~", paste(others, collapse = " + "))),
            data = data)
    tibble(Indicator = v, VIF = max(car::vif(m)))
  }) %>% bind_rows()}

form_vif_base <- bind_rows(
  calc_vif(DB_PLSSEM, exposure_vars)  %>% mutate(Block = "Exposure"),
  calc_vif(DB_PLSSEM, proximity_vars) %>% mutate(Block = "Proximity"),
  calc_vif(DB_PLSSEM, target_vars)    %>% mutate(Block = "TargetSuit"),
  calc_vif(DB_PLSSEM, guardian_vars)  %>% mutate(Block = "Guardian")
)

print(form_vif_base, n = Inf)

# 5.11 - INNER VIF (LOC model): collinearity among predictor constructs
calc_inner_vif <- function(model){
  # latent variable scores
  scores <- as.data.frame(model$construct_scores)
  # path coefficient matrix: rows = predictors, cols = outcomes
  B <- model$path_coef
  # endogenous constructs: those with at least one incoming path
  endo <- colnames(B)[colSums(abs(B), na.rm = TRUE) > 0]
  # for each endogenous construct, compute VIF for its predictor set
  res <- purrr::map_dfr(endo, function(y){
    preds <- rownames(B)[abs(B[, y, drop = TRUE]) > 0]
    preds <- setdiff(preds, y)  # safety, should not happen
    # VIF only makes sense if at least 2 predictors
    if(length(preds) < 2){
      return(tibble(
        Outcome = y,
        Predictor = preds,
        VIF = NA_real_
      ))
    }
    m <- lm(as.formula(paste(y, "~", paste(preds, collapse = " + "))),
            data = scores)
    tibble(
      Outcome   = y,
      Predictor = names(car::vif(m)),
      VIF       = as.numeric(car::vif(m))
    )
  })
  res
}

inner_vif_base <- calc_inner_vif(mod_base)
print(inner_vif_base, n = Inf)

# ===============================================================================
# 6. PLS-SEM: SECOND-ORDER MODEL

# 6.0 - Construction of LRAT as HOC
lv_scores_s1 <- as.data.frame(mod_base$construct_scores)

# Dataset final model
DB_PLSSEM_SECOND <- DB_PLSSEM %>%
  mutate(LV_Exposure    = lv_scores_s1$Exposure,
         LV_Proximity   = lv_scores_s1$Proximity,
         LV_TargetSuit  = lv_scores_s1$TargetSuit,
         LV_Guardian    = lv_scores_s1$Guardian )


# 6.1 - MEASUREMENT MODEL (second-order model)
mod_mm_second <- constructs(
  # FORMATIVE HOC
  composite("LRAT",
          multi_items("LV_", c("Exposure","Proximity","TargetSuit","Guardian")),
          weights = mode_B),
  # REFLECTIVES
  composite("SA",  add_vars, weights = mode_A),
  composite("SDP", sdp_vars, weights = mode_A),
  composite("SS",  ss_vars,  weights = mode_A),
  # SINGLE ITEMS
  composite("Gambling",  single_item("jugador")),
  composite("Age",       single_item("cont_edad_int")),
```

```
  composite("Gender",    single_item("cont_sexo")),
  composite("Education", single_item("cont_estudio")),
  composite("Output",    single_item("pc1a_Riesgo"))
)

# 6.2 - STRUCTURAL MODEL (second-order model)
mod_sm_second <- relationships(
  paths(from = c("SS","SDP"), to = "SA"),
  paths(from = "SA", to = "Gambling"),
  paths(from = c("SA","Gambling"), to = "LRAT"),
  paths(from = c("SA","Gambling","LRAT","Age","Gender","Education"),
        to = "Output")
)

# 6.3 - ESTIMATION OF PLS-SEM (second-order model)
mod_second <- estimate_pls(data = DB_PLSSEM_SECOND,
                           measurement_model = mod_mm_second,
                           structural_model  = mod_sm_second)

# 6.4 - Bootstrap of the model to obtain significance levels
boot_second <- bootstrap_model(mod_second,
                               nboot = 10000,
                               cores = parallel::detectCores())
summary(boot_second)

# 6.5 - OUTER VIF for HOC formative indicators (LV_ -> LRAT)
form_vif_hoc <- calc_vif(DB_PLSSEM_SECOND,
           c("LV_Exposure", "LV_Proximity", "LV_TargetSuit", "LV_Guardian")) %>%
  mutate(Block = "LRAT (HOC)")

print(form_vif_hoc, n = Inf)

# 6.6 - INNER VIF (HOC model): collinearity among predictor constructs
inner_vif_second <- calc_inner_vif(mod_second)
print(inner_vif_second, n = Inf)

# 6.7 - Bias-corrected bootstrap Conf interval
tab_base_2 <- bc_excel_ci_all_paths(boot_second,
                                    conf = 0.95,
                                    only_nonzero = TRUE,
                                    index_round = "ceiling_floor")
tab_base_2

# 6.8 - R2 bias-corrected bootstrap CI (second-order model)
endo_bis <- all_endogenous(boot_second$path_coef)
r2_results_bis <- lapply(endo_bis,
                         function(y){
                           tmp <- bootstrap_R2_from_paths_and_lvcor(boot_second,
                                                                    target = y,
                                                                    use_abs = TRUE)
                           ci <- bc_excel_ci(tmp$R2_star, tmp$R2_0, conf = 0.95)
                           data.frame(target = y, R2_0 = tmp$R2_0,
                                      bc_l= ci$lower, bc_u = ci$upper, a = ci$a,
                                      z0 = ci$z0, R = ci$R, row.names = NULL)
                         })
r2_table_bis <- do.call(rbind, r2_results_bis)
r2_table_bis

# 6.9 - Effect size (second-order model)
f2_results_bis <- compute_f2_from_r2_df(boot_obj = boot_second,
                                        r2_df    = r2_table_bis,
                                        endo_col = "target", r2_col = "R2_0")
f2_results_bis

# 6.10 - Mediation Analysis (second-order model)
boot_t_value <- function(t0, t_star) {
  t0 / sd(t_star, na.rm = TRUE)
}

total_bc_ci <- function(boot_obj, x = "SA", m = "LRAT",
                        y = "Output", conf = 0.95) {
  # point estimates
  b_dir0 <- boot_obj$path_coef[x, y]
  b1_0   <- boot_obj$path_coef[x, m]
  b2_0   <- boot_obj$path_coef[m, y]
  ie_0   <- b1_0 * b2_0
```

```r
  te_0   <- b_dir0 + ie_0
  # bootstrap draws
  b_dir_star <- as.numeric(boot_obj$boot_paths[x, y, ])
  b1_star    <- as.numeric(boot_obj$boot_paths[x, m, ])
  b2_star    <- as.numeric(boot_obj$boot_paths[m, y, ])
  ie_star    <- b1_star * b2_star
  te_star    <- b_dir_star + ie_star
  # BC CIs
  direct_ci   <- bc_excel_ci(b_dir_star, b_dir0, conf = conf)
  indirect_ci <- bc_excel_ci(ie_star,    ie_0,   conf = conf)
  total_ci    <- bc_excel_ci(te_star,    te_0,   conf = conf)
  # t-values
  dir_t <- boot_t_value(b_dir0, b_dir_star)
  ind_t <- boot_t_value(ie_0,   ie_star)
  tot_t <- boot_t_value(te_0,   te_star)
  # Output
  path_label <- paste(x, m, y, sep = " -> ")
  data.frame(
    Type = c(path_label, "Direct", "Indirect", "Total"),
    Estimate = c(NA, round(c(b_dir0, ie_0, te_0), 3)),
    Lower =    c(NA, round(c(direct_ci$lower, indirect_ci$lower,
                             total_ci$lower), 3)),
    Upper =    c(NA, round(c(direct_ci$upper, indirect_ci$upper,
                             total_ci$upper), 3)),
    t_stat =   c(NA, round(c(dir_t, ind_t, tot_t), 3)),
    VAF =      c(round(ie_0/te_0*100, 3), NA, NA, NA)
  )
}
# SA -> LRAT -> Output
med_SA_LRAT_Output       <- total_bc_ci(boot_obj = boot_second, x = "SA",
                                        m = "LRAT", y = "Output", conf = 0.95)
med_SA_LRAT_Output
# Gambling -> LRAT -> Output
med_Gambling_LRAT_Output <- total_bc_ci(boot_obj = boot_second, x = "Gambling",
                                        m = "LRAT", y = "Output", conf = 0.95)
med_Gambling_LRAT_Output
# SA -> Gambling -> Output
med_SA_Gambling_Output   <- total_bc_ci(boot_obj = boot_second, x = "SA",
                                        m ="Gambling", y ="Output", conf = 0.95)
med_SA_Gambling_Output

# Serial mediation function
serial_med_complete <- function(boot_obj, x, m1, m2, y, conf = 0.95) {
  # Original coefficients
  b_x_m1  <- boot_obj$path_coef[x, m1]
  b_m1_m2 <- boot_obj$path_coef[m1, m2]
  b_m2_y  <- boot_obj$path_coef[m2, y]
  b_direct_0 <- boot_obj$path_coef[x, y]
  # Indirect and Total
  ind_ser_0 <- b_x_m1 * b_m1_m2 * b_m2_y
  tot_ser_0 <- ind_ser_0 + b_direct_0
  # Bootstrap extraction
  star1 <- as.numeric(boot_obj$boot_paths[x, m1, ])
  star2 <- as.numeric(boot_obj$boot_paths[m1, m2, ])
  star3 <- as.numeric(boot_obj$boot_paths[m2, y, ])
  star_direct <- as.numeric(boot_obj$boot_paths[x, y, ])
  ind_ser_star <- star1 * star2 * star3
  tot_ser_star <- ind_ser_star + star_direct
  # CI Bias-Corrected
  ci_ind <- bc_excel_ci(ind_ser_star, ind_ser_0, conf = conf)
  ci_dir <- bc_excel_ci(star_direct, b_direct_0, conf = conf)
  ci_tot <- bc_excel_ci(tot_ser_star, tot_ser_0, conf = conf)
  # Output
  path_label <- paste(x, m1, m2, y, sep = " -> ")
  data.frame(
    Type = c(path_label, "Direct", "Indirect", "Total"),
    Estimate = c(NA, round(c(b_direct_0,   ind_ser_0,    tot_ser_0), 3)),
    Lower =    c(NA, round(c(ci_dir$lower, ci_ind$lower, ci_tot$lower), 3)),
    Upper =    c(NA, round(c(ci_dir$upper, ci_ind$upper, ci_tot$upper), 3)),
    t_stat =   c(NA, round(c(b_direct_0 / sd(star_direct),
                             ind_ser_0  / sd(ind_ser_star),
                             tot_ser_0  / sd(tot_ser_star)), 3)),
    VAF =      c(round(ind_ser_0/tot_ser_0*100, 3), NA, NA, NA)
  )}
# SA -> Gambling -> LRAT -> Output
med_SA_G_LRAT_Output <- serial_med_complete(boot_second, "SA", "Gambling",
```

```
                                            "LRAT", "Output")
med_SA_G_LRAT_Output

# ============================================================================
# 7. ML - MODEL INITIALIZATION

# 7.0 - efine binary target (classification)
DB_ML <- DB_ML %>%
  mutate(
    RiskBin = factor(ifelse(pc1a_Riesgo > 0, "Victim", "NonVictim"),
                     levels = c("NonVictim", "Victim"))
  )
set.seed(123)

# 7.1 - Train/test split (80/20), stratified on the binary outcome
split_rf <- initial_split(DB_ML, prop = 0.80, strata = RiskBin)
train_data <- training(split_rf)
test_data  <- testing(split_rf)

# 7.2 10-fold CV on training set
folds_rf <- vfold_cv(train_data, v = 10, strata = RiskBin)

# 7.3 - Recipe
rec_rf <- recipe(RiskBin ~ ., data = train_data) %>%
  # Removing pc1a_Riesgo
  step_rm(pc1a_Riesgo) %>%
  # Removing zero variance varaibleszero
  step_zv(all_predictors())

# 7.4 - Random Forest specification (classification)
rf_spec <- rand_forest(
  mode  = "classification",
  trees = tune(),  # to be optimised
  mtry  = tune(),  # to be optimised
  min_n = tune()   # to be optimised
) %>%
  set_engine(
    "ranger",
    importance      = "permutation", # measurements of varying importance
    sample.fraction = 0.632,    # sampling fraction per tree (bootstrap ~ 63.2%)
    max.depth       = tune(),        # to be optimised
    replace         = TRUE           # sampling with replacement
  )

wf_rf <- workflow() %>%
  add_model(rf_spec) %>%
  add_recipe(rec_rf)

# 7.5 Parameter ranges
rf_params <- parameters(
  trees(range = c(500, 2000)),
  mtry() %>% finalize(select(train_data, -RiskBin, -pc1a_Riesgo)),
  min_n(range = c(2, 30)),
  tree_depth(range = c(2, 20))
)
rf_params$id[rf_params$id == "tree_depth"] <- "max.depth"

# 7.6 - Tuning grid
set.seed(123)
rf_grid <- grid_space_filling(rf_params, size = 30)
colnames(rf_grid)

# ============================================================================
# 8. ML - MODEL TUNING VIA CROSS-VALIDATION

set.seed(123)

# 8.1 - Model tuning
rf_tuned <- tune_grid(
  wf_rf,
  resamples = folds_rf,
  grid      = rf_grid,
  metrics   = metric_set(roc_auc, accuracy),
  control   = control_grid(save_pred = TRUE)
)
```

```r
# 8.2 - Selection
best_rf <- select_best(rf_tuned, metric = "roc_auc")
best_rf

wf_rf_final <- finalize_workflow(wf_rf, best_rf)

rf_final_fit <- last_fit(
  wf_rf_final,
  split = split_rf,
  metrics = metric_set(roc_auc, accuracy)
)

collect_metrics(rf_final_fit)

# ==============================================================================
# 9. ML - EVALUATION ON THE TEST SET

# 9.1 - Collect test predictions from last_fit()
rf_test_preds <- collect_predictions(rf_final_fit)

# Sanity check: what columns are available?
dplyr::glimpse(rf_test_preds)

# 9.2 - Confusion matrix (default threshold 0.5)
cat("\n\nConfusion matrix (default threshold 0.5)\n")
conf_mat(rf_test_preds, truth = RiskBin, estimate = .pred_class)

# 9.3 - Full set of classification metrics
metrics_full <- metric_set(
  yardstick::accuracy,
  yardstick::sens,      # recall (TPR)
  yardstick::spec,      # specificity
  yardstick::ppv,       # precision
  yardstick::f_meas,
  yardstick::roc_auc
)

metrics_full(
  rf_test_preds,
  truth = RiskBin,
  estimate = .pred_class,
  .pred_Victim,
  event_level = "second"
)

# 9.4 - ROC curve data (for plotting/reporting)
roc_df <- roc_curve(
  rf_test_preds,
  truth = RiskBin,
  .pred_Victim,
  event_level = "second"
)

roc_df

# 9.5 - Plot ROC curve
autoplot(roc_df)

library(ggplot2)

auc_val <- yardstick::roc_auc(
  rf_test_preds,
  truth = RiskBin,
  .pred_Victim,
  event_level = "second"
)$.estimate

ggplot(roc_df, aes(x = 1 - specificity, y = sensitivity)) +
  geom_line(linewidth = 1.2, color = "#8EA5B6") +
  geom_abline(
    linetype = "dashed",
    color = "grey60",
    linewidth = 0.8
  ) +
  annotate(
    "text",
```

```
    x = 0.55,
    y = 0.20,
    label = paste0("AUC = ", round(auc_val, 3)),
    size = 5
  ) +
  labs(
    x = "1 - Specificity (False Positive Rate)",
    y = "Sensitivity (True Positive Rate)"
  ) +
  theme_minimal(base_size = 14) +
  theme(
    panel.grid.minor = element_blank()
  )

# ==============================================================================
# 10. ML - VARIABLE IMPORTANCE

# 10.1 - Final fit of the workflow
rf_final_workflow_full <- wf_rf_final %>%
  fit(data = DB_ML)

# 10.2 - Extraction of the ranger model
rf_engine <- extract_fit_engine(rf_final_workflow_full)

# 10.3 - Variable importance (permutation)
rf_var_imp <- ranger::importance(rf_engine)

rf_var_imp_tbl <- tibble::tibble(
  variable   = names(rf_var_imp),
  importance = as.numeric(rf_var_imp)
) %>%
  dplyr::arrange(desc(importance))

rf_var_imp_tbl

summary(rf_var_imp_tbl$importance) # Sanity check

# 10.4 - Selection of “relevant” variables
rf_var_imp_top <- rf_var_imp_tbl %>%
  dplyr::filter(importance > 0) %>%
  dplyr::slice_max(importance, n = 15)

rf_var_imp_top

# 10.5 - Plot
library(ggplot2)

ggplot(rf_var_imp_top,
       aes(x = reorder(variable, importance), y = importance)) +
  geom_col(fill = "#8EA5B6") +
  coord_flip() +
  labs(
    x = "Predictor",
    y = "Permutation Importance"
  ) +
  theme_minimal(base_size = 14) +
  theme(
    panel.grid.minor = element_blank()
  )
```