



**EDUCACIÓN**

SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO  
NACIONAL DE MÉXICO

# Tecnológico Nacional de México

**Centro Nacional de Investigación  
y Desarrollo Tecnológico**

## Tesis de Maestría

**Detección automática de trastorno de ansiedad  
generalizada utilizando estrategias de historia de vida**

presentada por

**Ing. Ana Luisa Islas Ávila**

como requisito para la obtención del grado de  
**Maestra en Ciencias de la Computación**

Director de tesis

**Dra. Alicia Martínez Rebollar**

Codirectores de tesis

**Dr. Hugo Estrada Esquivel**

**Dra. Angélica Fabiola Sánchez Gutiérrez**

**Cuernavaca, Morelos, México. Agosto de 2025.**



# Educación

Secretaría de Educación Pública



TECNOLÓGICO  
NACIONAL DE MÉXICO



Centro Nacional de Investigación y Desarrollo tecnológico  
Departamento de Ciencias Computacionales

Cuernavaca, Mor., 12/Junio/2025

OFICIO No. DCC/180/2025

Asunto: Aceptación de documento de tesis

CENIDET-AC-004-M14-OFICIO

**CARLOS MANUEL ASTORGA ZARAGOZA**  
**SUBDIRECTOR ACADÉMICO**  
**PRESENTE**

Por este conducto, los integrantes de Comité Tutorial de Ana Luisa Islas Ávila con número de control M23CE066, de la Maestría en Ciencias de la Computación, le informamos que hemos revisado el trabajo de tesis de grado titulado **"Detección automática de trastorno de ansiedad generalizada utilizando estrategias de historia de vida"** y hemos encontrado que se han atendido todas las observaciones que se le indicaron, por lo que hemos acordado aceptar el documento de tesis y le solicitamos la autorización de impresión definitiva.

**A T E N T A M E N T E**

*Excelencia en Educación Tecnológica®*  
*"Conocimiento y Tecnología al Servicio de México"*

  
\_\_\_\_\_  
Dra. Alicia Martínez Rebollar  
Directora de tesis

  
\_\_\_\_\_  
Dr. Hugo Estrada Esquivel  
Codirector de tesis

  
\_\_\_\_\_  
Dra. Angelica Fabiola Sánchez Gutiérrez  
Codirectora de tesis

  
\_\_\_\_\_  
Dr. Dante Mújica Vargas  
Revisor 1

  
\_\_\_\_\_  
Dra. Wendy Aracely Sánchez Gómez  
Revisor 2

C.c.p. Depto. Servicios Escolares.  
Expediente / Estudiante



**2025**  
Año de  
**La Mujer**  
**Indígena**

Interior Internado Palmira S/N, Col. Palmira,  
C. P. 62490, Cuernavaca, Morelos Tel. 01 (777) 3627770, ext. 3201,  
e-mail: dcc\_cenidet@tecnm.mx | tecnm.mx | cenidet.tecnm.mx

**cenidet**  
Centro Nacional de Investigación  
y Desarrollo Tecnológico





Centro Nacional de Investigación y Desarrollo tecnológico  
Subdirección Académica

Cuernavaca Mor, 16/junio/2025

Oficio No. SAC/145/2025

Asunto: Autorización de impresión de tesis

**ANA LUISA ISLAS ÁVILA**  
**CANDIDATA AL GRADO DE MAESTRA**  
**EN CIENCIAS DE LA COMPUTACIÓN**  
**P R E S E N T E**

Por este conducto, tengo el agrado de comunicarle que el Comité Tutorial asignado a su trabajo de tesis titulado **"Detección automática de trastorno de ansiedad generalizada utilizando estrategias de historia de vida"**, ha informado a esta Subdirección Académica, que están de acuerdo con el trabajo presentado. Por lo anterior, se le autoriza a que proceda con la impresión definitiva de su trabajo de tesis.

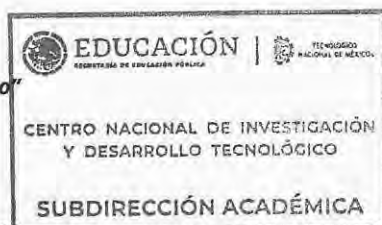
Esperando que el logro del mismo sea acorde con sus aspiraciones profesionales, reciba un cordial saludo.

**ATENTAMENTE**

*Excelencia en Educación Tecnológica®*

*"Conocimiento y Tecnología al Servicio de México"*

**CARLOS MANUEL ASTORGA ZARAGOZA**  
**SUBDIRECTOR ACADÉMICO**



c.c.p. Departamento de Ciencias Computacionales  
Departamento de Servicios Escolares

CMAZ/lmz



**2025**  
Año de  
**La Mujer**  
**Indígena**

Interior Internado Palmira S/N, Col. Palmira,  
C. P. 62490, Cuernavaca, Morelos Tel. 01 (777) 3627770, ext. 4104,  
e-mail: acad\_cenidet@tecnm.mx tecnm.mx | cenidet.tecnm.mx

**cenidet**  
Centro Nacional de Investigación  
y Desarrollo Tecnológico



# Dedicatoria

*A mi mamá, Laura,  
por su amor inmesurable que me sostiene  
incluso cuando el mundo se tambalea,  
por su apoyo incansable en los días grises  
y por motivarme a perseguir mis sueños.*

# Agradecimientos

Expreso mi gratitud a la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) por la beca otorgada para mi formación en el programa de maestría. Asimismo, agradezco al Tecnológico Nacional de México / Centro Nacional de Investigación y Desarrollo Tecnológico (TECNM/CENIDET) por facilitar los recursos y espacios necesarios para el desarrollo de mis estudios de posgrado.

A Dios, por brindarme la oportunidad de comenzar esta nueva etapa de mi vida y guiar mis decisiones.

A mi mamá, por nunca darse por vencida conmigo y apoyarme durante cada etapa de este posgrado. A mi familia, por creer en mí y motivarme a ser mejor cada día.

A mi directora de tesis, la Dra. Alicia Martínez, por la confianza depositada en mí al permitirme realizar mis estudios de maestría en esta institución bajo su supervisión. Su guía académica y apoyo continuo fueron esenciales durante estos dos años. Reconozco igualmente la valiosa colaboración de mis codirectores, el Dr. Hugo Estrada y la Dra. Fabiola Sánchez, quienes contribuyeron con su experiencia y dedicaron tiempo valioso al acompañamiento de esta investigación.

Expreso también mi gratitud a los revisores de este trabajo, el Dr. Dante Mújica y la Dra. Wendy Sánchez, por sus observaciones, retroalimentación enriquecedora y la atención brindada durante el proceso de evaluación.

Al cuerpo docente del Departamento de Ciencias de la Computación y al personal administrativo, por su continuo acompañamiento y apoyo.

A mi amigo, Ricardo Castrejón, por su guía y palabras de aliento en los tiempos difíciles.

# Resumen

Esta tesis presenta el desarrollo de un modelo para la detección del Trastorno de Ansiedad Generalizada (TAG) mediante algoritmos de aprendizaje automático supervisado. El objetivo principal de la investigación fue evaluar el desempeño de distintos modelos supervisados utilizando métricas clave bajo diversas configuraciones, incluyendo la selección de características y la optimización de hiperparámetros.

Para ello, se analizaron 244 registros obtenidos a partir de cuatro cuestionarios que evaluaron variables sociodemográficas, estrategias de historia de vida y niveles de ansiedad. Tras un proceso de selección de características relevantes, se evaluaron siete algoritmos supervisados en dos condiciones: configuración de hiperparámetros por defecto y configuración optimizada. Entre ellos, el modelo Random Forest, ajustado mediante submuestreo aleatorio y búsqueda aleatoria de hiperparámetros, obtuvo el mejor desempeño, alcanzando una exactitud del 81 % y una sensibilidad del 78 %.

Los resultados destacan la relevancia de las estrategias de historia de vida como variables predictoras del TAG, en particular dimensiones como las relaciones interpersonales, la maduración precoz y la impulsividad, las cuales mostraron asociaciones significativas con dicho trastorno. Estos hallazgos respaldan la viabilidad de incorporar variables de historia de vida en modelos predictivos de TAG, aportando una perspectiva integral que considera factores conductuales, cognitivos y sociodemográficos.

**Palabras clave:** aprendizaje automático supervisado; clasificación; estrategias de historia de vida; trastorno de ansiedad generalizada.

# Abstract

This thesis presents the development of a model for detecting Generalized Anxiety Disorder (GAD) using supervised machine learning algorithms. The main objective of the research was to evaluate the performance of various supervised models using key metrics under different configurations, including feature selection and hyperparameter optimization.

For this, 244 records were analyzed, collected through four questionnaires that assessed sociodemographic variables, life history strategies, and anxiety levels. After selecting relevant features, seven supervised algorithms were evaluated under two conditions: default hyperparameter configuration and optimized configuration. Among them, the Random Forest model, tuned using random undersampling and random search for hyperparameter optimization, achieved the best performance, with an accuracy of 81 % and a sensitivity of 78 %.

The results highlight the relevance of life history strategies as predictive variables for GAD, particularly dimensions such as interpersonal relationships, early maturation, and impulsivity, which showed significant associations with the disorder. These findings support the feasibility of incorporating life history variables into predictive models of GAD, providing a comprehensive perspective that includes behavioral, cognitive, and sociodemographic factors.

**Keywords:** classification; generalized anxiety disorder; life history strategies; supervised machine learning.

# Índice general

|                                                                       |             |
|-----------------------------------------------------------------------|-------------|
| <b>Lista de Figuras</b>                                               | <b>VI</b>   |
| <b>Lista de Tablas</b>                                                | <b>VIII</b> |
| <b>1. Introducción</b>                                                | <b>1</b>    |
| 1.1. Contexto general . . . . .                                       | 1           |
| 1.2. Planteamiento del problema . . . . .                             | 2           |
| 1.3. Objetivos . . . . .                                              | 2           |
| 1.3.1. Objetivo general . . . . .                                     | 2           |
| 1.3.2. Objetivos específicos . . . . .                                | 3           |
| 1.4. Metodología de solución . . . . .                                | 3           |
| 1.5. Organización de la tesis . . . . .                               | 3           |
| <b>2. Marco teórico</b>                                               | <b>5</b>    |
| 2.1. Fundamentos del aprendizaje automático . . . . .                 | 5           |
| 2.2. Trastorno de Ansiedad Generalizada . . . . .                     | 13          |
| 2.3. Estrategias de historia de vida . . . . .                        | 14          |
| <b>3. Estado del arte</b>                                             | <b>17</b>   |
| 3.1. Clasificación de los trabajos del estado del arte . . . . .      | 17          |
| 3.2. Predicción de los trastornos de ansiedad . . . . .               | 17          |
| 3.3. Predicción de ansiedad, estrés o depresión . . . . .             | 20          |
| 3.4. Resumen de los trabajos revisados . . . . .                      | 23          |
| <b>4. Modelo para detección de trastorno de ansiedad generalizada</b> | <b>27</b>   |
| 4.1. Introducción a la metodología . . . . .                          | 27          |
| 4.2. Recolección de datos . . . . .                                   | 28          |
| 4.3. Preprocesamiento del conjunto de datos . . . . .                 | 31          |
| 4.4. Desarrollo del modelo . . . . .                                  | 38          |
| 4.5. Optimización del modelo . . . . .                                | 43          |
| <b>5. Validación y análisis de resultados</b>                         | <b>47</b>   |
| 5.1. Metodología de validación . . . . .                              | 47          |
| 5.2. Recolección del segundo conjunto de datos . . . . .              | 48          |
| 5.3. Validación del modelo . . . . .                                  | 48          |
| 5.4. Interpretación del modelo Random Forest . . . . .                | 55          |

|                                                        |           |
|--------------------------------------------------------|-----------|
| <b>6. Conclusiones y trabajos futuros</b>              | <b>61</b> |
| 6.1. Conclusiones . . . . .                            | 61        |
| 6.2. Trabajos futuros . . . . .                        | 62        |
| 6.3. Productos derivados de la investigación . . . . . | 62        |
| <b>A. Consentimiento informado</b>                     | <b>70</b> |
| <b>B. Instrumentos utilizados</b>                      | <b>72</b> |

# Índice de figuras

|                                                                                                                                                                 |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1.1. Metodología de solución . . . . .                                                                                                                          | 4  |
| 2.1. Diagrama de flujo para realizar un Análisis Factorial Exploratorio . . . . .                                                                               | 10 |
| 2.2. Continuo rápido-lento en estrategias de historia de vida y perfiles . . . . .                                                                              | 15 |
| 2.3. Estructura de clasificación del modelo FSD. . . . .                                                                                                        | 15 |
| 4.1. Metodología implementada para el desarrollo del modelo de detección de trastorno de ansiedad generalizada . . . . .                                        | 27 |
| 4.2. Histograma y gráfico de caja de la variable Edad . . . . .                                                                                                 | 33 |
| 4.3. Histograma y gráfico de caja de la variable Factor K normalizado . . . . .                                                                                 | 33 |
| 4.4. Diagrama del análisis factorial exploratorio de cuestionario Marcadores de estrategias de historia de vida con 2 factores con cargas factoriales . . . . . | 36 |
| 4.5. Comparativa de distribución de las clases antes y después de utilizar la técnica de balanceo submuestreo aleatorio (Random Undersampling) . . . . .        | 37 |
| 4.6. Comparativa de resultados de las métricas de desempeño en los modelos implementados utilizando el conjunto de datos sin balancear . . . . .                | 40 |
| 4.7. Comparativa de resultados de las métricas de desempeño en los modelos implementados utilizando el conjunto de datos balanceado . . . . .                   | 41 |
| 4.8. Matrices de confusión de los modelos que utilizan el conjunto de datos balanceado con submuestreo aleatorio . . . . .                                      | 42 |
| 4.9. Comparación de curvas ROC del modelo Random Forest . . . . .                                                                                               | 43 |
| 5.1. Metodología implementada para la validación modelo de detección de trastorno de ansiedad generalizada . . . . .                                            | 47 |
| 5.2. Matriz de confusión de la validación del Modelo Random Forest sin optimizar                                                                                | 49 |
| 5.3. Evaluación del modelo Random Forest sin optimizar mediante curvas ROC macro-promedio y Precision-Recall ( <i>One-vs-Rest</i> ) . . . . .                   | 50 |
| 5.4. Matriz de confusión de la validación del Modelo Random Forest optimizado con Búsqueda en Malla . . . . .                                                   | 51 |
| 5.5. Evaluación del modelo Random Forest optimizado con búsqueda en malla mediante curvas ROC y Precision-Recall ( <i>One-vs-Rest</i> ) . . . . .               | 52 |
| 5.6. Matriz de confusión de la validación del Modelo Random Forest optimizado con Búsqueda Aleatoria . . . . .                                                  | 53 |
| 5.7. Evaluación del modelo Random Forest optimizado con búsqueda aleatoria mediante curvas ROC y Precision-Recall ( <i>One-vs-Rest</i> ) . . . . .              | 53 |

|                                                                                                                                                        |    |
|--------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 5.8. Matriz de confusión de la validación del Modelo Random Forest optimizado con Optimización Bayesiana . . . . .                                     | 54 |
| 5.9. Evaluación del modelo Random Forest optimizado con optimización bayesiana mediante curvas ROC y Precision-Recall ( <i>One-vs-Rest</i> ) . . . . . | 54 |
| 5.10. Importancia global de características del modelo Random Forest optimizado con búsqueda aleatoria . . . . .                                       | 56 |
| 5.11. Importancia de características por clase del modelo Random Forest optimizado con búsqueda aleatoria . . . . .                                    | 56 |
| 5.12. Resumen de los valores SHAP obtenidos del modelo Random Forest optimizado con búsqueda aleatoria para la clase TAG . . . . .                     | 57 |
| 5.13. Gráficos de dependencia de las características sociodemográficas del modelo y los valores SHAP . . . . .                                         | 58 |
| 5.14. Gráficos de dependencia de las características de estrategias de historia de vida del modelo y los valores SHAP . . . . .                        | 59 |
| 5.15. Gráficos de dependencia de las variables de preocupación y los valores SHAP . . . . .                                                            | 60 |
| A.1. Consentimiento informado (parte 1) . . . . .                                                                                                      | 70 |
| A.2. Consentimiento informado (parte 2) . . . . .                                                                                                      | 71 |
| B.1. Vista parcial de cuestionario sociodemográfico . . . . .                                                                                          | 72 |
| B.2. Vista parcial del cuestionario Marcadores de estrategias de historia de vida . . . . .                                                            | 73 |
| B.3. Vista parcial de cuestionario K-SF-42 (parte 1) . . . . .                                                                                         | 74 |
| B.4. Vista parcial de cuestionario K-SF-42 (parte 2) . . . . .                                                                                         | 75 |
| B.5. Vista parcial de cuestionario PSWQ . . . . .                                                                                                      | 76 |

# Índice de tablas

|                                                                                                                                                        |    |
|--------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2.1. Resumen de las métricas de evaluación utilizadas en la validación del modelo                                                                      | 7  |
| 3.1. Resumen comparativo de los trabajos del estado del arte . . . . .                                                                                 | 24 |
| 4.1. Resumen de las variables recolectadas de los instrumentos Marcadores de estrategias de historia de vida, K-SF-42 y PSWQ . . . . .                 | 30 |
| 4.2. Resumen de características sociodemográficas, factor K y categoría de ansiedad de los participantes . . . . .                                     | 32 |
| 4.3. Análisis bivariado de las variables categóricas sociodemográficas . . . . .                                                                       | 34 |
| 4.4. Análisis bivariado de las variables cuantitativas (edad y factor K) . . . . .                                                                     | 34 |
| 4.5. Resumen de resultados de las medidas de bondad de ajuste del análisis factorial confirmatorio . . . . .                                           | 36 |
| 4.6. Resumen de las variables de entrada utilizadas en el modelo para detección .                                                                      | 38 |
| 4.7. Configuración de hiperparámetros inicial de los algoritmos de aprendizaje supervisado seleccionados . . . . .                                     | 39 |
| 4.8. Resultados de las pruebas con los modelos implementados utilizando el conjunto de datos sin balancear . . . . .                                   | 39 |
| 4.9. Resultados de las pruebas con los modelos implementados utilizando el conjunto de datos balanceado . . . . .                                      | 40 |
| 4.10. Espacio de búsqueda utilizado para la optimización de hiperparámetros de los modelos de aprendizaje supervisado seleccionados . . . . .          | 44 |
| 4.11. Configuración de hiperparámetros obtenida con la técnica de búsqueda en malla ( <i>Grid Search</i> ) . . . . .                                   | 44 |
| 4.12. Configuración de hiperparámetros obtenida con la técnica de búsqueda aleatoria ( <i>Random Search</i> ) . . . . .                                | 45 |
| 4.13. Configuración de hiperparámetros obtenida con la técnica de optimización Bayesiana . . . . .                                                     | 45 |
| 4.14. Resultados de las pruebas con los modelos optimizados utilizando el conjunto de datos balanceado . . . . .                                       | 46 |
| 5.1. Resumen de características sociodemográficas, factor K, y categoría de ansiedad de los participantes de la segunda recolección de datos . . . . . | 48 |
| 5.2. Resultados de las métricas de evaluación en la validación del modelo Random Forest sin optimizar . . . . .                                        | 49 |
| 5.3. Resultados de la validación con el segundo conjunto de datos de cada modelo Random Forest optimizado . . . . .                                    | 51 |

# Capítulo 1

## Introducción

El presente capítulo tiene como propósito establecer los fundamentos conceptuales y metodológicos de esta investigación. Se presenta el contexto general del problema de estudio, se detallan los objetivos específicos establecidos, se delimitan los alcances y limitaciones del trabajo, se describe la metodología de solución implementada y, finalmente, se explica la organización estructural del documento completo.

### 1.1. Contexto general

En México, el proyecto Encovid-19 reportó una prevalencia del Trastorno de Ansiedad Generalizada del 31-33 % en 2020 (Gaitán-Rossi et al., 2021). Las estrategias de historia de vida, un enfoque de la psicopatología evolutiva, pueden ser utilizadas para realizar predicciones de la psicología de un individuo y el posible desarrollo de una psicopatología (Giudice et al., 2015; Kavanagh y Kahl, 2016). Aunque el aprendizaje automático ha ganado relevancia en la detección de trastornos mentales, existe poca evidencia sobre modelos que utilicen variables de estrategias de historia de vida. Por tal razón, el propósito de este proyecto es desarrollar un modelo de detección automática del trastorno de ansiedad generalizada basado en información relacionada con las estrategias de historia de vida. Este enfoque busca mejorar la comprensión y detección temprana de los trastornos de ansiedad, proporcionando nuevas perspectivas para el diagnóstico clínico. Se implementó una variedad de modelos de aprendizaje automático, incorporando selección de características, estrategias de balanceo de clases y optimización de hiperparámetros para garantizar un rendimiento óptimo, confirmado mediante la validación del mejor modelo. Los resultados de esta investigación no solo contribuirán al campo de la salud mental desde una perspectiva técnica, sino que también aportarán conocimientos valiosos sobre los factores clave que influyen en el desarrollo y manifestación de la ansiedad generalizada.

## 1.2. Planteamiento del problema

Los trastornos de salud mental son un desafío global importante, con aproximadamente uno de cada ocho individuos afectados en todo el mundo (WHO, 2023) y una alta prevalencia en México (Instituto Mexicano del Seguro Social, 2023). El Trastorno de Ansiedad Generalizada (TAG) destaca por su impacto significativo en la calidad de vida y, a pesar de los avances en la comprensión y el tratamiento de los trastornos de ansiedad, aún persisten desafíos importantes en la detección temprana y la predicción precisa de este trastorno.

En este contexto, el uso de algoritmos de aprendizaje automático en salud mental es crucial debido al crecimiento de los datos y la necesidad de mejorar la precisión en diagnóstico y predicción. Se han comparado varios algoritmos en estudios de casos, incluyendo Bosque Aleatorio, Redes Neuronales Artificiales y XGBoost (Baba y Bunji, 2023; Kumar et al., 2023; Mutalib y Al, 2021), pero aún no se han identificado las características de estrategias de historia de vida con mayor importancia para la detección del trastorno de ansiedad generalizada.

El problema central de esta investigación radica en determinar el modelo de aprendizaje automático más efectivo para detectar el trastorno de ansiedad generalizada al considerar las estrategias de historia de vida de un individuo. Sin embargo, esta tarea se complica debido a la falta de conjuntos de datos que combinen información detallada sobre las estrategias de historia de vida con la presencia o ausencia del trastorno de ansiedad generalizada, así como la identificación de variables relevantes en este contexto.

Para abordar este problema, se plantean dos preguntas clave: *¿Qué características de las estrategias de historia de vida se consideran importantes para evaluar el desarrollo del trastorno de ansiedad generalizada?*, *¿Cuál es el rendimiento de los modelos de aprendizaje automático para predecir el trastorno de ansiedad generalizada a partir de aspectos de la historia de vida de una persona?*. La respuesta a estas preguntas de investigación permitirá identificar el modelo de aprendizaje automático más eficaz en la detección del trastorno de ansiedad generalizada utilizando características relevantes de la historia de vida de un individuo.

## 1.3. Objetivos

### 1.3.1. Objetivo general

Construir un modelo de aprendizaje automático interpretable para la detección del Trastorno de Ansiedad Generalizada mediante algoritmos supervisados y técnicas de inteligencia artificial explicable aplicados a variables psicométricas, conductuales y sociodemográficas, con el propósito de identificar perfiles de riesgo.

### 1.3.2. Objetivos específicos

1. Construir un conjunto de datos estructurado y preprocesado mediante la aplicación en población mexicana de instrumentos psicométricos validados para obtener mediciones confiables del factor K, marcadores de estrategias de historia de vida, niveles de preocupación y variables sociodemográficas relevantes.
2. Desarrollar múltiples modelos supervisados de aprendizaje automático mediante algoritmos de clasificación a partir del conjunto de datos de variables psicométricas y sociodemográficas para detectar la presencia del Trastorno de Ansiedad Generalizada.
3. Evaluar la eficacia del modelo de detección del Trastorno de Ansiedad Generalizada mediante métricas de desempeño específicas, utilizando técnicas de validación cruzada y pruebas en un conjunto de datos independiente, con el fin de determinar su capacidad de generalización.

## 1.4. Metodología de solución

La metodología de solución propuesta para abordar el desarrollo del proyecto se representa en la Figura 1.1 e implica cuatro fases distintas: **(1)** Recolección de datos para el caso de estudio de trastorno de ansiedad generalizada, **(2)** Preprocesamiento de datos obtenidos de los instrumentos psicológicos, **(3)** Desarrollo del modelo para detección de trastorno de ansiedad generalizada, y **(4)** Validación del modelo para detección de trastorno de ansiedad generalizada.

## 1.5. Organización de la tesis

El presente documento se estructura en seis capítulos, anexos y referencias. El Capítulo 2 presenta los fundamentos teóricos relacionados con esta investigación, incluyendo una descripción detallada de los algoritmos de clasificación utilizados, las métricas de evaluación aplicadas, así como una revisión conceptual del trastorno de ansiedad generalizada y de las estrategias de historia de vida. El Capítulo 3 presenta la revisión de los trabajos reportados en la literatura que sirvieron de base para la investigación. En el Capítulo 4 se detalla el proceso implementado para el desarrollo del modelo de detección para el trastorno de ansiedad generalizada, abarcando la recolección y el preprocesamiento de los datos, los resultados preliminares obtenidos con los primeros modelos implementados y el proceso de optimización de hiperparámetros. El Capítulo 5 abarca la validación del modelo seleccionado con un nuevo conjunto de datos. Se presentan los resultados obtenidos y se ofrece una interpretación del modelo construido. Finalmente, en el Capítulo 6 se presentan las conclusiones derivadas de este proyecto, los trabajos futuros y los productos académicos generados a partir de la investigación.

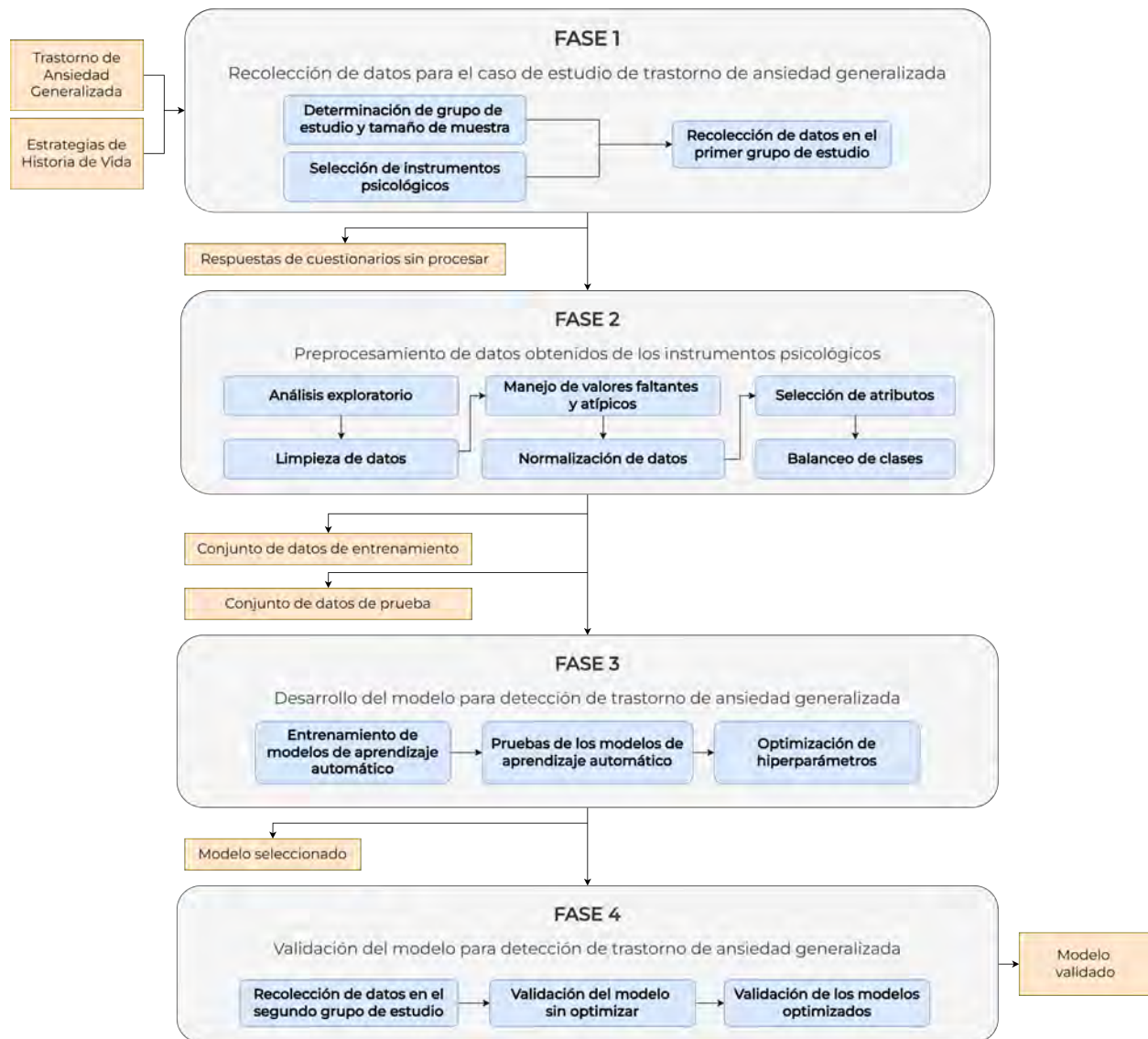


Figura 1.1: Metodología de solución

# Capítulo 2

## Marco teórico

En este capítulo se presentan los fundamentos teóricos que sustentan esta investigación y que permiten comprender la relevancia del enfoque adoptado. Se describen los conceptos clave del aprendizaje automático, explicando los principios básicos de los algoritmos de clasificación supervisados utilizados y las métricas de evaluación de desempeño. Además, se abordan los aspectos teóricos relacionados con el trastorno de ansiedad generalizada y las estrategias de historia de vida.

### 2.1. Fundamentos del aprendizaje automático

El aprendizaje automático (*Machine Learning*) es un subconjunto de la inteligencia artificial que se centra en utilizar conjuntos de datos que pueden ser limitados o extensos y algoritmos escalables (Burkov, 2019). En el ámbito de la salud mental, se ha utilizado para desarrollar modelos que predicen o detectan trastornos a partir de variables específicas y los algoritmos representativos de aprendizaje supervisado que se utilizan en varios trabajos relacionados con la salud mental son Máquinas de Soporte Vectorial, Random Forest, Naive Bayes, K Vecinos más Cercanos, XGBoost, Regresión Logística, AdaBoost y Perceptrón Multicapa (Cho et al., 2019; Chung y Teo, 2023).

#### Algoritmos de aprendizaje automático supervisado

La **Máquina de Soporte Vectorial** (SVM, por sus siglas en inglés) es un algoritmo de aprendizaje supervisado utilizado para tareas de clasificación. En su versión lineal, el objetivo principal es encontrar un hiperplano que separe de manera óptima dos clases en un espacio de características, maximizando el margen entre los datos más cercanos de cada clase, denominados vectores de soporte (Shalev-Shwartz y Ben-David, 2014). El modelo es definido como:

$$f(x) = w^T x + b \quad (2.1)$$

donde  $w \in \mathbb{R}^d$  es el vector de pesos y  $b \in \mathbb{R}$  es el sesgo. El objetivo es encontrar los parámetros  $w$  y  $b$  que definan un hiperplano de separación con el máximo margen posible.

**Random Forest** es un algoritmo de aprendizaje supervisado basado en el principio de ensamble de modelos, cuyo objetivo es mejorar la precisión y la capacidad de generalización mediante la combinación de múltiples árboles de decisión construidos sobre subconjuntos aleatorios de los datos (Ordóñez, 2006). En problemas de clasificación, el modelo construye un conjunto de  $T$  árboles de decisión  $\{h_t(x)\}_{t=1}^T$ , cada uno entrenado de forma independiente, y la predicción final para una nueva observación  $x$  se obtiene mediante votación mayoritaria entre los árboles:

$$\hat{y} = \text{modo}\{h_t(x)\}_{t=1}^T \quad (2.2)$$

**Naive Bayes** es un clasificador probabilístico basado en el teorema de Bayes, que asume independencia entre las variables predictivas dado el valor de la clase (Cho et al., 2019). El objetivo es estimar la probabilidad posterior  $P(x | y)$  para predecir la clase más probable de una nueva observación  $x$ :

$$P(x | y) = \prod_{j=1}^d P(x_j | y) \quad (2.3)$$

**K Vecinos más Cercanos** (KNN, por sus siglas en inglés) es un algoritmo basado en instancias que consiste en asignar una clase a una nueva observación  $x$  en función de las clases de sus  $k$  vecinos más cercanos en el espacio de características (Burkov, 2019). La distancia entre observaciones se calcula generalmente mediante la distancia euclidiana:

$$\text{dist}(x, x_i) = \sqrt{\sum_{j=1}^d (x_j - x_{ij})^2} \quad (2.4)$$

**XGBoost** (*Extreme Gradient Boosting*) es un algoritmo de aprendizaje supervisado basado en árboles de decisión donde el principio fundamental está basado en la construcción de un modelo aditivo mediante la combinación secuencial de múltiples modelos débiles, donde cada modelo nuevo corrige los errores cometidos por el conjunto de modelos anteriores (Chen y Guestrin, 2016). Sea  $x_i \in \mathbb{R}^d$  un vector de características y  $y_i$  la variable de salida, el modelo XGBoost busca predecir  $y_i$  como una suma de funciones de predicción  $f_t \in \mathcal{F}$ , donde  $\mathcal{F}$  representa el espacio de árboles posibles:

$$\hat{y}_i = \sum_{t=1}^T f_t(x_i), \quad f_t \in \mathcal{F} \quad (2.5)$$

La **regresión logística multinomial** es una extensión de la regresión logística binaria utilizada para modelar problemas de clasificación con múltiples clases mutuamente excluyentes. Es un modelo de aprendizaje supervisado que estima directamente la probabilidad de pertenencia de una observación a cada una de las  $k$  clases posibles, basándose en un conjunto de variables predictoras (Hastie et al., 2009). El modelo define una función discriminante para cada clase  $k$  de la forma:

$$P(y_i = k | x_i) = \frac{\exp(w_k^\top x_i + b_k)}{\sum_{j=1}^K \exp(w_j^\top x_i + b_j)} \quad (2.6)$$

donde  $w_k$  es el vector de coeficientes correspondiente a la clase  $k$  y  $b_k$  es el intercepto para la clase  $k$ .

El **Perceptrón Multicapa** es una arquitectura de red neuronal artificial de tipo *feedforward* que incorpora una o más capas ocultas con funciones de activación no lineales, lo que le permite modelar relaciones complejas y no lineales entre las variables de entrada y salida (Goodfellow et al., 2016). Esta red neuronal aprende los parámetros (pesos y sesgos) mediante la minimización de una función de pérdida a través del algoritmo de retropropagación del error, usualmente optimizado con métodos de descenso por gradiente (Rumelhart et al., 1986). Un Perceptrón Multicapa con una sola capa oculta se define mediante las siguientes operaciones:

$$z_i = \phi \left( W^{(1)} x_i + b^{(1)} \right) \quad (2.7)$$

$$y^i = \sigma \left( W^{(2)} z_i + b^{(2)} \right) \quad (2.8)$$

donde  $W^{(1)}$  y  $b^{(1)}$  son los pesos y sesgos de la capa oculta,  $W^{(2)}$  y  $b^{(2)}$  son los pesos y sesgos de la capa de salida,  $\phi$  es la función de activación de la capa oculta y  $\sigma$  es la función de activación de la capa de salida.

## Métricas de evaluación en modelos de clasificación

La evaluación de los modelos de aprendizaje es importante ya que nos permite comparar su rendimiento entre ellos y con los ya existentes a través de métricas que estiman si la predicción del modelo iguala o supera el desempeño de un profesional con experiencia. Esta evaluación se realiza a través de métricas derivadas de una matriz de confusión. Esta matriz de confusión tabula la relación entre los resultados predichos y los resultados reales, obteniendo el número de verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos. Los modelos fueron evaluados mediante múltiples métricas de rendimiento, incluyendo exactitud, precisión, sensibilidad y F1 Score (Tabla 2.1), con el fin de proporcionar una visión integral de su desempeño.

Tabla 2.1: Resumen de las métricas de evaluación utilizadas en la validación del modelo

| Métrica      | Definición                                                                  | Fórmula                                                                                               |
|--------------|-----------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------|
| Exactitud    | Proporción de predicciones correctas entre todas las observaciones          | $\frac{VP + VN}{VP + VN + FP + FN}$                                                                   |
| Precisión    | Proporción de verdaderos positivos sobre el total de predicciones positivas | $\frac{VP}{VP + FP}$                                                                                  |
| Sensibilidad | Proporción de verdaderos positivos sobre el total de casos positivos reales | $\frac{VP}{VP + FN}$                                                                                  |
| F1 Score     | Media armónica entre la precisión y la sensibilidad                         | $\frac{2 \times \text{Precisión} \times \text{Sensibilidad}}{\text{Precisión} + \text{Sensibilidad}}$ |

*Nota:* VP: Verdaderos positivos. FP: Falsos positivos. FN: Falsos negativos. VN: Verdaderos negativos.

Además, los modelos se evaluaron con los valores de área bajo la curva ROC (Características Operativas del Receptor) - o también conocida por sus siglas en inglés, ROC-AUC - y área bajo la curva de Precisión - Sensibilidad - por sus siglas en inglés, PR-AUC.

El área bajo la curva ROC es una métrica que muestra en un gráfico los verdaderos positivos versus la proporción de falsos negativos de todos los umbrales definidos. La puntuación puede ser de 0 a 1, con un valor de 1 que indica un rendimiento perfecto y un valor de 0.5 que indica poder discriminatorio aleatorio o nulo (Rashidi et al., 2023). Por otro lado, la curva de precisión-sensibilidad es una herramienta que evalúa el rendimiento de los clasificadores representando el valor predictivo positivo frente a la tasa de verdaderos positivos, siendo una alternativa para la curva ROC en conjuntos de datos desbalanceados (Brodersen et al., 2010; Davis y Goadrich, 2006).

## Selección de características

La selección de características es un proceso en el aprendizaje automático que consiste en elegir un subconjunto de características relevantes a partir de un conjunto más amplio, con el fin de mejorar el rendimiento y la eficiencia del modelo (Rawat, 2021; Srivastava et al., 2013). Su objetivo es eliminar características irrelevantes, redundantes y ruidosas, reduciendo así la dimensionalidad y la complejidad computacional (Büyükeğeci y Okur, 2023). La importancia de este proceso radica en su capacidad para incrementar la precisión del modelo, simplificar su estructura, mejorar la calidad y comprensión de los datos, y aumentar la eficiencia computacional. La selección de características es aplicable tanto a contextos de aprendizaje supervisado como no supervisado, y puede potenciar de manera significativa el desempeño de diversos algoritmos de aprendizaje automático (Srivastava et al., 2013).

Existen tres enfoques principales para la selección de características: **(i)** métodos de filtrado (*filter methods*), que evalúan las características de manera independiente al modelo predictivo; **(ii)** métodos de envoltura (*wrapper methods*), que utilizan el rendimiento de un modelo específico para seleccionar subconjuntos óptimos; y **(iii)** métodos embebidos (*embedded methods*), que integran la selección de características dentro del proceso de entrenamiento del modelo (Kohavi y John, 1997). En esta sección se describen tres técnicas representativas utilizadas en el trabajo de investigación: SelectKBest con prueba de chi-cuadrada, la Eliminación Recursiva de Características (RFE) y el Análisis Factorial.

### SelectKBest con prueba de Chi-cuadrada

La técnica SelectKBest basada en la prueba de chi-cuadrada es un método de filtrado comúnmente utilizado cuando tanto las características como la variable objetivo son categóricas. Este procedimiento evalúa la dependencia estadística entre cada característica y la variable a predecir, seleccionando las  $k$  variables que presentan los mayores valores de chi-cuadrada, los cuales indican una mayor asociación con la variable objetivo.

La estadística utilizada se define como:

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (2.9)$$

donde  $O_{ij}$  representa la frecuencia observada en la celda  $i, j$  de la tabla de contingencia, y  $E_{ij}$  la frecuencia esperada bajo la hipótesis de independencia. Esta técnica es eficiente y fácilmente interpretable, aunque presenta limitaciones importantes, como su aplicabilidad exclusiva a variables discretas y su incapacidad para capturar interacciones entre características (Huan Liu y Lei Yu, 2005).

## Eliminación Recursiva de Características (RFE)

La Eliminación Recursiva de Características (*Recursive Feature Elimination*, RFE) es un método de envoltura que selecciona variables en función de su contribución al rendimiento de un modelo predictivo. El procedimiento de RFE comienza con el entrenamiento de un modelo supervisado sobre el conjunto completo de características. A continuación, se calcula un ranking de importancia para cada variable —comúnmente derivado de los coeficientes del modelo o de medidas como la importancia de los atributos en árboles de decisión—. La característica con menor puntuación se elimina, y el modelo se reentrena con el subconjunto restante (Priyatno y Widiyaningtyas, 2024). Este proceso se repite hasta alcanzar un número predefinido de variables o hasta que se cumpla un criterio de detención.

Este enfoque es particularmente compatible con algoritmos que permiten una interpretación directa de la importancia de cada variable, como:

- Máquinas de vectores de soporte (SVM) con kernel lineal
- Regresión logística
- Árboles de decisión y bosques aleatorios

Este enfoque tiene como ventaja principal su capacidad para considerar interacciones y relaciones no lineales entre variables. Sin embargo, su costo computacional es elevado, especialmente en conjuntos de datos de alta dimensionalidad (Guyon et al., 2002).

## Análisis Factorial

El Análisis Factorial es una técnica estadística multivariante utilizada para identificar estructuras latentes en un conjunto de variables observadas. A través de la extracción de factores, busca explicar la variabilidad de un conjunto de datos mediante un número reducido de dimensiones subyacentes. Aunque no constituye un método de selección de características en sentido estricto, sí permite reducir la dimensionalidad conservando la información esencial.

Según De la Fuente Fernández, 2011, el Análisis Factorial (AF) contempla dos métodos distintos:

- **Análisis que incluye toda la varianza, tanto común como no común:** El método más común es el Análisis de Componentes Principales (ACP) que utiliza la matriz de correlaciones donde los valores fuera de la diagonal principal, elevados al cuadrado, muestran la varianza compartida entre las variables correspondientes, y se incluye el valor uno en la diagonal principal.

- **Análisis que se centra únicamente en la varianza común:** El método tradicional es el Análisis de Factores Comunes (AFC) que reemplaza los unos de la diagonal por estimaciones de la varianza que cada ítem tiene en común con los demás (denominado comunalidad).

En particular, el Análisis Factorial Exploratorio no asume cuántos factores se deben extraer, permitiendo que el número se defina durante el análisis. Uno de los elementos más importantes en este análisis son las cargas factoriales, que indican la fuerza de la relación entre cada variable observada y cada factor latente. Generalmente, las cargas factoriales se interpretan según su magnitud: valores cercanos a 1 o -1 denotan una relación fuerte, mientras que valores cercanos a 0 indican una relación débil. Se considera que cargas factoriales por encima de 0.4 son significativas para determinar la asociación entre variables y factores (Yaminfirooz y Gholinia-Ahangar, 2014).

Los pasos a seguir para realizar un análisis factorial sugeridos por Watkins, 2020 se encuentran en la Figura 4.4.

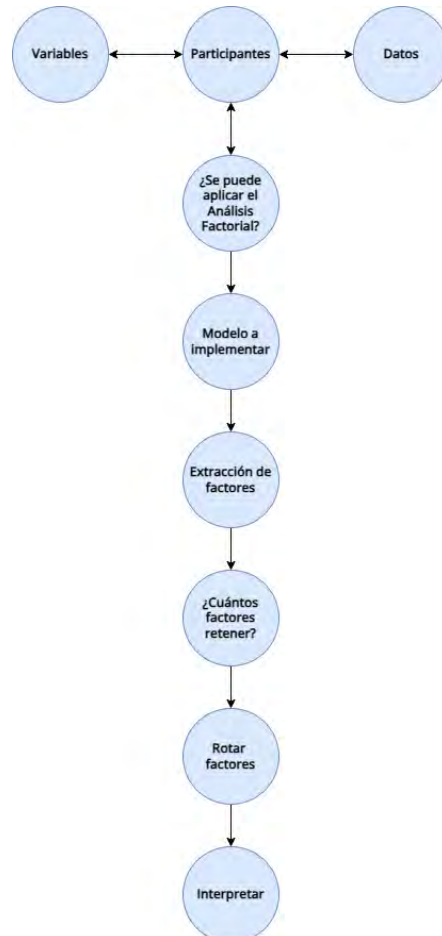


Figura 2.1: Diagrama de flujo para realizar un Análisis Factorial Exploratorio

*Nota:* Adaptado de Watkins, 2020.

Se recomienda realizar un análisis factorial confirmatorio después de un análisis factorial exploratorio para validar la estructura dimensional de una medida (Floyd y Widaman, 1995). El análisis factorial confirmatorio se utiliza para evaluar una teoría existente y examina si dicho modelo se ajusta adecuadamente a los datos (Bandalos, 1996). El autor Matsunaga, 2010 propone que un modelo de análisis factorial confirmatorio sea evaluado considerando su ajuste exacto (es decir, el valor de  $\chi^2$ ), el RMSEA, uno de los índices de ajuste incremental (como el CFI, TLI o RNI) y el SRMR. Un buen ajuste del modelo se puede considerar cuando el valor calculado de  $\chi^2$  no es estadísticamente significativo, el RMSEA es menor a 0.06, el CFI/TLI/RNI es mayor a 0.95, y el SRMR es menor a 0.10.

## Optimización de hiperparámetros

La optimización de hiperparámetros es un aspecto importante del aprendizaje automático, cuyo objetivo principal es identificar los valores óptimos de los parámetros que deben fijarse antes del proceso de entrenamiento del modelo (Monica y Agrawal, 2024). A diferencia de los parámetros que se ajustan durante el entrenamiento, los hiperparámetros controlan aspectos clave del comportamiento del algoritmo, tales como la tasa de aprendizaje, la profundidad de los árboles de decisión o el número de neuronas ocultas.

El ajuste adecuado de estos hiperparámetros tiene un impacto directo en diversos aspectos del modelo, entre ellos: el rendimiento predictivo, la tasa de convergencia durante el entrenamiento y la capacidad de generalización a datos no vistos (Lytra, 2021). Una elección subóptima de hiperparámetros puede resultar en modelos ineficaces, sobreajustados o con pobre desempeño, aun cuando el algoritmo subyacente sea sólido.

Para llevar a cabo esta tarea, se han desarrollado múltiples técnicas, entre las que se destacan:

- **Búsqueda en malla** (*Grid Search*): Este método consiste en especificar un conjunto de valores posibles para cada hiperparámetro y, a continuación, realizar una búsqueda exhaustiva en una rejilla multidimensional para evaluar el rendimiento del modelo para cada combinación de hiperparámetros (Andonie, 2019). La secuencia de acciones que se llevan a cabo en esta técnica se describen en el Algoritmo 1.
- **Búsqueda aleatoria** (*Random Search*): Este enfoque selecciona aleatoriamente un número predefinido de muestras dentro de los límites superior e inferior como valores candidatos para los hiperparámetros (Bergstra y Bengio, 2012). Estos candidatos se entrenan hasta que se agotan las configuraciones al azar. La idea de esta técnica es encontrar hiperparámetros óptimos o una aproximación si el espacio de búsqueda es lo suficientemente amplio. La secuencia de acciones que se llevan a cabo en esta técnica se describen en el Algoritmo 2.
- **Optimización bayesiana**: modela la relación entre hiperparámetros y desempeño del modelo mediante procesos gaussianos, lo que permite tomar decisiones más informadas al seleccionar combinaciones prometedoras (Wu et al., 2019). El objetivo general de la optimización bayesiana es encontrar la configuración óptima  $x^*$  que maximiza el valor de la función  $f(x)$  utilizando una función objetivo desconocida  $f$  que recibe un valor

de entrada  $x$ , donde la función objetivo real es desconocida. La secuencia de acciones que se llevan a cabo en esta técnica se describen en el Algoritmo 3.

---

**Algoritmo 1:** Búsqueda en Rejilla

---

**Data:**

Espacio de búsqueda de hiperparámetros  $\Theta$

Modelo de aprendizaje automático  $M$

Conjunto de datos de entrenamiento  $D_{train}$

**Result:** Mejor configuración de hiperparámetro  $\theta^*$

$\theta^* \leftarrow \emptyset$ ;

$best\_performance \leftarrow -\infty$ ;

**for**  $\theta$  **in**  $\Theta$  **do**

$M_\theta \leftarrow$  Inicializar el modelo  $M$  con los hiperparámetros  $\theta$ ;

    Entrenar  $M_\theta$  con  $D_{train}$ ;

$performance \leftarrow$  Evaluar el desempeño de  $M_\theta$  con el conjunto de prueba;

**if**  $performance > best\_performance$  **then**

$best\_performance \leftarrow performance$ ;

$\theta^* \leftarrow \theta$ ;

---



---

**Algoritmo 2:** Búsqueda Aleatoria

---

**Data:**

Espacio de búsqueda de hiperparámetros  $\Theta$

Modelo de aprendizaje automático  $M$

Conjunto de datos de entrenamiento  $D_{train}$

Número de iteraciones  $n$

**Result:** Mejor configuración de hiperparámetro  $\theta^*$

$\theta^* \leftarrow \emptyset$ ;

$best\_performance \leftarrow -\infty$ ;

**for**  $i \leftarrow 1$  **to**  $n$  **do**

$\theta \leftarrow$  Muestreo aleatorio de  $\Theta$ ;

$M_\theta \leftarrow$  Inicializar el modelo  $M$  con los hiperparámetros  $\theta$ ;

    Entrenar  $M_\theta$  con  $D_{train}$ ;

$performance \leftarrow$  Evaluar el desempeño de  $M_\theta$  con el conjunto de prueba;

**if**  $performance > best\_performance$  **then**

$best\_performance \leftarrow performance$ ;

$\theta^* \leftarrow \theta$ ;

---

---

**Algoritmo 3:** Optimización Bayesiana

---

**Data:**

Proceso gaussiano a priori sobre  $f$ ;

Observar  $f$  en  $n_0$  puntos de acuerdo con un diseño experimental inicial de llenado de espacio.;

Inicializar  $n = n_0$ .

**while**  $n \leq N$  **do**

    Actualizar la distribución de probabilidad posterior sobre  $f$  utilizando todos los datos disponibles.;

    Sea  $x_n$  un maximizador de la función de adquisición sobre  $x$ , donde la función de adquisición se calcula utilizando la distribución posterior actual.;

    Obsérvese que  $y_n = f(x_n)$ .;

    Incrementar  $n$ ;

**Result:** ya sea el punto evaluado con la mayor  $f(x)$ , o el punto con la mayor media posterior.

---

## 2.2. Trastorno de Ansiedad Generalizada

En la “Clasificación Internacional de Enfermedades, undécima revisión (CIE-11), Organización Mundial de la Salud (OMS)”, 2024 se define el trastorno de ansiedad generalizada como una preocupación excesiva por diversos eventos cotidianos relacionados con la familia, salud, finanzas, escuela o trabajo, persistente durante meses. Las personas que padecen este trastorno suelen reaccionar exageradamente ante situaciones percibidas como amenazantes, enfrentando dificultades para manejar las emociones negativas derivadas. La preocupación, lejos de ser beneficiosa, actúa como mecanismo compensatorio que perpetúa las emociones negativas, mermando la capacidad de sentir emociones positivas. Esta preocupación sobre la preocupación es especialmente prevalente en el trastorno.

### Diagnóstico

El Manual Diagnóstico y Estadístico de los Trastornos Mentales, Quinta Edición (American Psychiatric Association, 2013), establece que:

- Se debe sentir anticipación aprensiva mayoritariamente durante al menos seis meses.
- Es difícil para el individuo controlar la constante preocupación.
- Debe haber tres o más de los siguientes síntomas más días de los que no en los últimos seis meses: inquietud, fatiga, dificultad para concentrarse, irritabilidad, tensión muscular y problemas de sueño.
- Los síntomas causan malestar significativo o deterioro en áreas importantes de la vida del individuo.
- La condición no se atribuye a sustancias o a condiciones médicas.
- No se explica mejor por otro trastorno mental.

## Prevalencia, desarrollo y curso

De acuerdo a lo reportado en el proyecto de Encovid-19 en México (Gaitán-Rossi et al., 2021), la prevalencia del trastorno de ansiedad generalizada fue del 31 % al 33 % entre abril y junio de 2020. Las mujeres tienen el doble de riesgo que los hombres de experimentar ansiedad generalizada. Este trastorno suele diagnosticarse alrededor de los 20 años, con una edad media de inicio de aproximadamente 30 años (American Psychiatric Association, 2013). Su curso es generalmente crónico, con frecuentes recaídas.

## Factores de riesgo

- **Temperamentales:** La inhibición conductual, afectividad negativa (neuroticismo) y evitación del daño están vinculadas al trastorno.
- **Ambientales:** Las adversidades en la infancia y sobre protección parental se han relacionado, aunque no se han identificado factores específicos esenciales para el diagnóstico.
- **Genéticos y Fisiológicos:** Un tercio del riesgo es genético, superponiéndose con el neuroticismo y compartido con otros trastornos del ánimo, principalmente el trastorno depresivo mayor.

## 2.3. Estrategias de historia de vida

Las estrategias de historia de vida constituyen formas complejas de adaptación que han evolucionado en respuesta a las exigencias del entorno de los seres humanos. Estas estrategias comprenden una combinación de características conductuales y fisiológicas orientadas a enfrentar diversas presiones ambientales (Kahl et al., 2022). Para explicar cómo las condiciones del entorno y las experiencias durante la infancia influyen en el desarrollo de dichas estrategias, se ha propuesto un continuo denominado rápido-lento (Figura 2.2). Este continuo sugiere que los individuos ajustan su comportamiento en función del contexto ambiental con el fin de maximizar su éxito reproductivo.

El factor  $K$  sintetiza este constructo en un índice unificado que puede evaluarse a través de variables relacionadas con el desarrollo, la conducta reproductiva, el autocontrol, la vinculación social y la salud mental (Figueredo et al., 2004). Este factor se basa en la teoría ecológica de la selección  $r/K$ , la cual plantea que los organismos se sitúan en un espectro que va desde las estrategias  $r$  —caracterizadas por una reproducción rápida y una baja inversión parental— hasta las estrategias  $K$ , definidas por una reproducción más lenta, una mayor inversión en cada descendiente y una orientación hacia el largo plazo (Giudice et al., 2015).

La evidencia sugiere que los eventos adversos en etapas tempranas de la vida predisponen a los individuos a desarrollar estrategias más rápidas, mientras que los entornos estables favorecen estrategias más lentas, enfocadas en el bienestar a largo plazo (Chua et al., 2017). Estas variaciones permiten identificar perfiles de riesgo para distintos trastornos mentales desde una perspectiva evolucionista (Kahl et al., 2022).

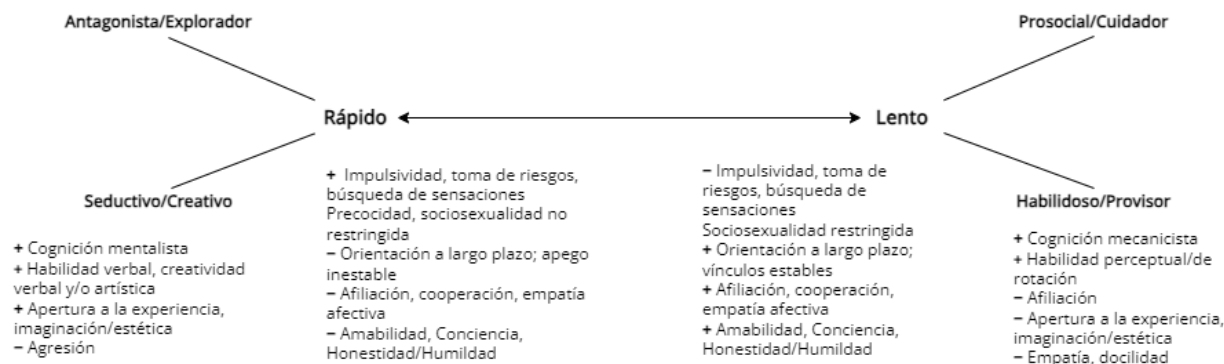


Figura 2.2: Continuo rápido-lento en estrategias de historia de vida y perfiles

*Nota:* Adaptado de Del Giudice, 2018.

## Modelo FSD (Fast-Slow-Defense)

Del Giudice (2018) propuso el modelo FSD como marco teórico para comprender la psicopatología evolucionista, clasificando los trastornos mentales según las estrategias de historia de vida en un continuo rápido-lento y una categoría de trastornos de activación de defensa. Los trastornos del espectro rápido, asociados con estrategias de historia de vida rápida, implican impulsividad y comportamientos antisociales, mientras que los trastornos del espectro lento, vinculados a estrategias de historia de vida lenta, conllevan una autorregulación exagerada y menor motivación sexual (Del Giudice, 2018). La categoría de trastornos de defensa se asocia con respuestas a amenazas percibidas, enfocadas en la supervivencia. La Figura 2.3 muestra la estructura general de este modelo.



Figura 2.3: Estructura de clasificación del modelo FSD.

*Nota:* Adaptado de Del Giudice, 2018.

### **Trastornos del espectro rápido (Tipo F)**

El espectro rápido se define por estrategias de historia de vida enfocadas en la reproducción temprana, respondiendo a ambientes percibidos como adversos. Las personas con estas estrategias muestran mayor impulsividad, toman más riesgos y tienen menor autorregulación. Este espectro incluye trastornos como el Trastorno Antisocial de la Personalidad, Trastorno por Déficit de Atención e Hiperactividad, y Trastorno Bipolar, todos asociados a impulsividad y, en ocasiones, conductas antisociales. Los factores de riesgo incluyen adversidades en la infancia, entornos familiares inestables, y exposición a violencia. La incertidumbre ambiental y ciertos factores genéticos también contribuyen al desarrollo de estrategias de vida rápida.

### **Trastornos del espectro lento (Tipo S)**

El espectro lento se define por adoptar estrategias de vida cuidadosas y meditadas, orientadas hacia el éxito reproductivo a largo plazo. Individuos con enfoques de vida lentos exhiben alta autorregulación y baja impulsividad, ponderando detenidamente las repercusiones futuras de sus actos. Los entornos estables y predecibles, que promueven la inversión a largo plazo y la precaución, pueden fomentar el desarrollo de estas estrategias. No obstante, una preocupación excesiva por la prudencia y la planificación podría derivar en trastornos marcados por inhibición y ansiedad anticipatoria. Dentro de este espectro se encuentran trastornos como el Trastorno del Espectro Autista, el Trastorno Bipolar, Trastornos Alimenticios y el Trastorno Obsesivo Compulsivo.

### **Trastornos de activación de defensa (Tipo D)**

Estos trastornos representan respuestas adaptativas ante el estrés agudo o crónico, preparando al organismo para acciones de "lucha o huida". Factores como la exposición a traumas, abuso, estresores vitales y predisposición genética hacia la ansiedad o sensibilidad al estrés pueden incrementar el riesgo de estos trastornos. Además, estrategias de afrontamiento deficientes y escaso apoyo social potencian dicho riesgo. El neuroticismo elevado es clave en los trastornos de tipo defensivo. Los individuos altamente neuróticos suelen tener una baja tolerancia a la incertidumbre, actitudes pesimistas, susceptibilidad a la crítica y propensión a la preocupación y rumiación. El espectro defensivo abarca trastornos como el Trastorno de Estrés Postraumático, trastornos de ansiedad generalizada, fobias específicas y el trastorno de pánico, caracterizados por una respuesta de defensa exagerada ante amenazas.

# Capítulo 3

## Estado del arte

En esta sección, se presenta el estado del arte respecto al uso de técnicas de aprendizaje automático en el ámbito de la salud mental. Se abordan los trastornos mentales identificados y los algoritmos empleados en estos estudios. Con el fin de organizar el análisis de estos casos de estudio, se han establecido dos categorías según su nivel de generalidad. Estos criterios se explicarán en detalle en las siguientes secciones.

### 3.1. Clasificación de los trabajos del estado del arte

Las categorías utilizadas para clasificar las técnicas de aprendizaje automático implementadas en los casos de estudio de salud mental son:

1. **Predicción del trastorno de ansiedad generalizada:** es una condición en la que las personas experimentan preocupaciones excesivas e inapropiadas que son persistentes y no están limitadas a circunstancias específicas.
2. **Predicción de ansiedad, estrés o depresión:** la relación entre la ansiedad, el estrés y la depresión es compleja y polifacética con causas subyacentes compartidas, síntomas superpuestos y desencadenantes comunes.

### 3.2. Predicción de los trastornos de ansiedad

En el contexto de los trastornos de ansiedad, la aplicación de modelos de aprendizaje automático para la detección y predicción ha resultado prometedora. La investigación en este campo ha implementado diversos enfoques, utilizando modelos supervisados tradicionales y combinándolos con técnicas de selección de características, técnicas de balanceo de clases e integrando fuentes de datos mixtas, para optimizar la precisión de las predicciones.

## **Predictive Modeling of Anxiety Levels in Bangladeshi University Students: A Voting-Based Approach with LIME and SHAP Explanations (Islam et al., 2024)**

En este artículo se desarrolló un modelo de votación constituido por tres algoritmos de aprendizaje automático para identificar los niveles de ansiedad en estudiantes universitarios de Bangladesh. Se recolectaron 569 registros donde se incluyeron variables sociodemográficas, preguntas de la escala DASS-21 y aspectos de la personalidad. Se aplicó preprocesamiento, balanceo de clases con SMOTE (Técnica de Sobre-muestreo Sintético para la Clase Minoritaria) y selección de características usando RFE (Eliminación Recursiva de Características). El modelo de votación, conformado por el algoritmo de Máquina de Soporte Vectorial, Random Forest y Naive Bayes, alcanzó una exactitud del 96.5 %, F1-score del 97 %, sensibilidad del 96 % y precisión del 97 % al emplear 20 características seleccionadas. Las variables más relevantes para la predicción, según los análisis de LIME y SHAP, fueron “temblor en las manos”, “preocupación por situaciones embarazosas”, “sensación de pánico” y “latidos cardíacos acelerados”.

## **SADXAI: Predicting social anxiety disorder using multiple interpretable artificial intelligence techniques (Chadaga et al., 2024)**

En este artículo se presentó un sistema de apoyo clínico para predecir el trastorno de ansiedad social (SAD) utilizando algoritmos de aprendizaje automático. Se utilizó un conjunto de datos de 214 pacientes, que incluyó 30 variables demográficas, físicas y emocionales, basado en el test DSM-5. Tras un análisis estadístico con Jamovi, se seleccionaron las ocho variables más relevantes mediante información mutua. Los modelos de Regresión Logística y AdaBoost obtuvieron el mejor desempeño, con una exactitud del 88 % y un AUC de 0.94. Se aplicaron cuatro técnicas de inteligencia artificial explicable (XAI) donde se identificaron como más importantes las variables de “miedo a hablar en público”, “entrar a fiestas”, “sentarse con extraños”, “recibir atención” y “conocer gente nueva”.

## **Analysis of Risk Factors and Diagnosis for Anxiety Disorder in Older People with the Aid of Artificial Intelligence: Observational Study (Wang et al., 2023)**

En esta investigación se desarrollaron modelos de aprendizaje automático para diagnosticar trastorno de ansiedad e identificar los factores de riesgo. Se utilizó un conjunto de datos con 5,186 registros de adultos mayores que incluía variables sociodemográficas, clínicas, bioquímicas, nutricionales y de estilo de vida. Se seleccionaron 83 variables predictoras y se aplicaron técnicas de sobremuestreo (SMOTE y ADASYN). Los modelos de Random Forest y Máquina de Soporte Vectorial obtuvieron los mejores resultados en la exactitud con 85.4 % y 83.4 % respectivamente al aplicar SMOTE. Entre las variables más relevantes se identificaron el sexo femenino, vivir solo, estado civil separado/divorciado, enfermedades crónicas (hipertensión, hiperlipidemia), antecedentes familiares de enfermedades, bajo estatus socioeconómico, consumo de alcohol, tabaquismo y preocupaciones sobre la memoria.

## **A Machine Learning Study to Predict Anxiety on Campuses in Lebanon Stacking ensemble learning model for predict anxiety level in university students using balancing methods (Mahalingam et al., 2023)**

En esta investigación se desarrolló un modelo de aprendizaje automático para predecir el nivel de ansiedad, determinada a partir del Inventario de Ansiedad de Beck (BAI), en estudiantes universitarios de Líbano. Se recolectaron datos de 329 estudiantes, incluyendo variables como salud autopercebida, edad, nivel educativo, hábitos de sueño, consumo de sustancias, actividad física y características sociodemográficas. El modelo con mejor rendimiento en términos de AUC fue la red neuronal Perceptrón Multicapa (MLP), con un valor de 80.70 %, mientras que el modelo con mejor exactitud fue SVM (69.70 %). La variable más importante fue la salud autopercebida, seguida por la edad, según el análisis de importancia de características con Random Forest.

## **Investigating anxiety levels in the Quebec university community during the COVID-19 pandemic using machine learning and data exploration techniques (Maitre et al., 2023)**

En este artículo se desarrolló un modelo de aprendizaje automático para clasificar los niveles de ansiedad en la comunidad universitaria de Quebec. Se recolectaron datos longitudinales a través de encuestas en línea en cuatro fases entre 2020 y 2021, con una muestra total de 3,878 participantes. Se utilizó la escala GAD-7 como variable objetivo para medir los niveles de ansiedad, y se incluyeron 23 variables predictoras sociodemográficas, de salud mental, espiritualidad, estrés diario, entre otras. El modelo CatBoost obtuvo mejor desempeño con un coeficiente de determinación  $R^2 = 0.5656$  y un MAE de 2.995. Los resultados del análisis SHAP mostraron que las variables más influyentes fueron el cambio en el estrés diario, el cambio en la salud mental, los trastornos del sueño, la calidad de vida espiritual (SRPB) y el género.

## **Stacking ensemble learning model for predict anxiety level in university students using balancing methods (Daza et al., 2023)**

El objetivo de esta investigación fue desarrollar un modelo de aprendizaje automático basado en la técnica de Stacking para predecir los niveles de ansiedad en estudiantes universitarios. Se utilizó un conjunto de datos de 284 estudiantes de una universidad pública en Perú, evaluados mediante el cuestionario GAD-7. Se aplicó preprocesamiento, codificación y balanceo de clases con sobremuestreo, seguido de validación cruzada con 10 repeticiones. El modelo con mejor desempeño fue Stacking 4A (KNN como meta-modelo), que logró una exactitud del 97.83 %, sensibilidad del 98.44 %, F1-score del 97.88 %, MCC de 97.08 % y especificidad del 99.32 %, superando a todos los modelos individuales.

## **Predicting the level of generalized anxiety disorder of the coronavirus pandemic among college age students using artificial intelligence technology (Alharthi, 2020)**

En este artículo se desarrolló un modelo de aprendizaje automático para predecir el nivel de trastorno de ansiedad generalizada (GAD) en estudiantes universitarios durante la pandemia de COVID-19. Se recolectaron 917 encuestas aplicadas a estudiantes de una universidad en Arabia Saudita, que incluyeron información general, estresores relacionados con la pandemia y el cuestionario GAD-7. El nivel de ansiedad fue clasificado en cuatro categorías: ninguno, leve, moderado y severo. Se entrenaron y evaluaron varios clasificadores de aprendizaje automático utilizando partición 70/30 y validación cruzada. El mejor modelo fue el Perceptrón Multicapa, obteniendo una exactitud de 75.4 %, F1-score de 74.9 % y un valor de AUC de 0.936. El análisis de importancia de variables identificó como principales predictores: el género, el apoyo familiar/social y la estabilidad del ingreso familiar.

### **3.3. Predicción de ansiedad, estrés o depresión**

A continuación se presentan los trabajos que han aplicado modelos de aprendizaje automático para realizar la detección y predicción de trastornos de ansiedad en conjunto con niveles de estrés, depresión u otros trastornos del ánimo.

## **Evaluating Machine Learning Stability in Predicting Depression and Anxiety Amidst Subjective Response Errors (Ku y Min, 2024)**

En este artículo se evaluó el desempeño de cinco algoritmos de aprendizaje automático (CNN, Random Forest, XGBoost, Regresión Logística y Naive Bayes) en la predicción de depresión (MDD) y trastorno de ansiedad generalizada (GAD). Se utilizó un conjunto de datos de 4,184 estudiantes universitarios franceses con variables sociodemográficas, biométricas y de estilo de vida, junto con diagnósticos clínicos. Se evaluaron los modelos con validación cruzada de 10 pliegues, usando métricas como AUC, exactitud, F1, kappa de Cohen y precisión. El modelo Random Forest obtuvo los mejores valores de AUC en los conjuntos de datos de depresión y ansiedad (0.58 en ambos). Las variables más importantes para predecir MDD incluyeron presión diastólica, estado de vacunación y entorno parental, mientras que para GAD destacaron la autopercepción de necesidad de control médico, género y estatura.

## **Construction of a machine learning-based risk prediction model for depression in middleaged and elderly hypertensive people in China: a longitudinal study (Ai et al., 2024)**

En este artículo se construyó un modelo predictivo basado en aprendizaje automático para estimar el riesgo de depresión en adultos mayores con hipertensión en China. Se analizaron datos longitudinales de 2,733 participantes, aplicando LASSO para seleccionar las 10 variables más relevantes. Se entrenaron tres modelos: regresión logística, Random Forest y

XGBoost, donde el modelo de regresión logística mostró el mejor desempeño con un AUC de 0.712, sensibilidad de 0.789 y valor predictivo negativo de 0.716, superando a los otros modelos. Las variables más importantes para la predicción fueron: edad, número de enfermedades crónicas, dolor, satisfacción con la vida, salud autopercebida, memoria autopercebida, género y tipo de registro domiciliario.

## **Enhancing the Diagnosis of Depression and Anxiety Through Explainable Machine Learning Methods (Marey et al., 2024)**

El objetivo de esta investigación fue desarrollar un modelo para la detección de depresión y ansiedad. Se utilizó un conjunto de datos digital con 783 muestras de jóvenes universitarios que incluyó variables como edad, género, índice de masa corporal, puntuaciones en los cuestionarios PHQ y GAD, y niveles de somnolencia. Se aplicó el algoritmo de selección de características RFE, seguido por la evaluación de cinco clasificadores: SVM, árbol de decisión, Random Forest, regresión logística y XGBoost. Tras la selección de características y la optimización de modelos con Grid Search, los mejores resultados para ansiedad se obtuvieron con XGBoost (exactitud = 98 %, F1-score = 99 %) y para depresión con regresión logística y Random Forest (exactitud = 95 %, F1-score = 97 %).

## **Machine learning models predict the emergence of depression in Argentinean college students during periods of COVID-19 quarantine (López Steinmetz et al., 2024)**

En este artículo se desarrollaron modelos de aprendizaje automático para predecir la aparición de depresión en estudiantes universitarios argentinos durante la cuarentena por COVID-19, utilizando un enfoque longitudinal. El estudio analizó datos de 1,492 estudiantes en dos momentos (T1 y T2), incorporando variables como puntuaciones de depresión y ansiedad, historial de trastornos mentales y conducta suicida, sexo, edad y subperiodos de cuarentena. Se implementaron modelos de clasificación (regresión logística, Random Forest y SVM) y regresión (ridge regression, Random Forest y SVR), evaluados mediante métricas como AUPRC, F1-score,  $R^2$  y MAE. Los mejores resultados se obtuvieron con SVM y regresión logística (AUC-PR = 0.76), así como con SVR y ridge regression ( $R^2 = 0.56$ ).

## **Comparison of the Performance of Machine Learning-based Algorithms for Predicting Depression and Anxiety among University Students in Bangladesh (Rahman et al., 2023)**

En esta investigación se desarrollaron modelos de aprendizaje automático para predecir los niveles de depresión y ansiedad en estudiantes universitarios de Bangladesh durante el COVID-19. Se aplicó una encuesta en línea a 2,121 estudiantes de universidades públicas y privadas, que incluía variables sociodemográficas y los cuestionarios PHQ-9 y GAD-7 para evaluar síntomas de depresión y ansiedad. Se implementaron los modelos Regresión Logística, Random Forest (RF), Máquinas de Soporte Vectorial (SVM), Análisis Discriminante Lineal (LDA), K Vecinos más Cercanos (KNN) y Naive Bayes, aplicando validación cruzada de 10

pliegues. Los mejores resultados para predecir depresión se obtuvieron con RF (exactitud = 88.66 %, sensibilidad = 68.79 %, especificidad = 95.8 %), mientras que para ansiedad el mejor modelo fue SVM (exactitud = 91.49 %, sensibilidad = 67.77 %, especificidad = 98.53 %). Se identificó mayor prevalencia de síntomas severos en mujeres, estudiantes en los primeros años y con menor rendimiento académico.

### **Machine learning-based diagnosis support system for differentiating between clinical anxiety and depression disorders (Richter et al., 2021)**

En esta investigación se desarrolló un sistema de apoyo diagnóstico basado en aprendizaje automático para diferenciar entre trastornos clínicos de ansiedad y depresión. El estudio incluyó a 101 participantes (grupos con diagnóstico clínico de ansiedad, depresión, mixto y un grupo control sin diagnóstico), quienes completaron una conjunto de seis tareas cognitivas que evaluaban sesgos en atención, expectativas, memoria, interpretación y control ejecutivo. Se entrenó un modelo de clasificación con el algoritmo Random Forest, utilizando validación cruzada. El modelo logró una exactitud del 80 % para clasificar ansiedad y del 66 % para depresión en el modelo comparativo entre ambos trastornos. Las medidas que más contribuyeron a la clasificación fueron los tiempos de reacción en tareas que evaluaban sesgos de expectativa e interpretación.

### **Decoding rumination: A machine learning approach to a transdiagnostic sample of outpatients with anxiety, mood and psychotic disorders (Silveira et al., 2020)**

En este artículo se identificaron patrones de rumiación mediante algoritmos de aprendizaje automático en una muestra de 200 pacientes ambulatorios con trastornos del ánimo, ansiedad y psicosis. Se recolectaron 83 variables clínicas, sociodemográficas y biomarcadores inflamatorios. Se clasificó a los participantes en altos y bajos rumiadores según puntuaciones  $z$  estandarizadas. Se aplicó selección recursiva de características y se entrenaron modelos con SVM lineal, Random Forest y Redes Neuronales Artificiales (ANN), evaluados mediante validación cruzada de 10 pliegues. El modelo SVM logró un AUC de 0.83, sensibilidad del 75.2 % y especificidad del 70.6 %. Las variables más relevantes para predecir rumiación fueron: puntajes en el Penn State Worry Questionnaire, GAD-7, Hamilton Depression Rating Scale, Clinical Global Impression Scale, nivel socioeconómico y diagnóstico actual de trastorno de pánico.

### 3.4. Resumen de los trabajos revisados

Se compararon diferentes trabajos de acuerdo a las categorías de modelos de predicción para trastornos de ansiedad y modelos de predicción para ansiedad, estrés y/o depresión. La Tabla 3.1 muestra la población estudiada en cada trabajo, los tipos de datos utilizados, los modelos empleados y los resultados obtenidos en las métricas de desempeño en los clasificadores. Los aspectos a destacar son los siguientes:

- **Poblaciones:** la mayoría de las poblaciones representadas son del continente asiático o europeo, con una escasa representación de poblaciones latinoamericanas. Esto limita la generalización de los resultados ya que se tienen contextos socioculturales diferentes. Los tamaños muestrales predominantes cuentan con aproximadamente 1,000 participantes o menos, variando según la metodología de recolección empleada (encuestas presenciales o en línea). La mayoría de estudios mantiene un mínimo de 100 participantes para garantizar el entrenamiento adecuado de los modelos predictivos.
- **Instrumentos psicométricos:** la mayoría de los estudios utilizan cuestionarios estandarizados para evaluar los trastornos mentales como el GAD-7, STAI y LSAS para diagnosticar trastornos de ansiedad, BAI para depresión, y otros instrumentos más robustos como DASS-21 y DSM-5 evalúan estrés o somatización.
- **Tipos de datos:** los conjuntos de datos analizados se componen principalmente de información tabular que incluye variables sociodemográficas básicas. Algunos estudios amplían esta base incorporando variables de comportamiento, de personalidad, información clínica y medidas de autopercepción en dominios familiares, sociales y de salud. Se observa que existe una falta de estudios que integren las estrategias de historia de vida como variables predictoras para estos trastornos.
- **Algoritmos de aprendizaje supervisado:** algunos algoritmos implementados incluyen Máquinas de Soporte Vectorial (SVM), Random Forest, XGBoost, regresión logística y redes neuronales, particularmente perceptrones multicapa. La mayoría de los trabajos utiliza al menos un algoritmo de ensamble, siendo XGBoost o Random Forest los más comunes. Respecto al desempeño de los modelos, estos suelen alcanzar porcentajes superiores al 80 % de exactitud en la mayoría de los casos, donde SVM, Random Forest y XGBoost obtienen los mejores resultados en las diferentes métricas evaluadas.

Tabla 3.1: Resumen comparativo de los trabajos del estado del arte

| Trabajo              | Población                                    | Tipo de datos                                                                    | Instrumentos de evaluación | Tipo de modelos                                                                                   | Resultados de desempeño                                                            |
|----------------------|----------------------------------------------|----------------------------------------------------------------------------------|----------------------------|---------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------|
| Islam et al., 2024   | 569 estudiantes universitarios de Bangladesh | Datos sociodemográficas, personalidad, síntomas de ansiedad                      | DASS-21, TIPI              | Naive Bayes, Árbol de decisión, KNN, Random Forest, SVM, Clasificador de votación                 | Mejor modelo: SVM<br>Exactitud = 97.4 %<br>F1-score = 97 %<br>Precisión = 98 %     |
| Chadaga et al., 2024 | 214 personas                                 | Datos sociodemográficos, emocionales, evaluación física, y evaluaciones clínicas | DSM-5, LSAS, SPIN          | Regresión logística, Random Forest, KNN, Árbol de decisión, AdaBoost, CatBoost, LightGBM, XGBoost | Mejor modelo: Regresión logística, AdaBoost<br>Exactitud = 88 %<br>AUC = 0.94.     |
| Ku y Min, 2024       | 4184 estudiantes universitarios franceses    | Datos biomédicos, sociodemográficos y encuestas                                  | No especificado            | XGBoost, Bosque aleatorio, SVM, KNN, Red neuronal                                                 | Mejor modelo: XGBoost<br>Sensibilidad = 70 %<br>Especificidad = 70 %<br>AUC = 0.67 |
| Ai et al., 2024      | 2,733 adultos mayores con hipertensión       | Datos sobre salud, estilo de vida, sociodemográficos                             | CESD-10                    | Regresión logística, Random Forest, XGBoost                                                       | Mejor modelo: Regresión logística<br>Sensibilidad = 0.789<br>AUC = 0.712           |
| Marey et al., 2024   | 783 estudiantes universitarios de Egipto     | Datos sobre ansiedad y depresión                                                 | PHQ-9, GAD-7               | SVM, Árbol de decisión, Random Forest, Regresión logística, XGBoost                               | Mejor modelo: XGBoost<br>Exactitud = 98 %<br>F1-score = 99 %                       |

**Tabla 3.1 Resumen comparativo de los trabajos del estado del arte (continuación)**

| Trabajo                      | Población                                             | Tipo de datos                                                                                                  | Instrumentos de evaluación     | Tipo de modelos                                                                  | Resultados de desempeño                                                                                                  |
|------------------------------|-------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|--------------------------------|----------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------|
| López Steinmetz et al., 2024 | 1,492 estudiantes universitarios argentinos           | Síntomas depresivos, ansiedad, historial de trastornos mentales y suicidio, edad, sexo y periodo de cuarentena | BDI, STAI-T                    | Regresión logística, Random Forest, SVM                                          | Mejor modelo: SVM, Regresión logística<br>AUC = 0.76                                                                     |
| Wang et al., 2023            | 5186 adultos mayores europeos                         | Datos clínicos, sociodemográficos, estilo de vida                                                              | HADS                           | Random Forest, SVM, Gradient Boosting, Perceptrón multicapa, Regresión logística | Mejor modelo: Random Forest<br>Exactitud = 85.4 %<br>Sensibilidad = 67 %<br>Especificidad = 90.3 %                       |
| Mahalingam et al., 2023      | 329 estudiantes universitarios libaneses              | Datos demográficos y de salud                                                                                  | BAI                            | Regresión logística, Perceptrón multicapa, SVM, Random Forest, XGBoost           | Mejor modelo: Perceptrón multicapa<br>Exactitud = 67.5 %<br>AUC = 0.8                                                    |
| Maitre et al., 2023          | 3878 estudiantes y empleados universitarios en Quebec | Datos sociodemográficos, salud, espiritualidad, violencia doméstica                                            | GAD-7, SPS-5, WHOQOL-SRPB-BREF | CatBoost, XGBoost, Random Forest, LASSO, Regresión lineal                        | Mejor modelo: CatBoost<br>$R^2 = 0.5656$<br>MAE = 2.995                                                                  |
| Daza et al., 2023            | 284 estudiantes universitarios peruanos               | Respuestas al cuestionario GAD-7                                                                               | GAD-7                          | KNN, SVM, Árbol de decisión, Gradient Boosting, Stacking                         | Mejor modelo: Stacking KNN<br>Exactitud = 97.83 %<br>Sensibilidad = 98.44 %<br>Precisión = 97.50 %<br>F1-score = 97.88 % |

**Tabla 3.1 Resumen comparativo de los trabajos del estado del arte (continuación)**

| Trabajo               | Población                                                       | Tipo de datos                                                    | Instrumentos de evaluación | Tipo de modelos                                                                   | Resultados de desempeño                                                                                            |
|-----------------------|-----------------------------------------------------------------|------------------------------------------------------------------|----------------------------|-----------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------|
| Rahman et al., 2023   | 2121 estudiantes de licenciatura de universidades de Bangladesh | Datos sociodemográficos                                          | PHQ-9, GAD-7               | Regresión logística, Random Forest, SVM, LDA, KNN, Naive Bayes                    | Mejor modelo: SVM<br>Exactitud = 91.49 %<br>Especificidad = 98.53 %<br>Sensibilidad = 67.77 %                      |
| Richter et al., 2021  | 111 pacientes psiquiátricos y de control                        | Características psicométricas y de comportamiento                | STAI-T, RRS, PSWQ          | Random Forest                                                                     | Especificidad = 76.81 %<br>Sensibilidad = 69.66 %                                                                  |
| Alharthi, 2020        | 917 estudiantes universitarios en Arabia Saudita                | Datos demográficos y estresores por COVID-19                     | GAD-7                      | KNN, Árboles de decisión, SVM, Random Forest, Red neuronal, Naive Bayes, AdaBoost | Mejor modelo: Red neuronal<br>Exactitud = 75 %<br>Precisión = 76.9 %<br>Sensibilidad = 75.4 %<br>F1 score = 74.9 % |
| Silveira et al., 2020 | 220 pacientes brasileños                                        | Datos sociodemográficas, de comportamiento e información clínica | RRS, PSWQ, GAD-7           | SVM, Random Forest, Red neuronal                                                  | Mejor modelo: SVM<br>Sensibilidad = 75.2 %<br>Especificidad = 70.6 %<br>AUC = 0.825                                |

*Nota.* DASS-21, Escala de Depresión, Ansiedad y Estrés; TIPI, Inventario de Personalidad de Diez Ítems; DSM-5, Manual Diagnóstico y Estadístico de los Trastornos Mentales, 5ª edición; LSAS, Escala de Ansiedad Social de Liebowitz; SPIN, Inventario de Fobia Social; CESD-10, Escala del Centro para Estudios Epidemiológicos de la Depresión; PHQ-9, Cuestionario de Salud del Paciente-9; GAD-7, Escala para el Trastorno de Ansiedad Generalizada-7; BDI, Inventario de Depresión de Beck; STAI-T, Inventario de Ansiedad Estado-Rasgo - Versión Rasgo; HADS, Escala de Ansiedad y Depresión Hospitalaria; BAI, Inventario de Ansiedad de Beck; SPS-5, Escala de Fobia Social-5; WHOQOL-SRPB-BREF, Cuestionario de calidad de vida de la Organización Mundial de la Salud (BREF): espiritualidad, religión y creencias personales; RRS, Escala de Respuestas Rumiativas; PSWQ, Inventario de Preocupación de Pensilvania

## Capítulo 4

# Modelo para detección de trastorno de ansiedad generalizada

En esta sección se describe el desarrollo del modelo propuesto para la detección del trastorno de ansiedad generalizada (TAG), utilizando datos recopilados mediante cuestionarios psicométricos y sociodemográficos. Se detalla la metodología implementada, que comprende desde la recolección y preprocesamiento de los datos, hasta la construcción, evaluación y optimización de diversos algoritmos de aprendizaje automático supervisado. Finalmente, se presentan los resultados obtenidos por cada modelo en las pruebas realizadas, haciendo énfasis en el desempeño del mejor clasificador, tanto en su versión sin optimizar como en la versión optimizada.

### 4.1. Introducción a la metodología

La metodología implementada para el desarrollo del modelo de detección del trastorno de ansiedad generalizada (TAG) se organizó en tres fases principales, como se muestra en la Figura 4.1. Estas fases abarcan desde la recolección de datos hasta la selección del mejor modelo.

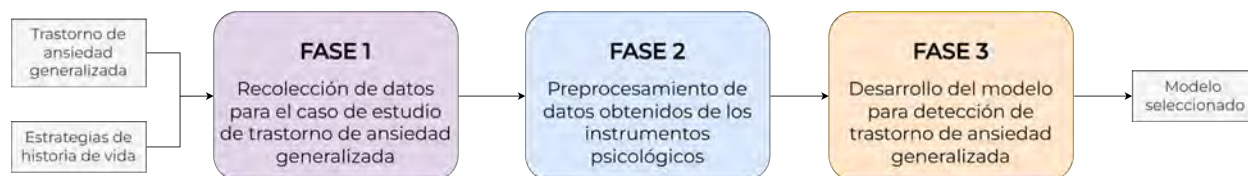


Figura 4.1: Metodología implementada para el desarrollo del modelo de detección de trastorno de ansiedad generalizada

La **Fase 1** estuvo conformada por varios procesos donde, entre ellos, se determinó la población de estudio y tamaño de muestra necesario para la recolección de datos. A continuación, se llevó a cabo una revisión de los instrumentos psicológicos adecuados para evaluar estrategias de historia y trastorno de ansiedad generalizada, seleccionando tres cuestionarios estandarizados y aplicando un cuestionario sociodemográfico. Finalmente, se llevó a cabo

la recolección de datos mediante la aplicación de instrumentos psicológicos previamente seleccionados. Las respuestas obtenidas constituyeron el conjunto de datos principal para las siguientes fases.

La **Fase 2** se centró en el preprocesamiento de los datos. Esta etapa comprendió el análisis exploratorio inicial, la limpieza de datos, el tratamiento de valores faltantes y atípicos, así como la normalización, selección de atributos relevantes y balanceo de clases. Al finalizar este proceso, se dividió el conjunto de datos en dos subconjuntos: uno de *entrenamiento* y otro de *prueba*. El conjunto de entrenamiento permite el ajuste de parámetros del modelo, mientras que el conjunto de prueba proporciona una evaluación independiente del rendimiento del modelo durante su desarrollo.

La **Fase 3** se centró en el desarrollo y optimización del modelo de clasificación. En esta fase se entrenaron distintos algoritmos de aprendizaje automático, se evaluó su desempeño mediante métricas clave y se llevó a cabo la optimización de hiperparámetros empleando una porción específica del conjunto de datos destinada a las pruebas. Estas actividades permitieron seleccionar el modelo con el mejor rendimiento para la detección del TAG, considerando tanto su exactitud, sensibilidad, precisión como su capacidad de distinción entre clases.

## 4.2. Recolección de datos

Se llevó a cabo un estudio con un diseño de encuesta transversal, empleando un enfoque metodológico analítico, retrospectivo y observacional. La población de estudio estuvo conformada por derechohabientes adultos adscritos a la Unidad de Medicina Familiar No. 20 del Instituto Mexicano del Seguro Social (IMSS), ubicada en Cuernavaca, Morelos.

Los criterios de selección fueron **(1)** hombres o mujeres de 18 a 59 años que pudieran leer y escribir; **(2)** sin diagnóstico médico de alguna enfermedad neurológica o enfermedades neurodegenerativas; y **(3)** no haber experimentado eventos adversos en el último año. En este estudio, la variable a detectar fue el nivel de ansiedad, teniendo tres posibles clasificaciones (*control sano*, *ansiedad no TAG* y *TAG*). Se aplicaron cuatro instrumentos autoadministrados, con una duración aproximada de 30 minutos por participante. El número total de personas encuestadas fue de  $n = 244$ , quienes conformaron la fuente de datos para este estudio.

### Instrumentos

Los instrumentos seleccionados para este estudio permitieron obtener información sociodemográfica de los participantes, aspectos de su estrategia de historia de vida y el nivel de preocupación en el momento en que se aplicaron. En total, se obtuvieron 94 variables de los instrumentos. Para llevar a cabo esta investigación, se aplicaron cuatro instrumentos de evaluación:

1. **Cuestionario sociodemográfico.** Incluyó las variables sociodemográficas de sexo, edad en años, estado civil, nivel educativo, nivel socioeconómico, municipio de residencia, personas con quienes habitan, número de hijos y edad en la que tuvo el primer hijo

(si aplicaba), así como la presencia de otras enfermedades crónicas. La vista parcial de este cuestionario puede visualizarse en el Apéndice B.1.

2. **Cuestionario de estrategias de historia de vida.** El instrumento New Short Form of the Arizona Life History Battery (K-SF-42) fue desarrollado por Figueredo en 2017 (Figueredo et al., 2017) y está conformado por 42 ítems de autoevaluación de escala Likert de 7 y 4 categorías, diseñados para evaluar las subescalas de Perspectiva, Planificación y Control, Calidad de Relaciones Parentales, Apoyo Familiar y de Amigos, Vínculo con Parejas Románticas, Altruismo General y Religiosidad. Su propósito es medir cuantitativamente el factor K donde valores altos indican una estrategia de vida lenta, mientras que los valores bajos indican una estrategia de vida rápida. El factor K se obtiene del promedio de la suma de los puntajes de cada aspecto evaluado y se normalizó para que tuviera un valor entre 0 y 1. La vista parcial de este cuestionario puede visualizarse en el Apéndice B.3 y B.4.
3. **Cuestionario de marcadores de estrategias de vida.** Las estrategias de historia de vida también se evaluaron utilizando los marcadores del espectro rápido y lento identificados por Del Giudice (Del Giudice, 2018), que incluyen: amabilidad, conciencia, humildad, honestidad, apertura a la experiencia, precocidad sexual, socio sexualidad, apego romántico, orientación a largo plazo, disgusto sexual, impulsividad, propensión a asumir riesgos, búsqueda de sensaciones placenteras, cognición mentalista, habilidad verbal, habilidad visoespacial y maduración sexual. Estos marcadores se midieron a través de 16 preguntas autoadministradas, utilizando una escala Likert de cinco categorías. La vista parcial de este cuestionario puede visualizarse en el Apéndice B.2.
4. **Cuestionario de trastorno de ansiedad generalizada.** El Inventario de Preocupación de Pensilvania (PSWQ) fue desarrollado por Meyer en 1990 (Meyer et al., 1990) y es un instrumento autoadministrable empleado para medir la gravedad de la preocupación crónica, frecuentemente observada en personas con trastorno de ansiedad generalizada. Se responden 16 enunciados relacionados con la preocupación utilizando una escala Likert de 5 puntos. De la suma de todos los ítems, se obtienen los siguientes resultados: (1) Menos de 40 puntos = Controles sanos de ansiedad, (2) Entre 40 y 60 puntos = Presencia de ansiedad (no TAG), (3) Más de 60 puntos = Presencia de Trastorno de Ansiedad Generalizada (TAG). La vista parcial de este cuestionario puede visualizarse en el Apéndice B.5.

Los instrumentos solamente se aplicaron con previa lectura y firma del consentimiento informado por cada participante. El consentimiento informado se puede visualizar en el Apéndice A.1 y A.2. En la Tabla 4.1 se muestran las variables recolectadas de los instrumentos de estrategias de historia de vida y el Inventario de Preocupación de Pensilvania (PSWQ).

Tabla 4.1: Resumen de las variables recolectadas de los instrumentos Marcadores de estrategias de historia de vida, K-SF-42 y PSWQ

| Marcadores de estrategias de historia de vida | Nuevo formato corto de la Batería de Historia de Vida de Arizona (K-SF-42) | Inventario de Preocupación de Pensilvania (PSWQ) |
|-----------------------------------------------|----------------------------------------------------------------------------|--------------------------------------------------|
|                                               | Mala situación                                                             |                                                  |
|                                               | No me rindo en situaciones adversas                                        |                                                  |
|                                               | Aprendo de situaciones adversas                                            |                                                  |
|                                               | Manera diferente de procesar situaciones adversas                          |                                                  |
|                                               | Lado bueno de las cosas                                                    |                                                  |
|                                               | Positivo en peores situaciones                                             |                                                  |
|                                               | Apoyo emocional a parientes                                                |                                                  |
|                                               | Bienestar parientes                                                        |                                                  |
|                                               | Apoyo emocional conocidos                                                  |                                                  |
|                                               | Bienestar amigos                                                           |                                                  |
|                                               | Voluntario                                                                 |                                                  |
|                                               | Contribuyo a causas                                                        |                                                  |
|                                               | Soy una persona religiosa                                                  |                                                  |
|                                               | Religión es importante                                                     |                                                  |
|                                               | Espiritualidad es importante                                               |                                                  |
|                                               | Miembro de una religión                                                    |                                                  |
|                                               | Asisto a servicios                                                         |                                                  |
|                                               | Qué hacer basado en creencias                                              |                                                  |
|                                               | Pareja no interés                                                          |                                                  |
|                                               | Incomodidad privacidad                                                     |                                                  |
|                                               | Marcar distancia                                                           |                                                  |
|                                               | Fusionarme pareja                                                          |                                                  |
|                                               | Demasiada intimidad                                                        |                                                  |
|                                               | Pareja no intimidad                                                        |                                                  |
|                                               | Atención mamá                                                              |                                                  |
|                                               | Esfuerzo mamá                                                              |                                                  |
|                                               | Enseñaba mamá                                                              |                                                  |
|                                               | Atención papá                                                              |                                                  |
|                                               | Esfuerzo papá                                                              |                                                  |
|                                               | Enseñaba papá                                                              |                                                  |
|                                               | Ayuda familiares                                                           |                                                  |
|                                               | Hiciste bien familiares                                                    |                                                  |
|                                               | Forma de ser familiares                                                    |                                                  |
|                                               | Familia muestra afecto                                                     |                                                  |
|                                               | Familia escuchar sentimientos                                              |                                                  |
|                                               | Familia se preocupa por bienestar                                          |                                                  |
|                                               | Ayuda amigos                                                               |                                                  |
|                                               | Hiciste bien amigos                                                        |                                                  |
|                                               | Forma de ser amigos                                                        |                                                  |
|                                               | Amigos muestran afecto                                                     |                                                  |
|                                               | Amigos escuchan sentimientos                                               |                                                  |
|                                               | Amigos se preocupan por bienestar                                          |                                                  |
|                                               | Factor K normalizado                                                       |                                                  |
| Amabilidad                                    |                                                                            | Preocupo si no tengo suficiente tiempo           |
| Consciencia                                   |                                                                            | Preocupaciones agobian                           |
| Humildad                                      |                                                                            | Preocupación por las cosas                       |
| Honestidad                                    |                                                                            | Muchas circunstancias me preocupan               |
| Apertura a la experiencia                     |                                                                            | No puedo evitar preocupación                     |
| Experiencias sexuales                         |                                                                            | Bajo tensión                                     |
| Socio sexualidad                              |                                                                            | Siempre preocupado/a                             |
| Amor inestable                                |                                                                            | Difícil dejar preocupaciones                     |
| Planes inmediatos                             |                                                                            | Me preocupo al terminar una tarea                |
| Disgusto sexual                               |                                                                            | Todo me preocupa                                 |
| Decisiones impulsivas                         |                                                                            | Sigo preocupándome                               |
| Asumir riesgos                                |                                                                            | Toda mi vida me he preocupado                    |
| Cognición mentalista                          |                                                                            | Preocupación excesiva                            |
| Habilidad verbal                              |                                                                            | No puedo detener preocupación                    |
| Habilidad visoespacial                        |                                                                            | Preocupación constante                           |
| Madurez sexual temprana                       |                                                                            | Preocupación por proyectos                       |

## Participantes

La primera recolección de datos se llevó a cabo en la Unidad de Medicina Familiar 20 del IMSS en Cuernavaca, Morelos. El número total de participantes fue de  $n = 244$ . El promedio de edad fue de 32.04 años ( $DE = 10.76$ ), con un rango de 18 a 59 años donde el 67 % de los participantes fueron mujeres. En términos de educación, el 41 % se había graduado de la preparatoria y un 36 % había completado sus estudios universitarios. En este conjunto de datos, 78 participantes tenían un control sano de ansiedad, 92 presentaban ansiedad diferente de TAG y 74 tenían indicios de TAG. La Tabla 4.2 presenta un resumen de las características sociodemográficas, el puntaje del factor K y la categoría de ansiedad de los participantes.

### 4.3. Preprocesamiento del conjunto de datos

Durante la fase del preprocesamiento de los datos obtenidos, se llevaron a cabo diversas actividades para entender las características del conjunto de datos y prepararlos para las fases posteriores. Las actividades involucraron la limpieza de los datos, el análisis preliminar, el análisis univariado, el análisis bivariado, la selección de características y el balanceo de clases.

#### Limpieza de datos y normalización de variables

El conjunto de datos requirió manejo de valores faltantes y atípicos, los cuales, en su mayoría, se originaron durante el proceso de transcripción de información de los cuestionarios físicos a formato digital. Para abordar los datos faltantes, se aplicó una imputación simple basada en la moda, dado que las variables afectadas eran cualitativas.

Por otro lado, el cálculo del factor K, un índice que integra diversas dimensiones del comportamiento y la cognición, se realizó a partir del promedio de la suma de los puntajes obtenidos en cada aspecto evaluado por el instrumento K-SF-42. Este factor fue normalizado para obtener valores en un rango de 0 a 1, facilitando así su interpretación.

#### Análisis preliminar

Se evaluó la normalidad de las variables cuantitativas mediante histogramas, gráficos de cajas y el estadístico de Shapiro-Wilk. La prueba de Shapiro-Wilk evalúa la hipótesis nula de que los datos siguen una distribución normal (Shapiro y Wilk, 1965). La variable edad no presentó normalidad ( $p < 0.05$ ), como se muestra en la Figura 4.2, por lo que se utilizó la prueba no paramétrica de Kruskal-Wallis para realizar la comparación con las categorías de ansiedad. En cambio, la variable factor K cumplió con la normalidad ( $p = 0.14$ ), como se muestra en la Figura 4.3, permitiendo el uso de ANOVA de una vía.

#### Análisis bivariado

En el análisis bivariado se utilizó el estadístico de Chi Cuadrada y Exacta de Fisher para comparar las frecuencias observadas con las frecuencias esperadas bajo la hipótesis nula de las variables categóricas. La prueba de Chi Cuadrada de independencia evalúa si existe una relación significativa entre dos variables categóricas (McHugh, 2013), mientras que la

Tabla 4.2: Resumen de características sociodemográficas, factor K y categoría de ansiedad de los participantes

| Variables                        |                                     | Frecuencia (Porcentaje) |
|----------------------------------|-------------------------------------|-------------------------|
| Sexo                             | Mujeres                             | 155 (63.52)             |
|                                  | Hombres                             | 89 (36.48)              |
| Edad, en años                    | Mediana (p25, p75)                  | 29 (24, 38)             |
| Nivel educativo                  | Primaria                            | 6 (2.46)                |
|                                  | Secundaria                          | 44 (18.03)              |
|                                  | Preparatoria                        | 100 (40.98)             |
|                                  | Universidad                         | 87 (35.66)              |
|                                  | Posgrado                            | 7 (2.87)                |
| Estado civil                     | Soltero/a                           | 108 (44.26)             |
|                                  | Casado/a                            | 70 (28.69)              |
|                                  | Unión libre                         | 46 (18.85)              |
|                                  | Divorciado/a                        | 8 (3.28)                |
|                                  | Viudo/a                             | 5 (2.05)                |
|                                  | Separado/a                          | 7 (2.05)                |
| Ocupación                        | Estudiante                          | 41 (16.80)              |
|                                  | Empleado sin estudios profesionales | 53 (21.72)              |
|                                  | Empleado con estudios profesionales | 79 (32.38)              |
|                                  | Negocio propio                      | 9 (3.69)                |
|                                  | Trabajador independiente            | 20 (8.20)               |
|                                  | Hogar                               | 29 (11.89)              |
|                                  | Jubilado/a                          | 4 (1.64)                |
|                                  | Desempleado/a                       | 9 (3.69)                |
| Nivel socioeconómico             | Alto                                | 0 (0)                   |
|                                  | Medio-alto                          | 42 (17.21)              |
|                                  | Medio                               | 155 (63.52)             |
|                                  | Medio-bajo                          | 43 (17.62)              |
|                                  | Bajo                                | 3 (1.23)                |
|                                  | Pobreza                             | 1 (0.41)                |
| Personas que habitan en el hogar | Familia nuclear                     | 132 (54.10)             |
|                                  | Padre, madre o ambos                | 31 (12.70)              |
|                                  | Hermanos                            | 9 (3.69)                |
|                                  | Pareja/Cónyuge                      | 37 (15.16)              |
|                                  | Hijos/as                            | 18 (7.38)               |
|                                  | Amigos                              | 3 (1.23)                |
|                                  | Solo/a                              | 14 (5.74)               |
| Tiene hijos                      | No                                  | 119 (48.77)             |
|                                  | Sí                                  | 125 (51.23)             |
| Hay comorbilidades               | No                                  | 183 (75)                |
|                                  | Sí                                  | 61 (25)                 |
| Factor K                         | Media (D.E.)                        | 0.53 ± 0.19             |
| Categoría de ansiedad            | Control sano de ansiedad            | 78 (31.97)              |
|                                  | Ansiedad no TAG                     | 92 (37.70)              |
|                                  | Trastorno de ansiedad generalizada  | 74 (30.33)              |

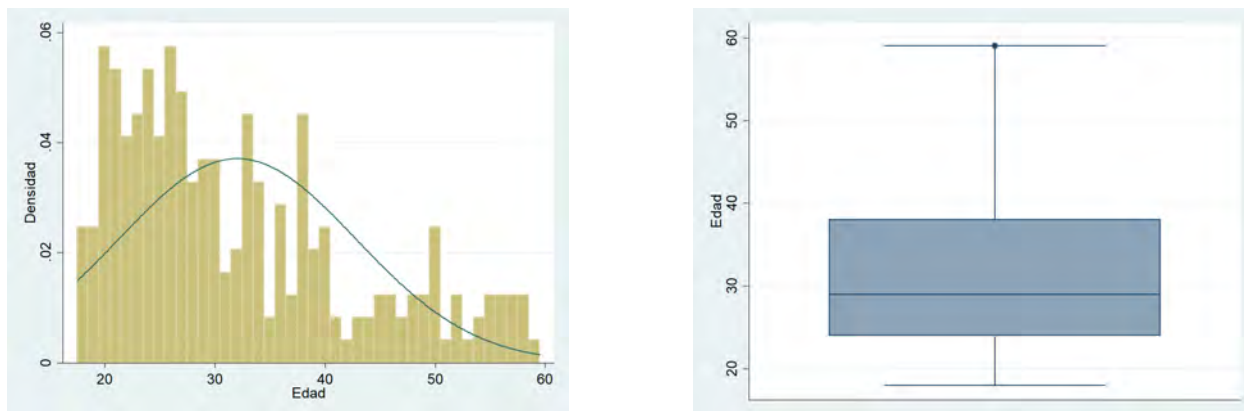


Figura 4.2: Histograma y gráfico de caja de la variable Edad

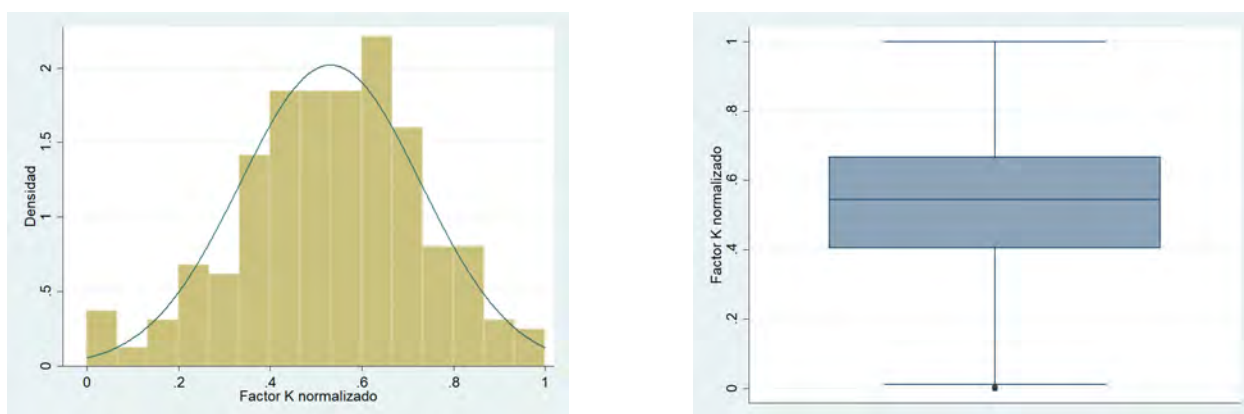


Figura 4.3: Histograma y gráfico de caja de la variable Factor K normalizado

prueba Exacta de Fisher evalúa si las proporciones de dos grupos difieren significativamente cuando las frecuencias esperadas son menores a 5 (Kim, 2017). Las variables categóricas se contrastaron con la categoría de ansiedad, medida por el puntaje total del PSWQ (Tabla 4.3).

Las variables cuantitativas se analizaron con los estadísticos de Kruskal Wallis y ANOVA. La prueba de Kruskal-Wallis es una prueba no paramétrica utilizada para evaluar si existen diferencias significativas entre las distribuciones de tres o más grupos independientes, sin asumir normalidad (Kruskal y Wallis, 1952). Por otra parte, la prueba de Análisis de Varianza (ANOVA) de una vía es una prueba estadística paramétrica que evalúa si existen diferencias significativas entre las medias de tres o más grupos independientes (Gelman, 2005). Las variables se contrastaron con la categoría de ansiedad, medida por el puntaje total del PSWQ (Tabla ??).

Para la variable edad, que obtuvo un  $p$ -valor significativo, las comparaciones post hoc con el método Dunn y ajuste Bonferroni indicaron diferencias significativas entre el grupo de *control sano* y el de *trastorno de ansiedad generalizada* ( $p = 0.03$ ), y entre *ansiedad no TAG* y *trastorno de ansiedad generalizada* ( $p = 0.05$ ). Para el factor K que también tuvo significancia estadística, la prueba post hoc de Tukey HSD señaló una diferencia significativa entre el grupo de *control sano* y el grupo con *trastorno de ansiedad generalizada* ( $p = 0.001$ ).

Tabla 4.3: Análisis bivariado de las variables categóricas sociodemográficas

| Variables                        |                                     | Control sano de ansiedad<br>(n = 78) | Ansiedad no TAG<br>(n = 92) | TAG<br>(n = 74) | p valor    |
|----------------------------------|-------------------------------------|--------------------------------------|-----------------------------|-----------------|------------|
| Sexo                             | Mujeres                             | 41 (52.56 %)                         | 55 (59.78 %)                | 59 (79.73 %)    | p < 0.01*  |
|                                  | Hombres                             | 37 (47.44 %)                         | 37 (40.22 %)                | 15 (20.27 %)    |            |
| Nivel educativo                  | Primaria                            | 1 (1.28 %)                           | 3 (3.26 %)                  | 2 (2.70 %)      | p = 0.89** |
|                                  | Secundaria                          | 17 (21.79 %)                         | 13 (14.13 %)                | 14 (18.92 %)    |            |
|                                  | Preparatoria                        | 31 (39.74 %)                         | 37 (40.22 %)                | 32 (43.24 %)    |            |
|                                  | Universidad                         | 27 (34.62 %)                         | 35 (38.04 %)                | 25 (33.78 %)    |            |
|                                  | Posgrado                            | 2 (2.56 %)                           | 4 (4.35 %)                  | 1 (1.35 %)      |            |
| Estado civil                     | Soltero/a                           | 32 (41.03 %)                         | 42 (45.65 %)                | 34 (45.95 %)    | p = 0.71** |
|                                  | Casado/a                            | 27 (34.62 %)                         | 21 (22.83 %)                | 22 (29.73 %)    |            |
|                                  | Unión libre                         | 13 (16.67 %)                         | 20 (21.74 %)                | 13 (17.57 %)    |            |
|                                  | Divorciado/a                        | 3 (3.85 %)                           | 4 (4.35 %)                  | 1 (1.35 %)      |            |
|                                  | Viudo/a                             | 0 (0 %)                              | 3 (3.26 %)                  | 2 (2.70 %)      |            |
|                                  | Separado/a                          | 3 (3.85 %)                           | 2 (2.17 %)                  | 2 (2.70 %)      |            |
| Ocupación                        | Estudiante                          | 11 (14.10 %)                         | 16 (17.39 %)                | 14 (18.92 %)    | p = 0.70** |
|                                  | Empleado sin estudios profesionales | 20 (25.64 %)                         | 18 (19.57 %)                | 15 (20.27 %)    |            |
|                                  | Empleado con estudios profesionales | 25 (32.05 %)                         | 32 (34.78 %)                | 22 (29.73 %)    |            |
|                                  | Negocio propio                      | 4 (5.13 %)                           | 2 (2.17 %)                  | 3 (4.05 %)      |            |
|                                  | Trabajador independiente            | 6 (7.69 %)                           | 10 (10.87 %)                | 4 (5.41 %)      |            |
|                                  | Hogar                               | 8 (10.26 %)                          | 10 (10.87 %)                | 11 (14.86 %)    |            |
|                                  | Jubilado/a                          | 3 (3.85 %)                           | 1 (1.09 %)                  | 0 (0 %)         |            |
| Nivel socioeconómico             | Desempleado/a                       | 1 (1.28 %)                           | 3 (3.26 %)                  | 5 (6.76 %)      | p = 0.22** |
|                                  | Medio-alto                          | 18 (23.08 %)                         | 16 (17.39 %)                | 8 (10.81 %)     |            |
|                                  | Medio                               | 46 (58.97 %)                         | 58 (63.04 %)                | 51 (68.92 %)    |            |
|                                  | Medio-bajo                          | 14 (17.95 %)                         | 14 (15.22 %)                | 15 (20.27 %)    |            |
|                                  | Bajo                                | 0 (0 %)                              | 3 (3.26 %)                  | 0 (0 %)         |            |
| Personas que habitan en el hogar | Pobreza                             | 0 (0 %)                              | 1 (1.09 %)                  | 0 (0 %)         | p = 0.97** |
|                                  | Familia nuclear                     | 41 (52.56 %)                         | 48 (52.17 %)                | 43 (58.11 %)    |            |
|                                  | Padre, madre o ambos                | 9 (11.54 %)                          | 11 (11.96 %)                | 11 (14.86 %)    |            |
|                                  | Hermanos                            | 4 (5.13 %)                           | 3 (3.26 %)                  | 2 (2.70 %)      |            |
|                                  | Pareja/Cónyuge                      | 10 (12.82 %)                         | 17 (18.48 %)                | 10 (13.51 %)    |            |
|                                  | Hijos/as                            | 7 (8.97 %)                           | 7 (7.61 %)                  | 4 (5.41 %)      |            |
|                                  | Amigos                              | 2 (2.56 %)                           | 1 (1.09 %)                  | 0 (0 %)         |            |
| Tiene hijos                      | Solo/a                              | 5 (6.41 %)                           | 5 (5.43 %)                  | 4 (5.41 %)      | p = 0.15*  |
|                                  | No                                  | 35 (44.87 %)                         | 41 (44.57 %)                | 43 (58.11 %)    |            |
| Hay comorbilidades               | Sí                                  | 43 (55.13 %)                         | 51 (55.43 %)                | 31 (41.89 %)    | p = 0.71*  |
|                                  | No                                  | 61 (78.21 %)                         | 67 (72.83 %)                | 55 (74.32 %)    |            |
|                                  | Sí                                  | 17 (21.79 %)                         | 25 (27.17 %)                | 19 (25.68 %)    |            |
|                                  | No                                  |                                      |                             |                 |            |

Nota: \* Estadístico Chi Cuadrada. \*\* Estadístico Exacta de Fisher.

Tabla 4.4: Análisis bivariado de las variables cuantitativas (edad y factor K)

| Variables       | Control sano de ansiedad<br>(n = 78) | Ansiedad no TAG<br>(n = 92) | TAG<br>(n = 74) | p valor     |
|-----------------|--------------------------------------|-----------------------------|-----------------|-------------|
| Edad, en años * | 31 (25, 39)                          | 31 (24, 41)                 | 26 (22, 35)     | p < 0.001 † |
| Factor K **     | 0.59 ± 0.21                          | 0.52 ± 0.17                 | 0.47 ± 0.18     | p < 0.01 †† |

Nota: \* Mediana (percentil 25, percentil 75). \*\* Media ± desviación estándar. † Kruskal-Wallis. †† ANOVA.

Como resultado del análisis bivariado, las variables sociodemográficas seleccionadas para las siguientes etapas fueron *edad*, *sexo*, *factor K*, *tiene hijos* y, por su relevancia reportada en la literatura, el *nivel socioeconómico* y *nivel educativo*.

Se evaluó también la significancia estadística de los 42 ítems que pertenecen al instrumento de K-SF-42 contrastando con las categorías de ansiedad medida por el instrumento PSWQ utilizando el estadístico de Chi Cuadrada y Exacta de Fisher. Se consideraron también las variables que estuvieran cercanas al nivel de significancia establecido ( $p < 0.05$ ). Se encontraron 19 variables que cumplieran con estos criterios, las cuales son: **(1)** 2 variables de la subescala Perspectiva, Planificación y Control, **(2)** 3 variables de la subescala Altruismo General, **(3)** 5 variables de la subescala Experiencia en Relaciones Cercanas, **(4)** 3 variables de la subescala Calidad de Relaciones Parentales, y **(5)** 6 variables de la subescala Apoyo Familiar y de Amigos.

## Selección de características

La selección de características es una actividad fundamental en el aprendizaje automático ya que mejora el desempeño del modelo al identificar y conservar las variables más importantes mientras que descarta las que no aportan información significativa (Cheng, 2024; Rawat, 2021). Para este proceso se utilizaron las técnicas de Análisis Factorial (AF) y Eliminación Recursiva de Características (RFE).

### Análisis Factorial

El Análisis Factorial Exploratorio (AFE) se emplea para examinar los datos y determinar la cantidad de factores cuando no se cuenta con suficiente evidencia previa. Por otro lado, el Análisis Factorial Confirmatorio (AFC) se utiliza para evaluar hipótesis específicas sobre la estructura factorial obtenida del AFE (Stapleton, 1997). Se realizaron los análisis factoriales, tanto exploratorio como confirmatorio, sobre el cuestionario de marcadores de estrategias de historia de vida. El objetivo fue validar los constructos subyacentes, confirmar la fiabilidad del instrumento y obtener un subconjunto de variables que representaran adecuadamente los marcadores.

Inicialmente, se planteó que el instrumento evaluaba cinco factores relacionados con las estrategias de historia de vida con los 16 ítems: personalidad, motivación, autorregulación, habilidad cognitiva y maduración sexual. En el Análisis Factorial Exploratorio (Figura 4.4), se identificó que el conjunto de ítems X8, X9, X11 y X12 está asociado al factor "Motivación", mientras que los ítems X6 y X16 corresponden al factor "Sexualidad".

La correspondencia entre el modelo hipotetizado y los datos observados se analiza mediante diversos estadísticos de bondad de ajuste. Se realizó el análisis confirmatorio donde se utilizó la siguiente hipótesis:

- La motivación es medida por los ítems 8, 9, 10 y 11.
- La sexualidad es medida por los ítems 6 y 16.

Los resultados de los estadísticos de bondad de ajuste del modelo se muestran en la Tabla 4.5. El  $p$ -valor del estadístico chi cuadrado es mayor a 0.05, lo que permite aceptar

## Análisis Factorial Exploratorio

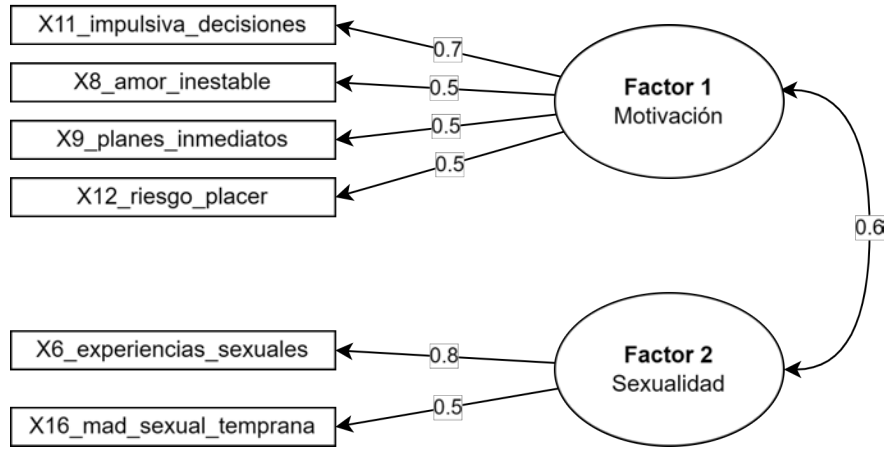


Figura 4.4: Diagrama del análisis factorial exploratorio de cuestionario Marcadores de estrategias de historia de vida con 2 factores con cargas factoriales

la hipótesis nula de que el modelo se ajusta correctamente a los datos. Los índices CFI y TLI presentan valores superiores a 0.95, indicando un buen ajuste del modelo. Además, el RMSEA muestra un valor inferior a 0.05, lo que refuerza la calidad del ajuste, al igual que el SRMR, que también es menor a 0.05. En conjunto, estos resultados confirman que los ítems seleccionados evalúan de manera adecuada los factores de 'Motivación' y 'Sexualidad'.

Tabla 4.5: Resumen de resultados de las medidas de bondad de ajuste del análisis factorial confirmatorio

| Medida de bondad de ajuste         | Resultado                    |
|------------------------------------|------------------------------|
| $\chi^2$                           | $\chi^2(8) = 8.64, p = 0.37$ |
| <i>Comparative Fit Index (CFI)</i> | 0.99                         |
| <i>Tucker-Lewis Index (TLI)</i>    | 0.99                         |
| <i>RMSEA</i>                       | 0.01                         |
| <i>SRMR</i>                        | 0.03                         |

### Eliminación Recursiva de Características (RFE)

La eliminación recursiva de características es un método de selección de características que tiene como propósito eliminar aquellas poco relevantes mientras conserva las más independientes y significativas para mejorar la capacidad de generalización del modelo (Misra y Yadav, 2020). Esta técnica emplea un procedimiento iterativo de clasificación de características, basado en la eliminación regresiva. Se implementó en las 19 variables seleccionadas del instrumento K-SF-42, el factor K normalizado, el subconjunto de 6 variables de los marcadores de estrategias de historia de vida y las 2 variables sociodemográficas estadísticamente

significativas del análisis bivariado.

Se emplearon 10 pliegues de validación para evitar el sobreajuste (Yates et al., 2023), y las características consideradas como no informativas fueron eliminadas. De las 19 variables seleccionadas del instrumento K-SF-42, se eliminaron *familia se preocupa por bienestar*, *forma de ser familiares*, y *familia muestra afecto*. Además, se agregaron dos variables referentes a la preocupación para complementar la información de las otras variables seleccionadas que son *bajo tensión* y *preocupación excesiva*. El resultado final después de utilizar RFE fueron 27 variables seleccionadas para el entrenamiento de los modelos.

## Balanceo de clases

En el primer conjunto de datos existió un ligero desbalance en las tres clases a detectar de ansiedad, por lo que para mejorar el desempeño del modelo se decidió utilizar la técnica de submuestreo aleatorio (*random undersampling*, en inglés). El submuestreo aleatorio es una técnica utilizada para abordar el desbalance de clases en conjuntos de datos mediante la reducción de instancias de la clase mayoritaria y ajustarla al número de instancias de la clase minoritaria (Goswami y Roy, 2021). Originalmente, el conjunto de datos tuvo 78 instancias en la clase 0 (*control sano de ansiedad*), 92 instancias en clase 1 (*ansiedad no TAG*), y 74 instancias en la clase 2 (*TAG*). Después de aplicar la técnica de submuestreo aleatorio, todas las clases tienen 74 instancias (Figura 4.5). En total, se eliminaron 22 instancias del conjunto de datos original.

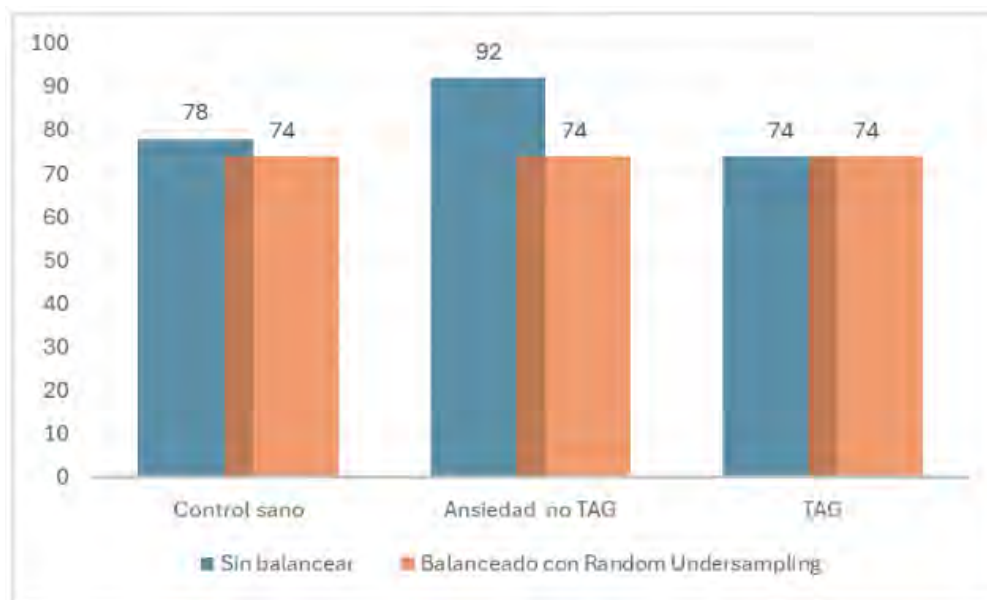


Figura 4.5: Comparativa de distribución de las clases antes y después de utilizar la técnica de balanceo submuestreo aleatorio (Random Undersampling)

## 4.4. Desarrollo del modelo

El conjunto de datos final que se utilizó para entrenar los modelos de aprendizaje está compuesto por 27 variables: **(1)** 2 variables sociodemográficas, **(2)** 6 ítems pertenecientes a los marcadores de estrategias de historia de vida, **(3)** 16 ítems del instrumento K-SF-42, **(4)** el factor K normalizado, y **(5)** dos ítems del instrumento PSWQ. Las variables están resumidas en la Tabla 4.6.

Tabla 4.6: Resumen de las variables de entrada utilizadas en el modelo para detección

| Instrumento al que pertenece                  | Variable                                | Descripción                                                                                                                        |
|-----------------------------------------------|-----------------------------------------|------------------------------------------------------------------------------------------------------------------------------------|
| Sociodemográfico                              | Edad                                    | Variable numérica que representa la edad del derechohabiente en años.                                                              |
|                                               | Sexo                                    | Variable binaria indicando el sexo del derechohabiente. '0' representa mujer, '1' representa hombre.                               |
| Marcadores de estrategias de historia de vida | Experiencias sexuales                   | Variable categórica con un rango de valores es de 1 a 5, donde 1 es 'No me describe en absoluto' y 5 es 'Me describe en absoluto'. |
|                                               | Amor inestable                          |                                                                                                                                    |
|                                               | Planes inmediatos                       |                                                                                                                                    |
|                                               | Decisiones impulsivas                   |                                                                                                                                    |
|                                               | Riesgo placer                           |                                                                                                                                    |
| K-SF-42                                       | Madurez sexual temprana                 | Variable categórica con un rango de valores de -3 a 3, donde -3 es 'Fuertemente en desacuerdo' y 3 es 'Fuertemente de acuerdo'.    |
|                                               | Manera diferente de ver las situaciones |                                                                                                                                    |
|                                               | Positivo en peores situaciones          |                                                                                                                                    |
|                                               | Bienestar parientes                     |                                                                                                                                    |
|                                               | Bienestar amigos                        | Variable categórica con un rango de valores de -3 a 3, donde -3 es 'Fuertemente en desacuerdo' y 3 es 'Fuertemente de acuerdo'.    |
|                                               | Contribuyo a causas                     |                                                                                                                                    |
|                                               | Pareja no interés                       |                                                                                                                                    |
|                                               | Marcar distancia                        |                                                                                                                                    |
|                                               | Fusionarme pareja                       |                                                                                                                                    |
|                                               | Demasiada intimidad                     |                                                                                                                                    |
|                                               | Pareja no intimidad                     |                                                                                                                                    |
|                                               | Atención mamá                           | Variable categórica con un rango de valores de 0 a 3, donde 0 es 'Nunca' y 3 es 'Mucho'.                                           |
|                                               | Atención papá                           |                                                                                                                                    |
|                                               | Enseñaba papá                           |                                                                                                                                    |
|                                               | Familia escuchar sentimientos           |                                                                                                                                    |
|                                               | Ayuda amigos                            |                                                                                                                                    |
|                                               | Hiciste bien amigos                     | Variable numérica obtenido como resultado del instrumento K-SF-42. El rango de valores es de 0 a 1.                                |
|                                               | Factor K normalizado                    |                                                                                                                                    |
| PSWQ                                          | Bajo tensión                            | Variable categórica con un rango de valores es de 1 a 5, donde 1 es 'No me describe en absoluto' y 5 es 'Me describe en absoluto'. |
|                                               | Preocupación excesiva                   |                                                                                                                                    |

Los modelos de aprendizaje automático supervisados desarrollados en este trabajo fueron **(1)** Máquina de Soporte Vectorial (SVM), **(2)** Random Forest, **(3)** K Vecinos más Cercanos (KNN), **(4)** Naive Bayes, **(5)** XGBoost, **(6)** Perceptrón Multicapa, y **(7)** Regresión Logística. Para llevar a cabo las pruebas, se empleó la validación cruzada estratificada de 10 pliegues, donde el primer conjunto de datos se utilizó para el entrenamiento y las pruebas en una proporción de 90 % y 10 % respectivamente por cada pliegue. Los algoritmos se entrenaron siguiendo la configuración inicial descrita en la Tabla 4.7.

Tabla 4.7: Configuración de hiperparámetros inicial de los algoritmos de aprendizaje supervisado seleccionados

| Modelo               | Configuración de hiperparámetros                                                                    |
|----------------------|-----------------------------------------------------------------------------------------------------|
| SVM                  | kernel='rbf', C=10, gamma=0.01, class_weight='balanced'                                             |
| KNN                  | n_neighbors=5, weights='uniform'                                                                    |
| Naive Bayes          | var_smoothing=1e-09                                                                                 |
| XGBoost              | eval_metric="mlogloss"                                                                              |
| Random Forest        | n_estimators=100, max_depth=None, max_features='sqrt'                                               |
| Regresión Logística  | max_iter=500                                                                                        |
| Perceptrón Multicapa | hidden_layer_sizes=(100,), max_iter=200, activation='relu', solver='adam', learning_rate_init=0.001 |

Los resultados obtenidos de la implementación de los modelos con el conjunto de datos sin balancear se muestran en la Tabla 4.8. El modelo Random Forest presentó el mejor desempeño general (exactitud de 80 %, F1 score de 79 %), seguido por XGBoost, Regresión Logística y Perceptrón Multicapa. KNN fue el modelo con el peor desempeño (exactitud de 49 %). La variabilidad promedio de cada modelo se puede observar en la Figura 4.6.

Tabla 4.8: Resultados de las pruebas con los modelos implementados utilizando el conjunto de datos sin balancear

| Modelo               | Exactitud   | Precisión   | Sensibilidad | F1 Score    |
|----------------------|-------------|-------------|--------------|-------------|
| SVM                  | 0.69        | 0.70        | 0.69         | 0.68        |
| KNN                  | 0.49        | 0.57        | 0.49         | 0.48        |
| Naive Bayes          | 0.76        | 0.77        | 0.76         | 0.75        |
| XGBoost              | 0.78        | 0.79        | 0.78         | 0.78        |
| <b>Random Forest</b> | <b>0.80</b> | <b>0.81</b> | <b>0.80</b>  | <b>0.79</b> |
| Regresión Logística  | 0.78        | 0.79        | 0.78         | 0.77        |
| Perceptrón Multicapa | 0.78        | 0.80        | 0.78         | 0.77        |

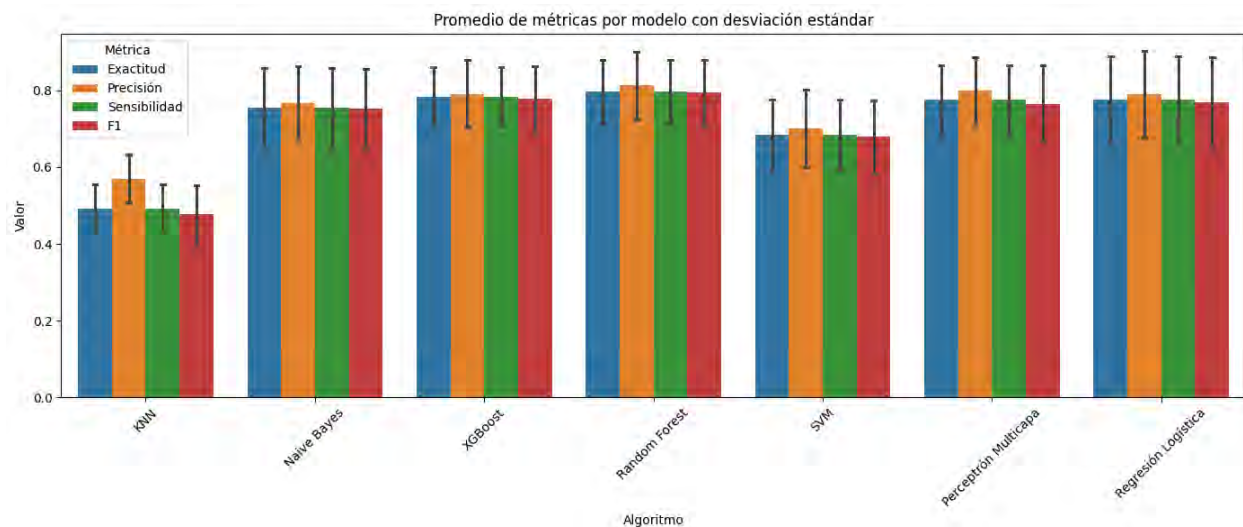


Figura 4.6: Comparativa de resultados de las métricas de desempeño en los modelos implementados utilizando el conjunto de datos sin balancear

Los resultados obtenidos de la implementación de los modelos con el conjunto de datos balanceado con submuestreo aleatorio se muestran en la Tabla 4.9. Las métricas de desempeño presentaron valores iguales o superiores al 70 % en la mayoría de los modelos, destacando el modelo Random Forest que logró 81 % en las métricas de exactitud, precisión, sensibilidad y F1 score, seguido del modelo XGBoost que obtuvo valores iguales o superiores al 80 % en las mismas métricas. La variabilidad promedio de cada modelo se puede observar en la Figura 4.7.

Tabla 4.9: Resultados de las pruebas con los modelos implementados utilizando el conjunto de datos balanceado

| Modelo               | Exactitud   | Precisión   | Sensibilidad | F1 Score    |
|----------------------|-------------|-------------|--------------|-------------|
| SVM                  | 0.72        | 0.73        | 0.72         | 0.72        |
| KNN                  | 0.48        | 0.51        | 0.48         | 0.48        |
| Naive Bayes          | 0.79        | 0.79        | 0.79         | 0.79        |
| XGBoost              | 0.81        | 0.80        | 0.81         | 0.80        |
| <b>Random Forest</b> | <b>0.81</b> | <b>0.81</b> | <b>0.81</b>  | <b>0.81</b> |
| Regresión Logística  | 0.80        | 0.79        | 0.80         | 0.80        |
| Perceptrón Multicapa | 0.77        | 0.76        | 0.77         | 0.77        |

Las matrices de confusión (Figura 4.8) de los modelos entrenados que obtuvieron resultados superiores al 70 % en las métricas de exactitud, precisión, sensibilidad y F1 score muestran su desempeño en tres clases equilibradas (74 instancias cada una): *Control sano*, *Ansiedad no TAG* y *TAG*. Random Forest logró un rendimiento sobresaliente, clasificando correctamente el 86.4 % de los casos *TAG* y el 90.5 % de *control sano*. La clase *ansiedad no TAG* presentó más clasificaciones incorrectas, confundándose con las otras dos categorías.

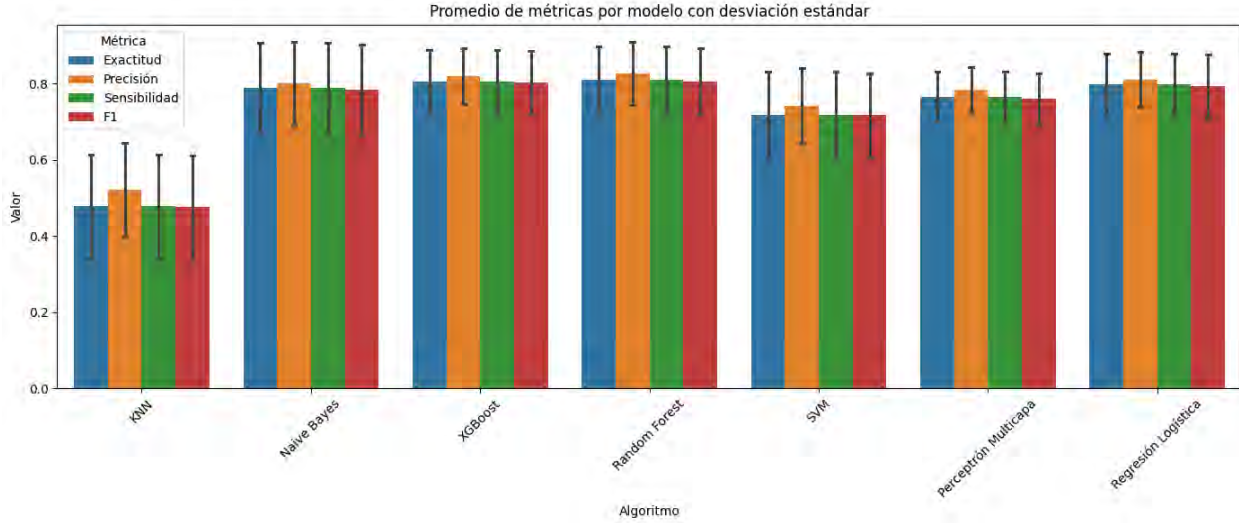


Figura 4.7: Comparativa de resultados de las métricas de desempeño en los modelos implementados utilizando el conjunto de datos balanceado

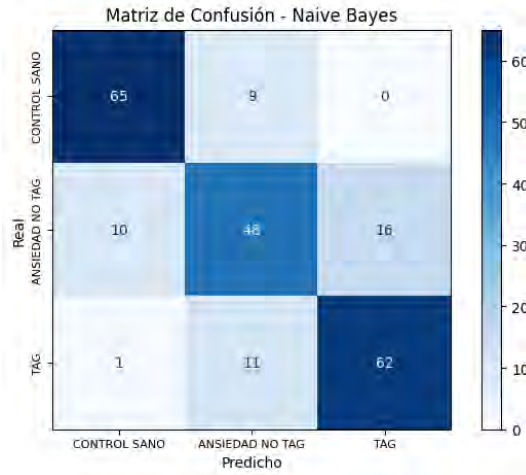
En general, Random Forest, XGBoost y Regresión Logística destacaron por su precisión en la identificación de casos TAG.

Se graficaron 10 curvas ROC micro-promedio, una por cada pliegue, y una curva ROC micro-promedio que representa el desempeño promedio de todos los pliegues. La curva ROC (*Receiver Operating Characteristic*) micro-promedio representa el rendimiento global del clasificador considerando todas las clases de manera conjunta (Rashidi et al., 2023). La curva ROC micro-promedio presentada en la Figura 4.9a muestra el rendimiento del modelo Random Forest balanceado con la técnica de submuestreo aleatorio y evaluado mediante validación cruzada de 10 pliegues, el cual obtuvo los mejores resultados en las métricas evaluadas.

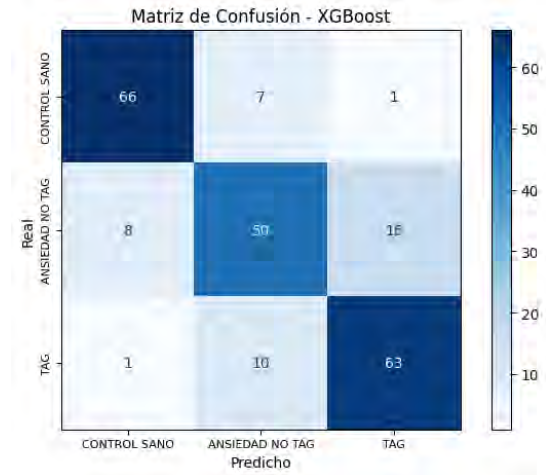
Se observó que la mayoría de los pliegues tiene un valor de AUC igual o mayor a 0.85, lo que sugiere un buen desempeño del modelo. La región sombreada alrededor de la curva de desempeño general representa la desviación estándar obtenida entre los diferentes pliegues, donde la amplitud de esta área con una desviación estándar de  $\pm 0.03$  indica que el modelo es estable en los diferentes pliegues.

La estrategia *One-vs-Rest* (también conocida como *One-vs-All*) permite transformar la matriz de confusión original de dimensiones  $K \times K$  en  $K$  matrices binarias de  $2 \times 2$ , cada una correspondiente a una clasificación individual (Rashidi et al., 2023). Es así como la métrica macro representa el valor promedio AUC para cada clase individual. Las curvas ROC generadas con la estrategia *One-vs-Rest* para el modelo Random Forest con submuestreo aleatorio y evaluado mediante validación cruzada de 10 pliegues se muestran en la Figura 4.9b.

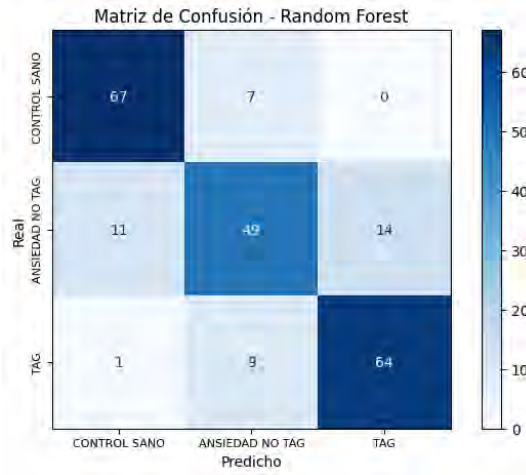
Las curvas ROC obtenidas por cada clase muestran valores AUC de 0.97, 0.87, y 0.93 para las clases *control sano*, *ansiedad no TAG*, y *TAG* respectivamente, calculados como promedios a través de las 10 iteraciones de validación cruzada. La clase *control sano* resalta con el valor AUC más elevado (0.97), indicando una capacidad discriminativa elevada para esta categoría. La variabilidad entre pliegues, representada por las regiones sombreadas alrededor de cada curva, muestra desviaciones estándar de  $\pm 0.03$ ,  $\pm 0.06$ , y  $\pm 0.05$ , revelando



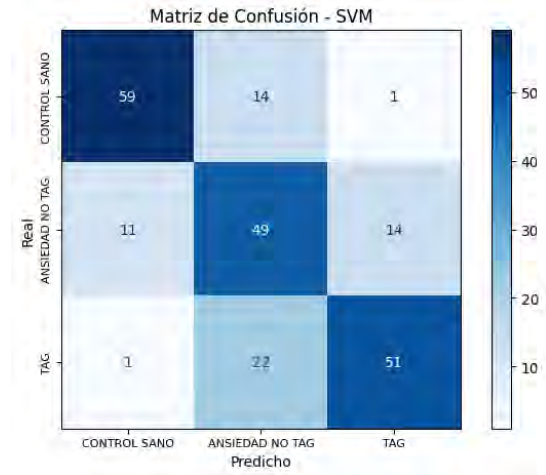
(i) Modelo Naive Bayes



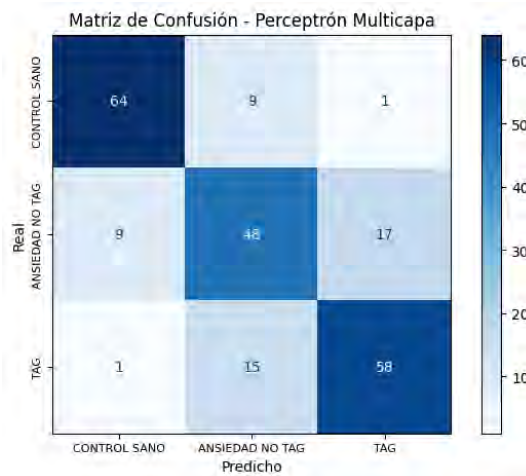
(ii) Modelo XGBoost



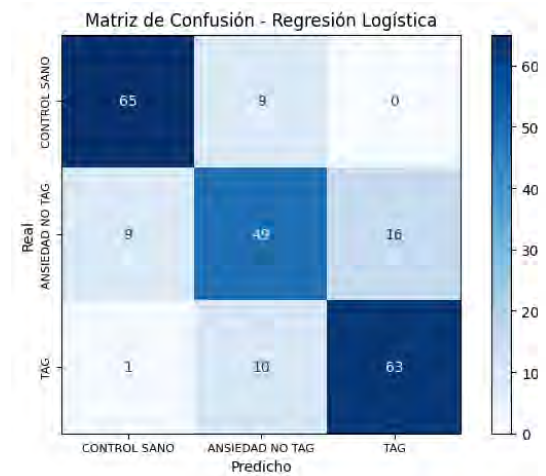
(iii) Modelo Random Forest



(iv) Modelo SVM

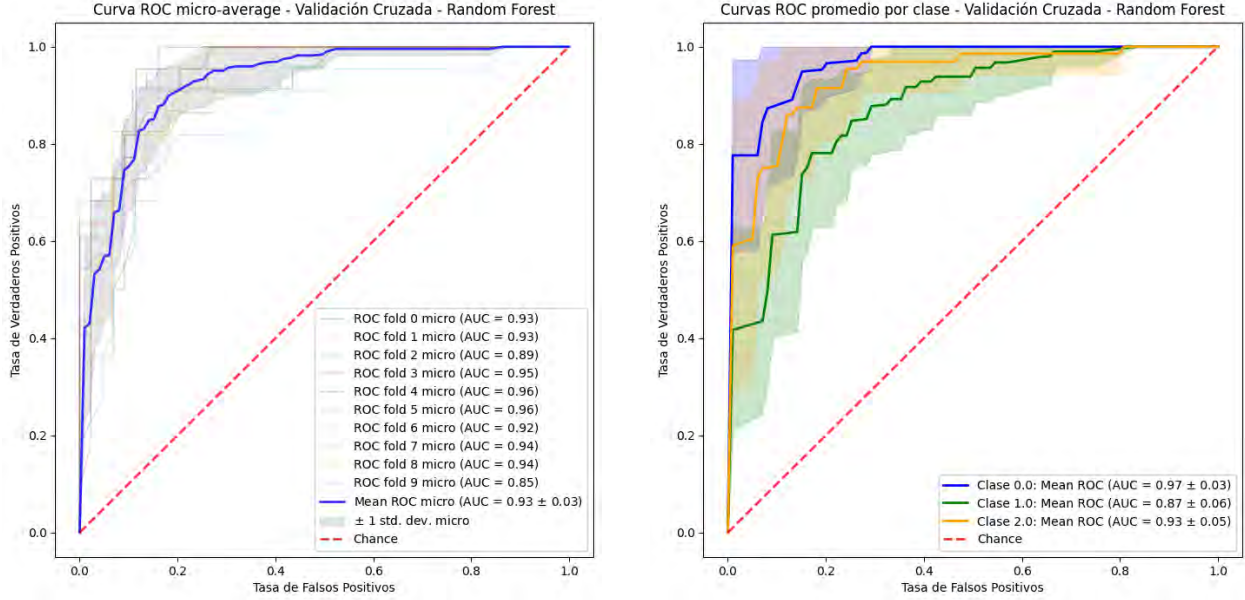


(v) Modelo Perceptrón Multicapa



(vi) Modelo Regresión Logística

Figura 4.8: Matrices de confusión de los modelos que utilizan el conjunto de datos balanceado con submuestreo aleatorio



(a) Curva ROC micro-promedio del rendimiento general y por pliegue

(b) Curva ROC *One-vs-Rest* macro-promedio por clase

Figura 4.9: Comparación de curvas ROC del modelo Random Forest

una estabilidad uniforme del modelo en la clasificación de las diferentes categorías. La mayor variabilidad se encuentra en la clase *ansiedad no TAG*, indicando que el modelo tiene dificultades para detectar correctamente esta clase a pesar de mostrar buen desempeño. En general, el desempeño del modelo es constante y con buenos resultados en los valores AUC (todos por encima de 0.85).

De forma general, Random Forest balanceado con submuestreo aleatorio fue el modelo con el mejor desempeño, alcanzando una exactitud del 81 %. Otros modelos como XGBoost y Regresión Logística obtuvieron valores máximos de exactitud del 81 % y 80 %, respectivamente.

## 4.5. Optimización del modelo

Los modelos implementados previamente fueron optimizados mediante tres técnicas: búsqueda en malla (*Grid Search*), búsqueda aleatoria (*Random Search*) y optimización bayesiana. Este proceso de optimización tuvo como objetivo identificar la configuración óptima de hiperparámetros que maximizara el rendimiento predictivo de cada algoritmo en la detección del Trastorno de Ansiedad Generalizada.

El espacio de búsqueda de hiperparámetros establecido para cada algoritmo se muestra en la Tabla 4.10, detallando los rangos y valores específicos considerados para cada parámetro relevante. Específicamente para Random Forest, se identificó que los hiperparámetros de `max_features`, `max_samples`, y `min_samples_leaf` son los más importantes y los que controlan la aleatoriedad de Random Forest (Bernardelli De Moraes y Palermo Coelho, 2023; Probst et al., 2019).

Tabla 4.10: Espacio de búsqueda utilizado para la optimización de hiperparámetros de los modelos de aprendizaje supervisado seleccionados

| Modelo               | Espacio de búsqueda de hiperparámetros                                                                                                                                   |
|----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| SVM                  | C: [0.1, 1, 10, 100], gamma: [0.001, 0.01, 0.1, 1], kernel: ['rbf', 'linear'], class_weight: [None, 'balanced']                                                          |
| KNN                  | n_neighbors: range(2, 11), weights: ['uniform', 'distance'], p: [1, 2]                                                                                                   |
| Naive Bayes          | var_smoothing: logspace(-11, -7, 5)                                                                                                                                      |
| XGBoost              | n_estimators: [50, 100, 200, 300], max_depth: [3, 5, 7, 10], learning_rate: [0.01, 0.05, 0.1, 0.2]                                                                       |
| Random Forest        | n_estimators: [100, 300, 500, 700, 1000], max_depth: [10, 30, 50, 70], max_samples: [0.5, 0.7, 0.9, 1.0], min_samples_leaf: [1, 2, 4, 8], max_features: ['sqrt', 'log2'] |
| Regresión Logística  | C: [0.01, 0.1, 1, 10]                                                                                                                                                    |
| Perceptrón Multicapa | hidden_layer_sizes: [(64,), (128,), (128, 64)], activation: ['relu', 'tanh'], alpha: [0.0001, 0.001, 0.01]                                                               |

La configuración óptima de hiperparámetros obtenida mediante la técnica de búsqueda en malla para cada uno de los modelos implementados se resume en la Tabla 4.11, mientras que en la Tabla 4.12 se presenta la configuración obtenida de la técnica de búsqueda aleatoria. Finalmente, la configuración óptima resultante de la técnica de optimización Bayesiana se muestra en la Tabla 4.13.

Los resultados obtenidos al evaluar el desempeño de los modelos optimizados con el conjunto de datos balanceado con submuestreo aleatorio se encuentran en la Tabla 4.14. Las métricas reportadas incluyen exactitud, precisión, sensibilidad y F1 Score.

Tabla 4.11: Configuración de hiperparámetros obtenida con la técnica de búsqueda en malla (*Grid Search*)

| Modelo               | Espacio de búsqueda de hiperparámetros                                                         |
|----------------------|------------------------------------------------------------------------------------------------|
| SVM                  | C: 100, gamma: 0.001, kernel: 'rbf', class_weight: None                                        |
| KNN                  | n_neighbors: 9, weights: 'distance', p: 1                                                      |
| Naive Bayes          | var_smoothing: 1e-11                                                                           |
| XGBoost              | n_estimators: 300, max_depth: 3, learning_rate: 0.05                                           |
| Random Forest        | n_estimators: 1000, max_depth: 30, max_samples: 1.0, min_samples_leaf: 1, max_features: 'log2' |
| Regresión Logística  | C: 1                                                                                           |
| Perceptrón Multicapa | hidden_layer_sizes: (128, 64), activation: 'relu', alpha: 0.0001                               |

Tabla 4.12: Configuración de hiperparámetros obtenida con la técnica de búsqueda aleatoria (*Random Search*)

| Modelo               | Espacio de búsqueda de hiperparámetros                                                         |
|----------------------|------------------------------------------------------------------------------------------------|
| SVM                  | C: 100, gamma: 0.001, kernel: 'rbf', class_weight: None                                        |
| KNN                  | n_neighbors: 9, weights: 'distance', p: 1                                                      |
| Naive Bayes          | var_smoothing: 1e-11                                                                           |
| XGBoost              | n_estimators: 300, max_depth: 3, learning_rate: 0.05                                           |
| Random Forest        | n_estimators: 1000, max_depth: 10, max_samples: 0.5, min_samples_leaf: 1, max_features: 'log2' |
| Regresión Logística  | C: 1                                                                                           |
| Perceptrón Multicapa | hidden_layer_sizes: (128, 64), activation: 'relu', alpha: 0.0001                               |

Tabla 4.13: Configuración de hiperparámetros obtenida con la técnica de optimización Bayesiana

| Modelo               | Espacio de búsqueda de hiperparámetros                                                        |
|----------------------|-----------------------------------------------------------------------------------------------|
| SVM                  | C: 100, gamma: 0.001, kernel: 'rbf', class_weight: None                                       |
| KNN                  | n_neighbors: 9, weights: 'distance', p: 1                                                     |
| Naive Bayes          | var_smoothing: 4.4e-11                                                                        |
| XGBoost              | n_estimators: 265, max_depth: 3, learning_rate: 0.05                                          |
| Random Forest        | n_estimators: 1000, max_depth: 5, max_samples: 0.9, min_samples_leaf: 1, max_features: 'sqrt' |
| Regresión Logística  | C: 3.3                                                                                        |
| Perceptrón Multicapa | hidden_layer_sizes: (100,), activation: 'relu', alpha: 0.01                                   |

Tabla 4.14: Resultados de las pruebas con los modelos optimizados utilizando el conjunto de datos balanceado

| Modelo               | Técnica de optimización       | Exactitud   | Precisión   | Sensibilidad | F1 Score    |
|----------------------|-------------------------------|-------------|-------------|--------------|-------------|
| SVM                  | Búsqueda en malla             | 0.80        | 0.80        | 0.80         | 0.80        |
|                      | Búsqueda aleatoria            | 0.80        | 0.80        | 0.80         | 0.80        |
|                      | Optimización Bayesiana        | 0.80        | 0.80        | 0.80         | 0.80        |
| KNN                  | Búsqueda en malla             | 0.65        | 0.65        | 0.65         | 0.65        |
|                      | Búsqueda aleatoria            | 0.65        | 0.65        | 0.65         | 0.65        |
|                      | Optimización Bayesiana        | 0.65        | 0.65        | 0.65         | 0.65        |
| Naive Bayes          | Búsqueda en malla             | 0.79        | 0.79        | 0.79         | 0.79        |
|                      | Búsqueda aleatoria            | 0.79        | 0.79        | 0.79         | 0.79        |
|                      | Optimización Bayesiana        | 0.79        | 0.79        | 0.79         | 0.79        |
| XGBoost              | Búsqueda en malla             | 0.83        | 0.83        | 0.83         | 0.83        |
|                      | Búsqueda aleatoria            | 0.83        | 0.83        | 0.83         | 0.83        |
|                      | Optimización Bayesiana        | 0.83        | 0.83        | 0.83         | 0.83        |
| <b>Random Forest</b> | <b>Búsqueda en malla</b>      | <b>0.84</b> | <b>0.84</b> | <b>0.84</b>  | <b>0.84</b> |
|                      | <b>Búsqueda aleatoria</b>     | <b>0.83</b> | <b>0.83</b> | <b>0.83</b>  | <b>0.83</b> |
|                      | <b>Optimización Bayesiana</b> | <b>0.83</b> | <b>0.83</b> | <b>0.83</b>  | <b>0.83</b> |
| Regresión Logística  | Búsqueda en malla             | 0.80        | 0.79        | 0.80         | 0.80        |
|                      | Búsqueda aleatoria            | 0.80        | 0.79        | 0.80         | 0.80        |
|                      | Optimización Bayesiana        | 0.80        | 0.79        | 0.80         | 0.80        |
| Perceptrón Multicapa | Búsqueda en malla             | 0.82        | 0.81        | 0.82         | 0.81        |
|                      | Búsqueda aleatoria            | 0.79        | 0.79        | 0.79         | 0.79        |
|                      | Optimización Bayesiana        | 0.79        | 0.79        | 0.79         | 0.79        |

De forma general, Random Forest fue el modelo con el mejor desempeño, alcanzando una exactitud del 84 % con búsqueda en malla. Otros modelos como XGBoost y Perceptrón Multicapa obtuvieron valores máximos de exactitud del 83 % y 82 %, respectivamente. El algoritmo con menor rendimiento fue KNN, lo cual sugiere una capacidad limitada de generalización para este conjunto de datos, incluso después de ajustar los hiperparámetros.

# Capítulo 5

## Validación y análisis de resultados

Este capítulo presenta los resultados de la validación del modelo de clasificación desarrollado para detectar el trastorno de ansiedad generalizada. La evaluación incluyó tanto el modelo Random Forest base (sin optimización de hiperparámetros) como sus versiones optimizadas mediante diferentes técnicas. El propósito fue identificar la configuración con el mejor rendimiento según las métricas de exactitud, sensibilidad, precisión y F1 score.

La validación se realizó utilizando un conjunto de datos independiente para evaluar de manera objetiva la capacidad de generalización del modelo. Como parte del proceso de evaluación, se generaron matrices de confusión, curvas ROC por clase y se calcularon las métricas de rendimiento correspondientes. Adicionalmente, se incorporaron los resultados del análisis de interpretabilidad mediante SHAP (*SHapley Additive exPlanations*), herramienta que permitió identificar las características con mayor influencia en las predicciones del modelo.

### 5.1. Metodología de validación

La metodología de validación del modelo de detección del trastorno de ansiedad generalizada (TAG) se estructuró en una fase principal, tal como se ilustra en la Figura 5.1. El proceso consistió en la recolección de datos mediante un grupo de estudio independiente, aplicando los mismos instrumentos psicológicos previamente seleccionados. Con estos nuevos datos se procedió a validar tanto el modelo Random Forest base sin optimización como las versiones con configuraciones optimizadas de hiperparámetros. Esta evaluación comparativa permitió identificar el modelo con superior capacidad de generalización basándose en el análisis de las métricas de rendimiento.

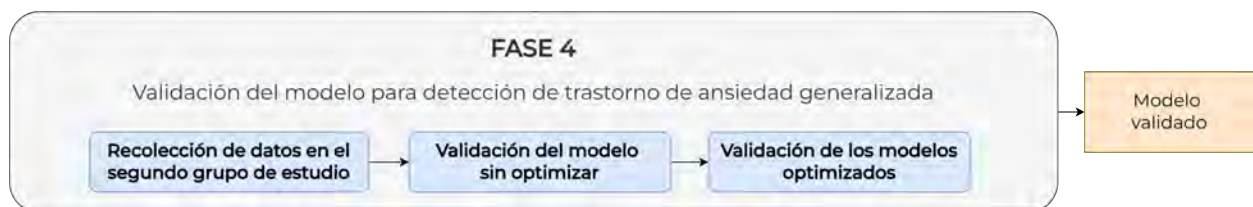


Figura 5.1: Metodología implementada para la validación modelo de detección de trastorno de ansiedad generalizada

## 5.2. Recolección del segundo conjunto de datos

La segunda recolección de datos se realizó en diversas ubicaciones de Cuernavaca, Morelos, con un total de 36 participantes. La edad promedio fue de 32.3 años ( $DE = 9.53$ ), con un rango de 18 a 56 años, y una distribución por sexo equilibrada, con un 52 % de mujeres. En cuanto al nivel educativo, el 47 % contaba con estudios universitarios concluidos y el 25 % con estudios de nivel medio superior. Respecto a la clasificación de ansiedad, 18 personas fueron identificadas con *control sano de ansiedad*, 10 con *ansiedad no de TAG* y 8 con indicios de *TAG*. La Tabla 5.1 presenta un resumen de las características sociodemográficas, el valor del Factor K y la categoría de ansiedad correspondiente a cada participante.

Tabla 5.1: Resumen de características sociodemográficas, factor K, y categoría de ansiedad de los participantes de la segunda recolección de datos

| Variables             |                                    | Frecuencia<br>n = 36 | Porcentaje (%) |
|-----------------------|------------------------------------|----------------------|----------------|
| Sexo                  | Mujeres                            | 19                   | 52.78          |
|                       | Hombres                            | 17                   | 47.22          |
| Edad, en años         | Media (D.E.)                       | $32.30 \pm 9.53$     |                |
| Factor K              | Media (D.E.)                       | $0.60 \pm 0.26$      |                |
| Categoría de ansiedad | Control sano de ansiedad           | 18                   | 50.00          |
|                       | Ansiedad no TAG                    | 10                   | 27.78          |
|                       | Trastorno de ansiedad generalizada | 8                    | 22.22          |

## 5.3. Validación del modelo

En esta sección se presentan los resultados obtenidos de las validaciones del modelo Random Forest sin optimizar y optimizado, utilizando el segundo conjunto de datos que no ha sido utilizado en el entrenamiento o las pruebas. Al desarrollar un modelo de aprendizaje automático es necesario realizar una validación con un nuevo conjunto de datos para confirmar su capacidad de generalización y evitar el sobreajuste (Vos et al., 2024). En general, este proceso permite evaluar el rendimiento del modelo en diferentes poblaciones y contextos.

### Validación del modelo Random Forest sin optimizar

La evaluación del modelo Random Forest se realizó inicialmente sin cambios en los ajustes de los hiperparámetros, con el objetivo de establecer una línea base de comparación respecto a las configuración optimizadas con diferentes técnicas. La validación se llevó a cabo con un conjunto de prueba independiente, obtenido de una segunda recolección, y manteniendo las clases originales a detectar: *control sano de ansiedad*, *ansiedad no TAG* y *TAG*.

La Figura 5.2 muestra la matriz de confusión correspondiente a los resultados de la validación. El modelo identificó correctamente 15 de los 18 casos de *control sano de ansiedad* (83.3 % de aciertos), con 3 casos clasificados erróneamente como *ansiedad no TAG*. Para la clase *ansiedad no TAG*, únicamente se clasificaron correctamente 4 de 10 casos (40 %), siendo 6 casos confundidos con *TAG*. Finalmente, la clase *TAG* fue detectada con mayor exactitud, detectando correctamente 7 de 8 casos (87.5 %), con un caso clasificado incorrectamente como *ansiedad no TAG*.

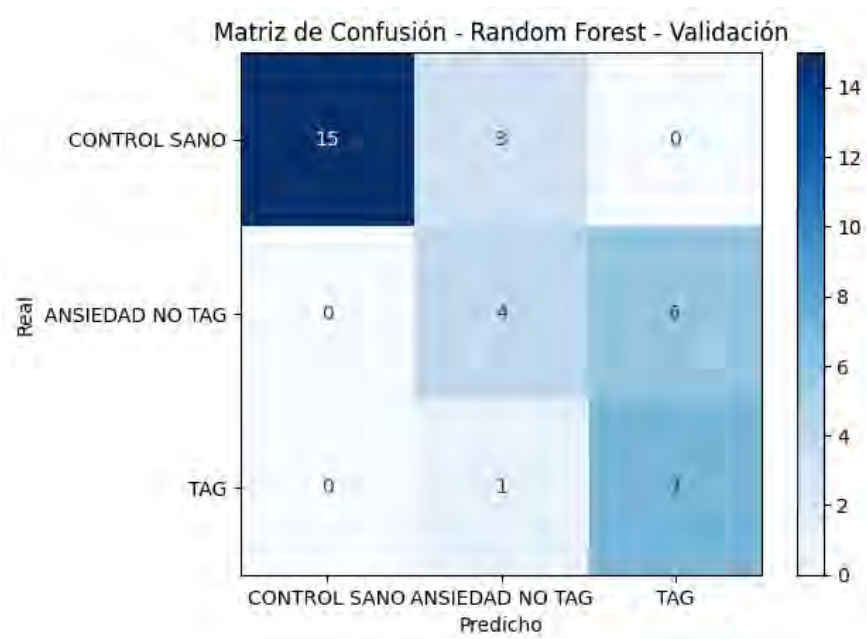


Figura 5.2: Matriz de confusión de la validación del Modelo Random Forest sin optimizar

Los resultados obtenidos en las métricas de evaluación se muestran en la Tabla 5.2. Este modelo alcanzó una exactitud del 72 % y una precisión del 76 %. La clase *control sano* no presentó falsos positivos, mientras que la clase *TAG* generó una cantidad considerable de falsos positivos y es la que acumuló el mayor número de errores de clasificación. A pesar de ello, las clases *control sano* y *TAG* lograron superar el 80 % de aciertos en sus predicciones.

Tabla 5.2: Resultados de las métricas de evaluación en la validación del modelo Random Forest sin optimizar

| Exactitud | Precisión | Sensibilidad | F1 Score | AUC-ROC (One vs Rest) | AUC-PR (One vs Rest) |
|-----------|-----------|--------------|----------|-----------------------|----------------------|
| 0.72      | 0.76      | 0.72         | 0.73     | 0.89                  | 0.76                 |

La Figura 5.3 muestra las curvas ROC individuales por clase, junto con el promedio macro. La clase *control sano* obtuvo un valor AUC-ROC de 0.99, evidenciando una alta capacidad para esta clase de las demás. La clase *TAG* también mostró un buen rendimiento (AUC = 0.95), mientras que la clase *ansiedad no TAG* fue la más difícil de clasificar correctamente, con un AUC de apenas 0.73. Esto sugiere que el modelo tiene mayor dificultad para identificar patrones distintivos en los casos intermedios de ansiedad, ya que comparten

características con las otras dos clases. El valor macro promedio de AUC-PR fue de  $0.80 \pm 0.16$ , lo cual muestra una variabilidad considerable en el modelo.

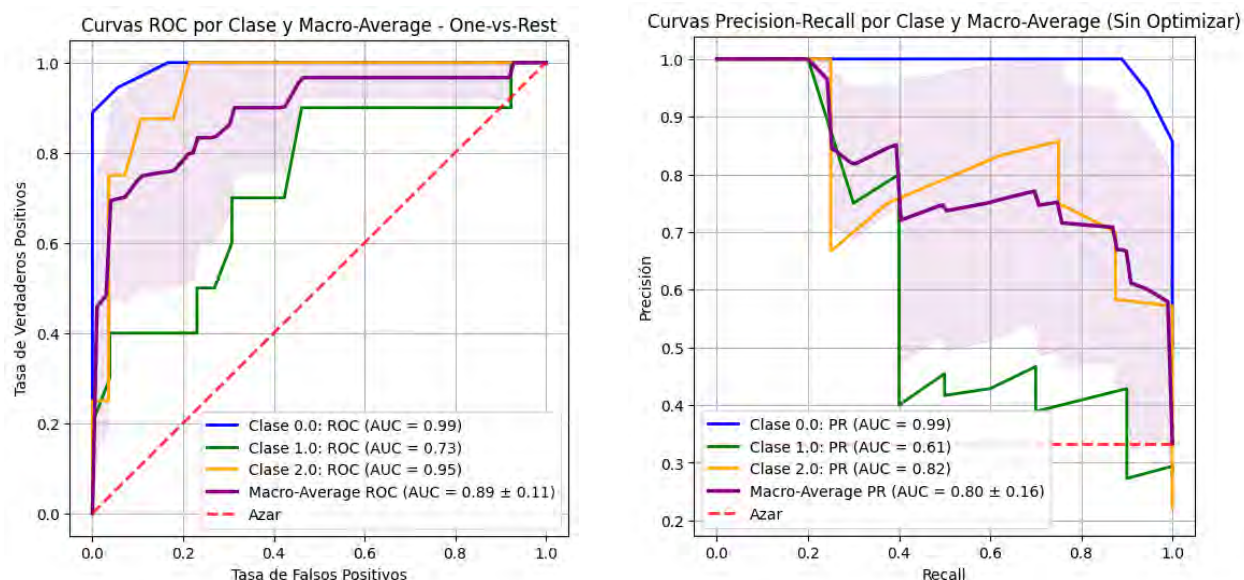


Figura 5.3: Evaluación del modelo Random Forest sin optimizar mediante curvas ROC macro-promedio y Precision-Recall (*One-vs-Rest*)

En resumen, aunque el modelo Random Forest sin optimizar muestra un rendimiento adecuado, especialmente en las clases de extremos (*control sano* y *TAG*), existen áreas de mejora, en particular para la clase *ansiedad no TAG*. Estos resultados justifican la implementación de técnicas de optimización de hiperparámetros para mejorar la sensibilidad y la exactitud del modelo.

## Validación del modelo Random Forest optimizado

Con el objetivo de mejorar el desempeño del modelo Random Forest, se aplicaron tres técnicas de optimización de hiperparámetros: búsqueda en malla (*grid search*), búsqueda aleatoria (*random search*) y optimización bayesiana. Estas estrategias permitieron ajustar de forma más precisa los parámetros clave del modelo, como la profundidad máxima de los árboles, el número de estimadores, y los criterios de división, entre otros.

La evaluación de los modelos optimizados se realizó sobre el mismo conjunto de prueba empleado en la validación del modelo base. En la Tabla 5.3 se presentan las métricas obtenidas por cada técnica, comparadas también con los resultados del modelo sin optimización.

La Figura 5.4 muestra la matriz de confusión del modelo optimizado con búsqueda en malla. Al igual que en el modelo base, se identificaron correctamente 15 de los 18 casos de la clase *control sano de ansiedad*. En contraste con el modelo sin optimización, se observó una mejora en la clasificación de los casos de *ansiedad no TAG*, con 5 instancias correctamente clasificadas y 5 confundidas con *TAG*, lo cual representa una ligera mejora en el equilibrio de clasificación de esta clase intermedia. Finalmente, los 8 casos correspondientes a la clase

Tabla 5.3: Resultados de la validación con el segundo conjunto de datos de cada modelo Random Forest optimizado

| Técnica de optimización   | Exactitud   | Precisión   | Sensibilidad | F1 Score    | AUC-ROC<br>(One vs Rest) | AUC-PR<br>(One vs Rest) |
|---------------------------|-------------|-------------|--------------|-------------|--------------------------|-------------------------|
| Sin técnica               | 0.72        | 0.76        | 0.72         | 0.73        | 0.89                     | 0.80                    |
| Búsqueda en Malla         | 0.78        | 0.75        | 0.78         | 0.74        | 0.90                     | 0.81                    |
| <b>Búsqueda Aleatoria</b> | <b>0.81</b> | <b>0.75</b> | <b>0.78</b>  | <b>0.74</b> | <b>0.91</b>              | <b>0.83</b>             |
| Optimización Bayesiana    | 0.78        | 0.75        | 0.78         | 0.74        | 0.91                     | 0.81                    |

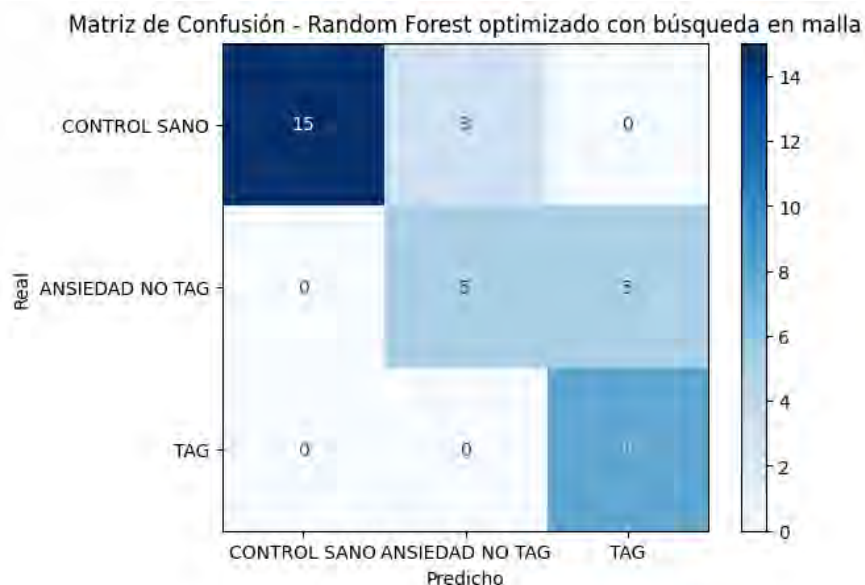


Figura 5.4: Matriz de confusión de la validación del Modelo Random Forest optimizado con Búsqueda en Malla

*TAG* fueron correctamente identificados, sin falsos negativos, lo cual representa una mejora notable frente a la versión no optimizada.

La Figura 5.5 presenta las curvas ROC y *Precision-Recall* generadas para cada clase. Se observan mejoras consistentes en el área bajo la curva (AUC), particularmente para la clase *TAG*, que alcanza un valor de 0.96, y la clase *control sano*, con 0.99. La clase *ansiedad no TAG* mantiene un AUC de 0.76, ligeramente superior al modelo base. El promedio macro de las curvas ROC es de  $0.90 \pm 0.10$ , lo que refleja un buen comportamiento general del modelo optimizado, especialmente en la correcta identificación de la clase *TAG*. El valor macro promedio de AUC-PR fue de  $0.81 \pm 0.15$ , lo cual representa una mejora respecto al modelo base y respalda la utilidad de la optimización.

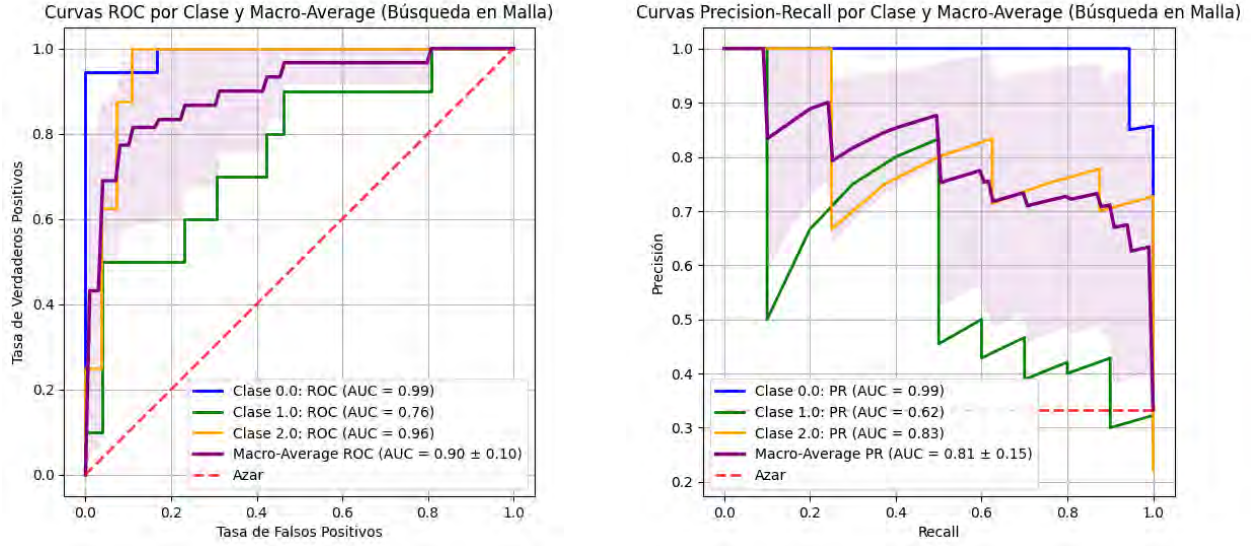


Figura 5.5: Evaluación del modelo Random Forest optimizado con búsqueda en malla mediante curvas ROC y Precision-Recall (*One-vs-Rest*)

En la Figura 5.6 se presenta la matriz de confusión correspondiente al modelo optimizado con búsqueda aleatoria. En comparación con las versiones anteriores, se observan mejoras importantes en la clasificación de la clase *control sano*, con 16 de 18 casos correctamente identificados, y solo 2 clasificaciones erróneas como *ansiedad no TAG*. La clase *ansiedad no TAG* mantuvo el mismo desempeño que en el modelo con búsqueda en malla, con 5 casos correctamente identificados y 5 clasificados erróneamente como *TAG*. Finalmente, todos los casos de la clase *TAG* fueron correctamente clasificados, mostrando una perfecta sensibilidad para esta clase.

La Figura 5.7 muestra las curvas ROC y PR para cada clase. Se observan altos valores de AUC, en particular en la clase de *control sano* (AUC = 0.99) y *TAG* (AUC = 0.96). Estas cifras muestran una ligera mejora respecto a la búsqueda en malla, especialmente en la clase *ansiedad no TAG*. La consistencia de estas curvas sugiere un buen desempeño para diferenciar las clases en el modelo optimizado. Se mantiene la dificultad para separar correctamente los casos de *ansiedad no TAG*, los cuales, debido a su posición intermedia entre las otras dos clases, presentan patrones más ambiguos. Sin embargo, el rendimiento en las clases más relevantes, como *TAG*, se mantiene elevado.

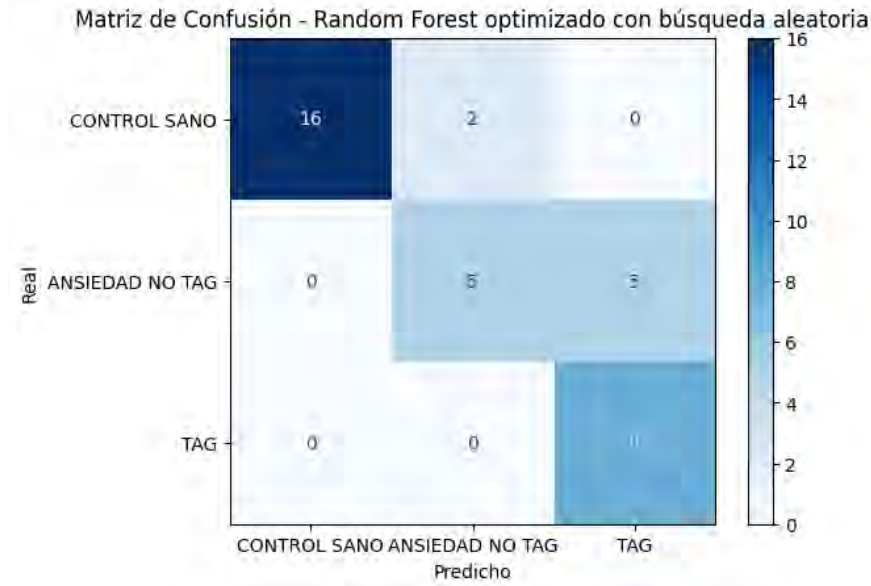


Figura 5.6: Matriz de confusión de la validación del Modelo Random Forest optimizado con Búsqueda Aleatoria

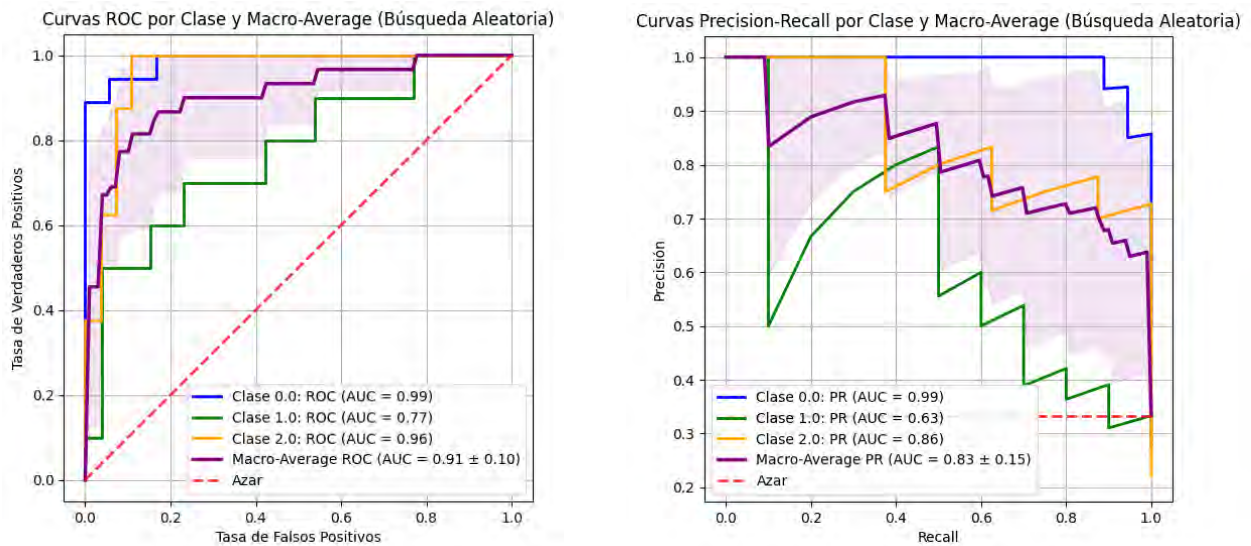


Figura 5.7: Evaluación del modelo Random Forest optimizado con búsqueda aleatoria mediante curvas ROC y Precision-Recall (*One-vs-Rest*)

En la Figura 5.8 se presenta la matriz de confusión generada por el modelo optimizado mediante optimización bayesiana. El desempeño es comparable al observado en la búsqueda en malla: se identificaron correctamente 15 de los 18 casos de la clase *control sano*, con 3 clasificaciones incorrectas como *ansiedad no TAG*. La clase *ansiedad no TAG* mostró una clasificación con 5 casos correctamente clasificados y 5 confundidos con la clase *TAG*. Finalmente, al igual que en los modelos anteriores, los 8 casos de la clase *TAG* fueron correctamente clasificados, lo que indica una sensibilidad perfecta para esta clase.

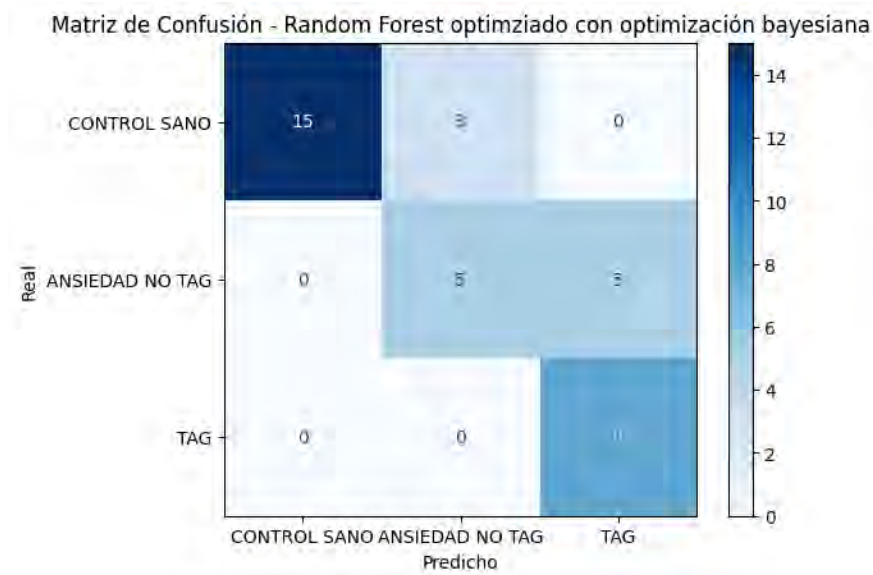


Figura 5.8: Matriz de confusión de la validación del Modelo Random Forest optimizado con Optimización Bayesiana

La Figura 5.9 muestra las curvas ROC y PR para cada clase. Los valores de AUC alcanzados fueron en la clase de *control sano* de 0.99, en la clase de *ansiedad no TAG* de 0.78 y, finalmente, en *TAG* se obtuvo un valor de 0.96. Estos resultados se encuentran al mismo nivel que los obtenidos mediante búsqueda aleatoria, mostrando que la optimización bayesiana logra un buen equilibrio en la clasificación multiclase. El macro-promedio del AUC-PR fue de  $0.81 \pm 0.15$ , muy similar al de la búsqueda en malla. La clase con mayor ambigüedad sigue siendo *ansiedad no TAG* (AUC-PR = 0.63).

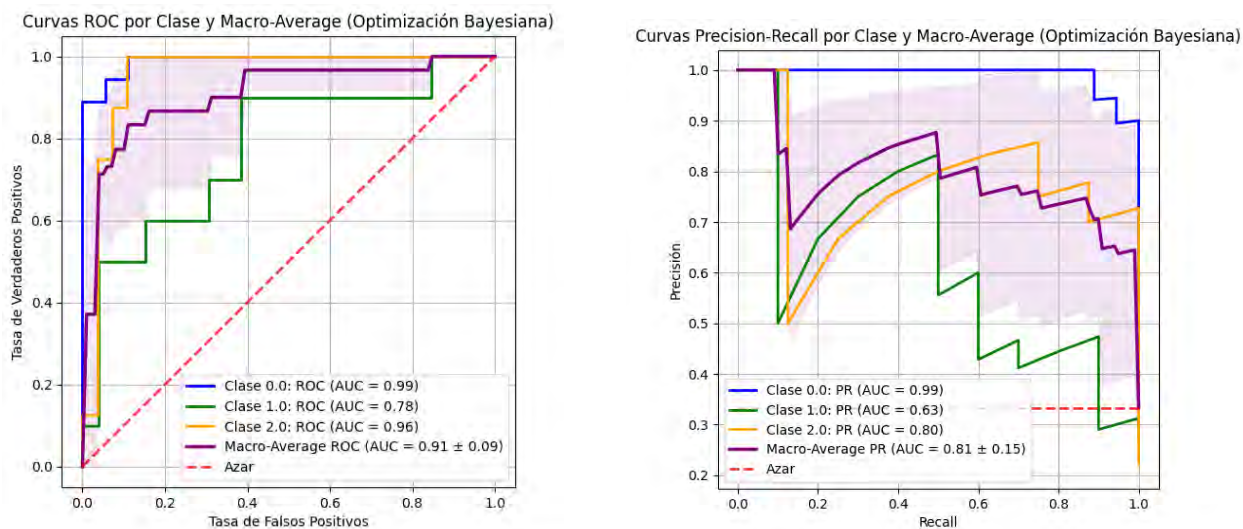


Figura 5.9: Evaluación del modelo Random Forest optimizado con optimización bayesiana mediante curvas ROC y Precision-Recall (*One-vs-Rest*)

Entre los modelos optimizados, la búsqueda aleatoria presentó el mejor rendimiento general, alcanzando una exactitud balanceada de 80 %, una AUC-ROC de 0.91 y una AUC-PR de 0.83, superando al modelo base en todas las métricas evaluadas. Aunque las diferencias en precisión y F1 score no fueron significativas, el aumento en la sensibilidad y las áreas bajo la curva indican una mejora en la capacidad del modelo para distinguir entre clases, especialmente en contextos con clases desbalanceadas.

## 5.4. Interpretación del modelo Random Forest

La interpretabilidad de los modelos en el aprendizaje automático se refiere a la capacidad explicar y entender cómo toman decisiones dichos modelos, siendo un elemento valioso para obtener información sobre el comportamiento del modelo y tomar decisiones basadas en los datos (Shah et al., 2020). Entre las técnicas de interpretabilidad utilizadas en diferentes dominios se encuentran LIME (*Local Interpretable Model-agnostic Explanations*) y SHAP (*SHapley Additive exPlanations*) y se ha mostrado que son efectivas en identificar las características más importantes en un modelo (Y y Challa, 2023). Además, Random Forest cuenta con una característica incorporada para estimar los puntajes de importancia de las variables utilizadas para el modelo, lo que ayuda a interpretar el comportamiento del modelo según el orden y tamaño del efecto de cada variable (Orlenko y Moore, 2021).

Se realizó un análisis del modelo para detección de TAG que fue optimizado con la técnica de Búsqueda Aleatoria para identificar las características más relevantes que contribuyen a la detección de esta clase. Para comprender cómo el modelo está tomando decisiones y los factores que se consideran más determinantes en la identificación de TAG, se utilizó el puntaje de importancia de características proporcionada por Random Forest y el métodos de SHAP (*SHapley Additive exPlanations*).

### Importancia de características

Se obtuvo tanto la medida de importancia global de las características como los valores por cada clase. El gráfico de importancia global (Figura 5.10) muestra que las características de *preocupación excesiva*, *estado de tensión* y *factor K* tienen el mayor peso en las predicciones del modelo, representado en conjunto aproximadamente el 40 % de la importancia total. Esto sugiere que el modelo se basa principalmente en estas tres variables para realizar sus predicciones.

Por otro lado, el gráfico de importancia por clases (Figura 5.11) reveló que las variables de *preocupación excesiva* y *estados de tensión* son las más importantes para las tres clases, siendo *preocupación excesiva* la de más peso para la clase de *control sano de ansiedad* y *estados de tensión* más importante para *ansiedad no TAG*. La variable *factor K* y *decisiones impulsivas* tienen mayor peso en la clase de *control sano de ansiedad*. Algunas características de relaciones interpersonales como el interés por el bienestar de los parientes sanguíneos, marcar distancia en una relación romántica y la madurez sexual también tienen relevancia en la clasificación.

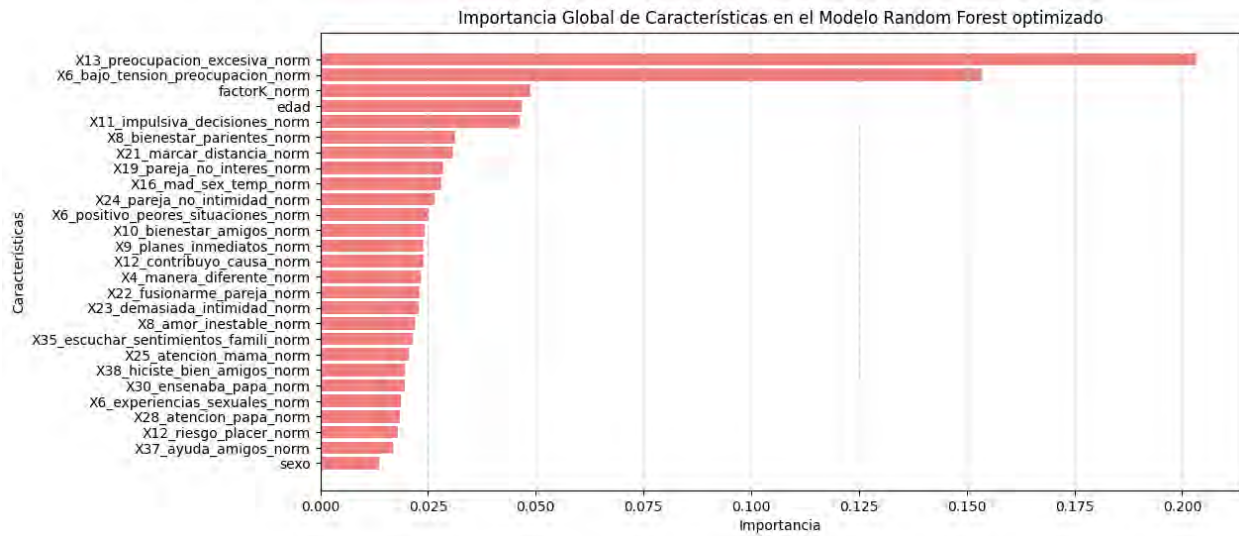


Figura 5.10: Importancia global de características del modelo Random Forest optimizado con búsqueda aleatoria

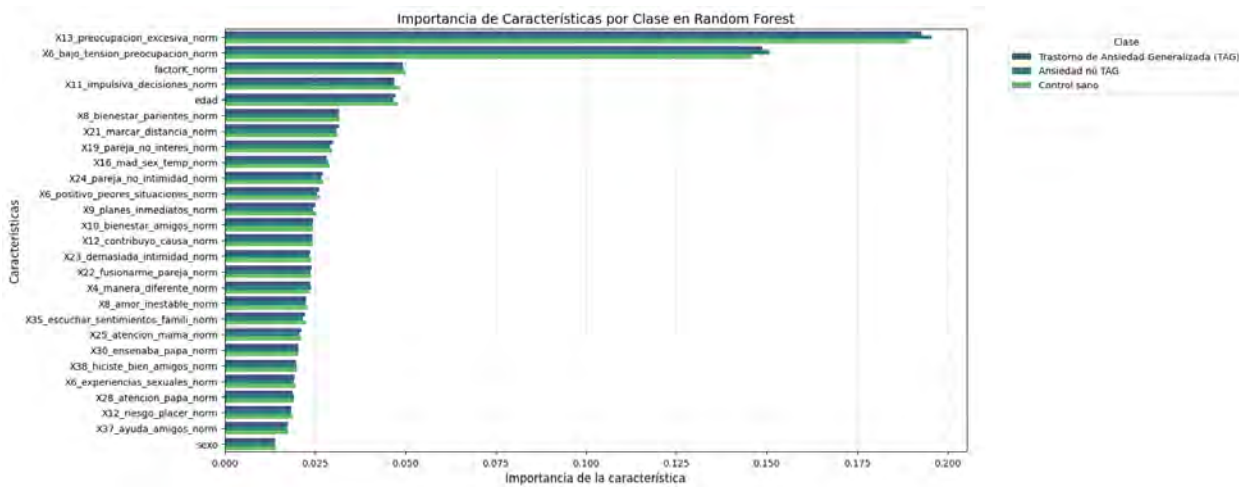


Figura 5.11: Importancia de características por clase del modelo Random Forest optimizado con búsqueda aleatoria

## Análisis SHAP

El enfoque en el análisis SHAP fue sobre la clase de TAG para identificar las características que tienen un impacto positivo o negativo en la clasificación. El gráfico de resumen de los valores SHAP (Figura 5.12) muestra que las cinco características más importantes y que impactan sobre las decisiones de clasificación en el modelo son *preocupación excesiva*, *estado de tensión*, *decisiones impulsivas*, *marcar distancia con la pareja* y *edad*.

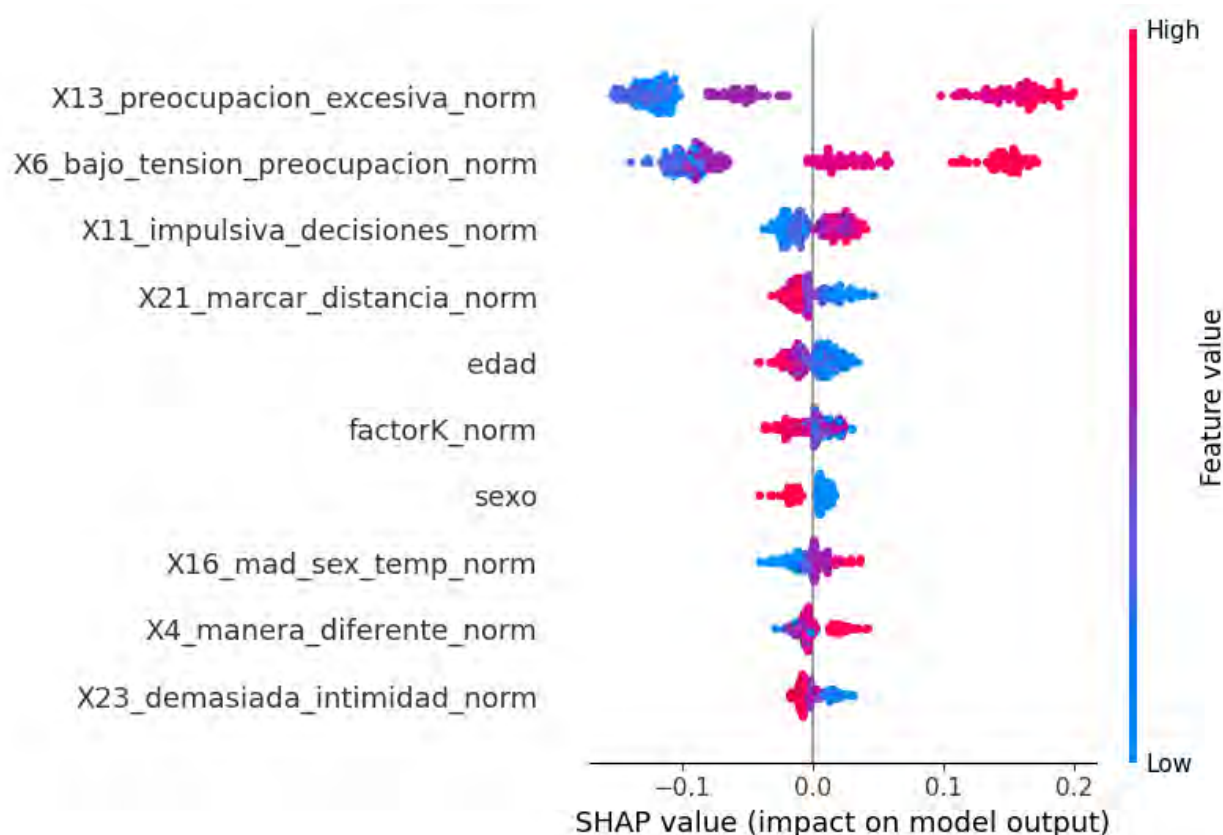


Figura 5.12: Resumen de los valores SHAP obtenidos del modelo Random Forest optimizado con búsqueda aleatoria para la clase TAG

Al realizar los gráficos de dependencia de las primeras diez variables con mayor impacto para la clase *TAG*, se encontró lo siguiente sobre las características sociodemográficas, estrategias de historia de vida y preocupación:

- **Características sociodemográficas:** los gráficos de dependencia de edad y sexo presentados en la Figura 5.13 muestran que, en cuanto al sexo, las mujeres (codificadas con valor 0) presentan valores SHAP positivos, lo que indica una mayor probabilidad de ser clasificadas con TAG según el modelo. Respecto a la edad, se identifica una tendencia en los adultos jóvenes (18-30 años), donde los valores SHAP son positivos, señalando un incremento en la probabilidad de TAG en este grupo. En contraste, a partir de los 30 años se observa una inversión de esta tendencia, con valores SHAP negativos que indican una disminución progresiva en la probabilidad de TAG conforme aumenta la edad.

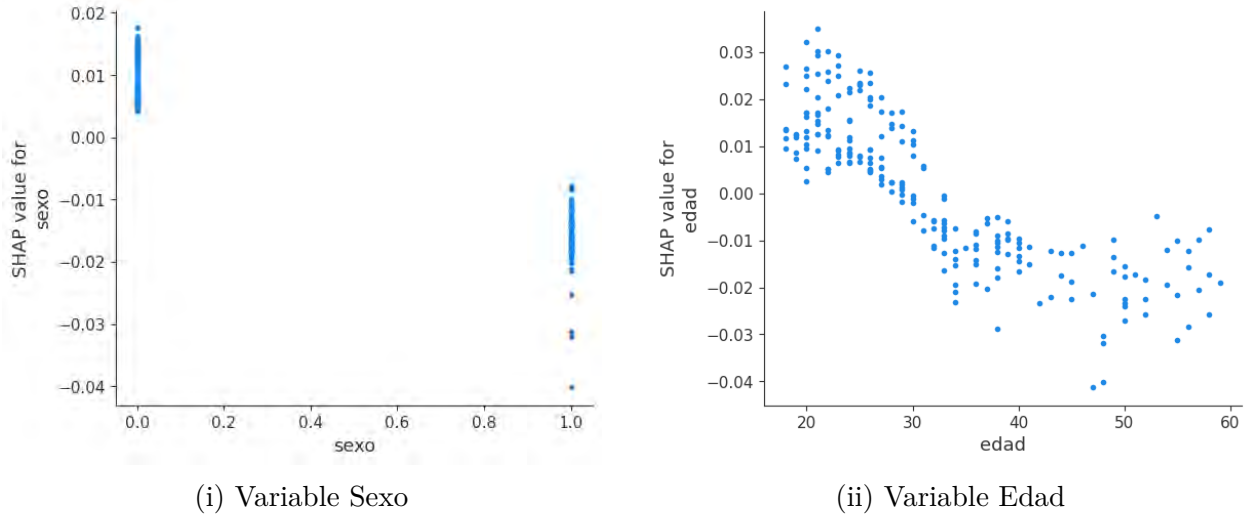
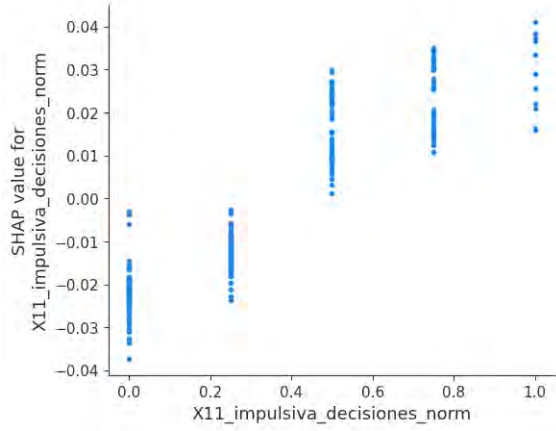
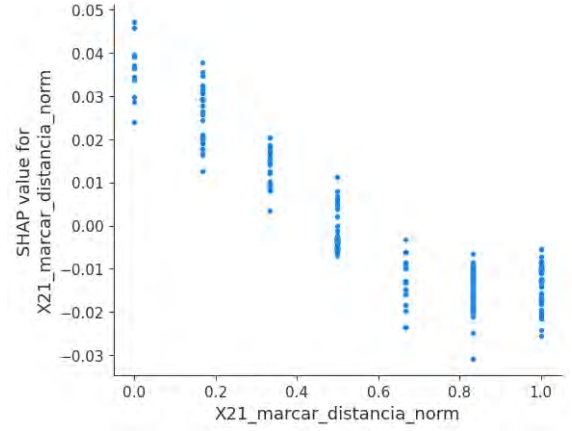


Figura 5.13: Gráficos de dependencia de las características sociodemográficas del modelo y los valores SHAP

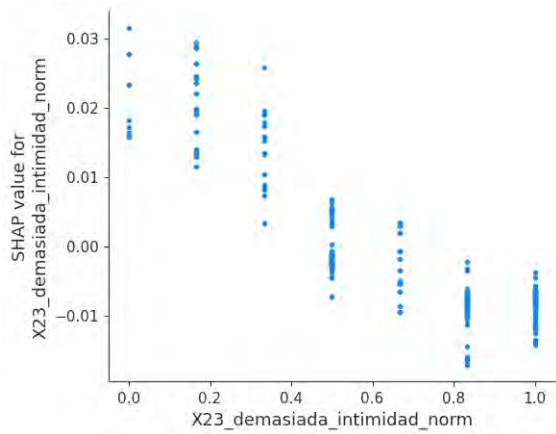
- Estrategias de historia de vida:** se generaron los gráficos de dependencia de seis variables (Figura 5.14). La probabilidad de TAG aumenta con mayor impulsividad (valores  $\geq 0.8$ ), mayor percepción de distancia afectiva con la pareja (valores  $\leq 0.2$ ), y temor a relaciones con demasiada intimidad afectiva (valores  $\leq 0.2$ ). La madurez sexual temprana muestra una tendencia incremental en valores SHAP, sugiriendo que los cambios físicos, hormonales y emocionales tempranos incrementan la probabilidad de TAG. La variable de frecuencia en que los amigos externaron que se hicieron bien las cosas muestra que las personas que indicaron que nunca les han dicho que hacen las cosas bien tienen mayor probabilidad de ser clasificados como TAG, mientras que aquellas que consideraron que se los dicen mucho tienen menor probabilidad de ser clasificados como TAG. Finalmente, para el factor K (indicador de estrategia de historia de vida), valores bajos ( $< 0.2$ , estrategia rápida) incrementan ligeramente la probabilidad de TAG, con otro pico entre 0.5-0.6, mientras que valores  $\geq 0.7$  (estrategia lenta) la disminuyen notablemente.



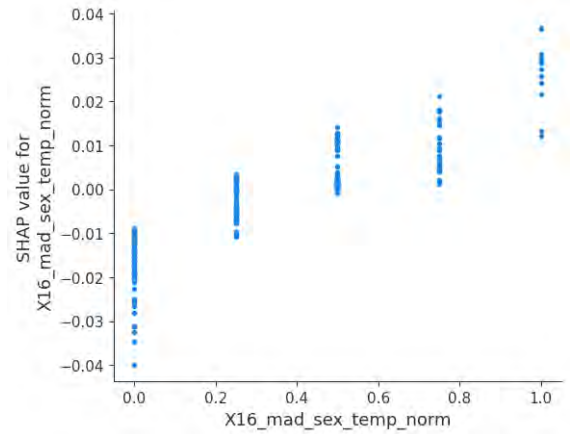
(i) Variable Decisiones impulsivas



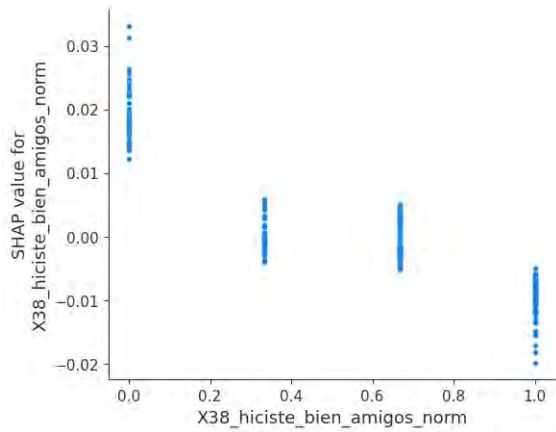
(ii) Variable Marcar distancia con la pareja



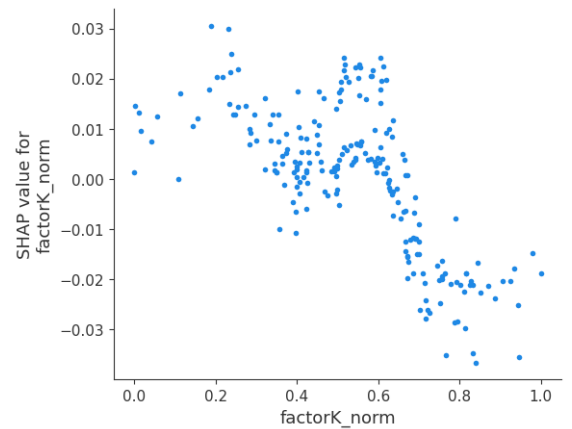
(iii) Variable Demasiada intimidad con la pareja



(iv) Variable Madurez sexual temprana



(v) Variable Frecuencia en que los amigos externaron que se hicieron bien las cosas



(vi) Variable Factor K normalizado

Figura 5.14: Gráficos de dependencia de las características de estrategias de historia de vida del modelo y los valores SHAP

- **Preocupación:** se generaron los gráficos de dependencia de preocupación excesiva y estado de tensión (Figura 5.15) donde se muestra que la percepción de preocuparse en exceso por cosas de la vida diaria y que el estrés o tensión sostenida son variables que impactan en la probabilidad de clasificar a la persona como TAG (valores de 0.8 a 1).

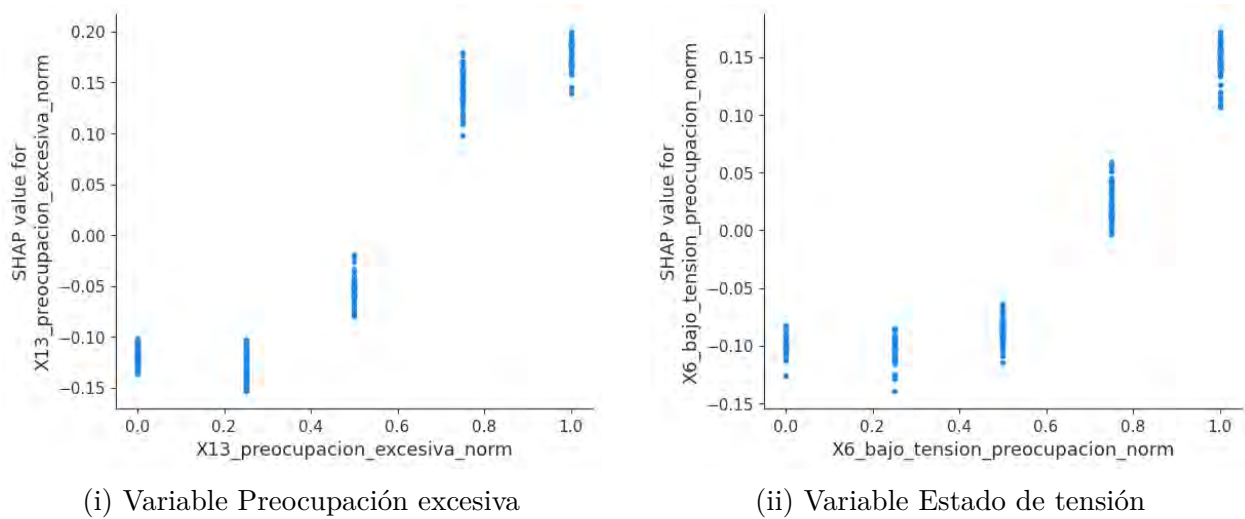


Figura 5.15: Gráficos de dependencia de las variables de preocupación y los valores SHAP

A partir de los resultados de la interpretabilidad del modelo, se identificó que las características más importantes que comparte el grupo que fue clasificado como TAG y que contribuyen a la detección de este trastorno en el modelo desarrollado son:

1. Sexo (en particular mujeres)
2. Edad entre 18 a 30 años
3. Rasgos de estrategias de historia de vida rápida como impulsividad, maduración sexual rápida, apego romántico inestable y factor K bajo.
4. Poca validación externa.

Además, la característica de la preocupación excesiva y sostenida es determinante para la clasificación en este trastorno de ansiedad específico.

# Capítulo 6

## Conclusiones y trabajos futuros

En este capítulo se explican las conclusiones derivadas de este trabajo de investigación, resaltando los hallazgos más relevantes, aportaciones en el área de ciencias computacionales y las limitaciones enfrentadas. Además, se describen las líneas futuras de investigación y los productos académicos obtenidos del desarrollo del tema de tesis.

### 6.1. Conclusiones

Esta tesis tuvo como objetivo desarrollar un modelo de aprendizaje automático capaz de detectar el trastorno de ansiedad generalizada a partir de datos obtenidos mediante instrumentos que evaluaran aspectos de estrategias de historia de vida, preocupación y variables sociodemográficas. Para ello, se implementó una metodología que incluyó la recolección, preprocesamiento y análisis de datos, así como la evaluación comparativa de diferentes algoritmos de clasificación.

El modelo sin optimizar que mostró mejor desempeño en la validación fue Random Forest, con una exactitud balanceada del 70 % y una sensibilidad del 72 %, logrando identificar correctamente el 87.5 % de los casos de TAG en el conjunto de 36 participantes. Al aplicar técnicas de optimización de hiperparámetros, se observó una mejora generalizada en el desempeño de todos los modelos. En particular, Random Forest y XGBoost destacaron con resultados de 84 % y 83 % de exactitud, respectivamente, utilizando configuraciones derivadas de búsqueda en malla. El modelo con mejor desempeño general en la validación fue Random Forest optimizado con búsqueda aleatoria, alcanzando una exactitud balanceada del 80 % y una sensibilidad del 78 % en la validación. Los análisis de comparación revelaron mejoras significativas en el rendimiento del modelo optimizado, especialmente en la detección de casos de TAG y control sano de ansiedad.

Las variables relacionadas con las estrategias de historia de vida, como el factor K, la impulsividad y la madurez sexual temprana, fueron especialmente relevantes para la detección del TAG. Entre las limitaciones de este trabajo se encuentran el tamaño reducido de la muestra, así como el uso de cuestionarios autoadministrados, lo cual puede introducir sesgos de respuesta. Además, los resultados obtenidos podrían no ser directamente generalizables a otras poblaciones sin una validación externa adicional.

La incorporación de variables asociadas a las estrategias de historia de vida ha demostrado ser un enfoque eficaz para la detección automatizada del trastorno de ansiedad generalizada (TAG) al aportar una visión integral que contempla factores conductuales, cognitivos y sociodemográficos. El modelo de Random Forest desarrollado mostró un desempeño adecuado que puede complementar el diagnóstico clínico y evaluaciones psicológicas.

## 6.2. Trabajos futuros

En función de los resultados obtenidos en esta investigación, se identificaron mejoras que podrían implementarse como trabajo a futuro que permita mejorar el desempeño del modelo de detección desarrollado. Se propone ampliar tanto el número como la diversidad de variables relacionadas con estrategias de historia de vida, con el objetivo de identificar aquellas que mejoren la precisión en la clasificación de personas con ansiedad intermedia (no TAG), categoría que actualmente presenta la mayor dificultad de detección.

Adicionalmente, se recomienda expandir el conjunto de datos incorporando muestras más diversas y poblaciones de contextos socioculturales distintos. Esto permitiría mejorar la efectividad en diferentes entornos culturales y demográficos. Finalmente, se sugiere explorar versiones optimizadas del modelo que utilicen un número reducido de predictores clave, manteniendo o superando los niveles actuales de exactitud y sensibilidad.

## 6.3. Productos derivados de la investigación

Como resultado del desarrollo de esta tesis se generaron diversos productos académicos que contribuyeron a la difusión del conocimiento sobre los hallazgos obtenidos, reflejando el trabajo realizado y su aplicación en el ámbito computacional y de salud mental. A continuación, se describen los principales productos derivados de esta investigación.

### Artículos

- Islas Ávila, A. L., Martínez Rebollar, A., Castrejón Salgado, R., Ávila Jiménez, L., & Estrada Esquivel, H. (2025). *Detección automática de trastorno de ansiedad generalizada: un enfoque de estrategias de historia de vida*. Revista Mexicana de Ingeniería Biomédica.
- Islas Ávila, A. L., Martínez Rebollar, A., Castrejón Salgado, R., Ávila Jiménez, L., & Sánchez Gutiérrez, A. F. (2024). *Analysis of Generalized Anxiety Disorder and Life History Strategies in beneficiaries affiliated with UMF No. 20, IMSS Morelos*. Tecnología y Ciencia Aplicada, 7(2), 452–457.  
[https://jcyta.cenidet.tecnm.mx/revistas/tycap/13-Vol\\_7\\_Num\\_2\\_Julio-Diciembre\\_2024.pdf](https://jcyta.cenidet.tecnm.mx/revistas/tycap/13-Vol_7_Num_2_Julio-Diciembre_2024.pdf)

## Póster

- Islas Avila, Ana L., Martínez Rebollar, Alicia, Estrada Esquivel, Hugo. (Noviembre 2023). Análisis comparativo de algoritmos de aprendizaje enfocado a un caso de estudio de salud mental [Póster]. 1er Congreso Internacional de Tecnología y Ciencia Aplicada, Cuernavaca, Morelos.

## Protocolo de investigación

- El trabajo de investigación fue evaluado y aprobado como protocolo de investigación por el Comité de Ética en Investigación y el Comité Local de Investigación del IMSS de Morelos, obteniendo el número de registro **R-2024-1701-029**.

# Bibliografía

- Ai, F., Li, E., Ji, Q., & Zhang, H. (2024). Construction of a machine learning-based risk prediction model for depression in middle-aged and elderly hypertensive people in China: a longitudinal study. *Frontiers in Psychiatry*, 15, 1398596. <https://doi.org/10.3389/fpsyt.2024.1398596>
- Alharthi, H. (2020). Predicting the level of generalized anxiety disorder of the coronavirus pandemic among college age students using artificial intelligence technology. *2020 19th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)*, 218-221. <https://doi.org/10.1109/DCABES50732.2020.00064>
- American Psychiatric Association. (2013, 22 de mayo). *Diagnostic and statistical manual of mental disorders* (Fifth Edition). <https://doi.org/10.1176/appi.books.9780890425596>
- Andonie, R. (2019). Hyperparameter optimization in learning systems. *Journal of Membrane Computing*, 1(4), 279-291. <https://doi.org/10.1007/s41965-019-00023-0>
- Baba, A., & Bunji, K. (2023). Prediction of Mental Health Problem Using Annual Student Health Survey: Machine Learning Approach [Company: JMIR Mental Health Distributor: JMIR Mental Health Institution: JMIR Mental Health Label: JMIR Mental Health Publisher: JMIR Publications Inc., Toronto, Canada]. *JMIR Mental Health*, 10(1), e42420. <https://doi.org/10.2196/42420>
- Bandalos, D. L. (1996). Confirmatory factor analysis. En *Applied multivariate statistics for the social sciences* (3.<sup>a</sup> ed., pp. 389-420). Lawrence Erlbaum Associates.
- Bergstra, J., & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, 13, 281-305. <https://jmlr.org/papers/volume13/bergstra12a/bergstra12a.pdf>
- Bernardelli De Moraes, M., & Palermo Coelho, G. (2023). Effects of the Random Forests Hyper-Parameters in Surrogate Models for Multi-Objective Combinatorial Optimization: A Case Study using MOEA/D-RFTS. *IEEE Latin America Transactions*, 21(5), 621-627. <https://doi.org/10.1109/TLA.2023.10130833>
- Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann, J. M. (2010). The Binormal Assumption on Precision-Recall Curves. *2010 20th International Conference on Pattern Recognition*, 4263-4266. <https://doi.org/10.1109/ICPR.2010.1036>
- Burkov, A. (2019). *The Hundred-Page Machine Learning Book* (1.<sup>a</sup> ed.).
- Büyükekeci, M., & Okur, M. C. (2023). A Comprehensive Review of Feature Selection and Feature Selection Stability in Machine Learning. *Gazi University Journal of Science*, 36(4), 1506-1520. <https://doi.org/10.35378/gujs.993763>

- Chadaga, K., Prabhu, S., Sampathila, N., Chadaga, R., Bhat, D., Sharma, A. K., & Swathi, K. (2024). SADXAI: Predicting social anxiety disorder using multiple interpretable artificial intelligence techniques. *SLAS Technology*, 29(2), 100129. <https://doi.org/10.1016/j.slast.2024.100129>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. <https://doi.org/10.1145/2939672.2939785>
- Cheng, X. (2024). A Comprehensive Study of Feature Selection Techniques in Machine Learning Models. *Insights in Computer, Signals and Systems*, 1(1), 65-78. <https://doi.org/10.70088/xpf2b276>
- Cho, G., Yim, J., Choi, Y., Ko, J., & Lee, S.-H. (2019). Review of Machine Learning Algorithms for Diagnosing Mental Illness. *Psychiatry Investigation*, 16(4), 262-269. <https://doi.org/10.30773/pi.2018.12.21.2>
- Chua, K. J., Lukaszewski, A. W., Grant, D. M., & Sng, O. (2017). Human life history strategies: Calibrated to external or internal cues? *Evolutionary Psychology*, 15(1), 1474704916677342. <https://doi.org/10.1177/1474704916677342>
- Chung, J., & Teo, J. (2023). Single classifier vs. ensemble machine learning approaches for mental health prediction. *Brain Informatics*, 10(1), 1. <https://doi.org/10.1186/s40708-022-00180-6>
- Clasificación Internacional de Enfermedades, undécima revisión (CIE-11), Organización Mundial de la Salud (OMS). (2024, enero). <https://icd.who.int/browse/2024-01/mms/es#1712535455>
- Davis, J., & Goadrich, M. (2006). The relationship between precision-recall and ROC curves. *Proceedings of the 23rd international conference on Machine learning - ICML '06*, 233-240. <https://doi.org/10.1145/1143844.1143874>
- Daza, A., Arroyo-Paz, Bobadilla, J., Apaza, O., & Pinto, J. (2023). Stacking ensemble learning model for predict anxiety level in university students using balancing methods. *Informatics in Medicine Unlocked*, 42, 101340. <https://doi.org/10.1016/j.imu.2023.101340>
- De la Fuente Fernández, S. (2011). Análisis Factorial. Consultado el 23 de mayo de 2024, desde <https://www.fuenterrebollo.com/Economicas/ECONOMETRIA/MULTIVARIANTE/FACTORIAL/analisis-factorial.pdf>
- Del Giudice, M. (2018, junio). The life history framework and the FSD model. En M. Del Giudice (Ed.), *Evolutionary psychopathology* (pp. 153-192). Oxford University Press. <https://doi.org/10.1093/med-psych/9780190246846.003.0006>
- Figueredo, A. J., Garcia, R. A., Menke, J. M., Jacobs, W. J., Gladden, P. R., Bianchi, J., Patch, E. A., Beck, C. J. A., Kavanagh, P. S., Sotomayor-Peterson, M., Jiang, Y., & Li, N. P. (2017). The k-SF-42: A new short form of the arizona life history battery. *Evolutionary Psychology*, 15(1), 147470491667627. <https://doi.org/10.1177/1474704916676276>
- Figueredo, A. J., & Gaxiola, R. J. (2007). The Arizona-Sonora Life History Battery [Cuestionario disponible en línea]. Consultado el 19 de junio de 2024, desde <https://arizona.app.box.com/s/bvzzxw3o7m43l2hw3bdmoeqf6n203s18>

- Figueredo, A. J., Vásquez, G., Brumbach, B. H., & Schneider, S. M. R. (2004). The heritability of life history strategy: The k-factor, covitality, and personality. *Biodemography and Social Biology*, 51(3), 121-143. <https://doi.org/10.1080/19485565.2004.9989090>
- Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7(3), 286-299. <https://doi.org/10.1037/1040-3590.7.3.286>
- Gaitán-Rossi, P., Pérez-Hernández, V., Vilar-Compte, M., & Teruel-Belismelis, G. (2021). Prevalencia mensual de trastorno de ansiedad generalizada durante la pandemia por Covid-19 en México. *Salud Pública de México*, 63(4), 478-485. <https://doi.org/10.21149/12257>
- Gelman, A. (2005). Analysis of variance—why it is more important than ever. *The Annals of Statistics*, 33(1). <https://doi.org/10.1214/009053604000001048>
- Giudice, M. D., Gangestad, S. W., & Kaplan, H. S. (2015, 18 de noviembre). Life history theory and evolutionary psychology. En D. M. Buss (Ed.), *The handbook of evolutionary psychology* (1.<sup>a</sup> ed., pp. 1-27). Wiley. <https://doi.org/10.1002/9781119125563.evpsych102>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Goswami, T., & Roy, U. B. (2021). Classification accuracy comparison for imbalanced datasets with its balanced counterparts obtained by different sampling techniques [Series Title: Lecture Notes in Electrical Engineering]. En A. Kumar & S. Mozar (Eds.), *ICCCE 2020* (pp. 45-54, Vol. 698). Springer Nature Singapore. [https://doi.org/10.1007/978-981-15-7961-5\\_5](https://doi.org/10.1007/978-981-15-7961-5_5)
- Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). [No title found]. *Machine Learning*, 46(1), 389-422. <https://doi.org/10.1023/A:1012487302797>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer New York. <https://doi.org/10.1007/978-0-387-84858-7>
- Huan Liu & Lei Yu. (2005). Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge and Data Engineering*, 17(4), 491-502. <https://doi.org/10.1109/TKDE.2005.66>
- Instituto Mexicano del Seguro Social. (2023, 4 de julio). *La punta del iceberg* [Opinión IMSS]. Consultado el 22 de abril de 2025, desde <https://www.gob.mx/imss/articulos/la-punta-del-iceberg?idiom=es#:~:text=Pero%20a%20C3%20BA%20no%20es%20suficiente,de%20ellos%20no%20recibe%20tratamiento>.
- Islam, M. T., Ashraf, K., Hosen, M., Nawar, S., & Asgar, S. (2024). Predictive Modeling of Anxiety Levels in Bangladeshi University Students: A Voting-Based Approach with LIME and SHAP Explanations. *2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS)*, 01-06. <https://doi.org/10.1109/iCACCESS61735.2024.10499576>
- Kahl, B. L., Kavanagh, P. S., & Gleaves, D. H. (2022). Testing a life history model of psychopathology: A replication and extension. *Current Psychology*, 41(9), 6233-6246. <https://doi.org/10.1007/s12144-020-01062-y>
- Kavanagh, P. S., & Kahl, B. L. (2016). Life history theory. En V. Weekes-Shackelford, T. K. Shackelford & V. A. Weekes-Shackelford (Eds.), *Encyclopedia of evolutionary psychological science* (pp. 1-12). Springer International Publishing. [https://doi.org/10.1007/978-3-319-16999-6\\_1914-1](https://doi.org/10.1007/978-3-319-16999-6_1914-1)

- Kim, H.-Y. (2017). Statistical notes for clinical researchers: Chi-squared test and fisher's exact test. *Restorative Dentistry & Endodontics*, 42(2), 152. <https://doi.org/10.5395/rde.2017.42.2.152>
- Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97(1), 273-324. [https://doi.org/10.1016/S0004-3702\(97\)00043-X](https://doi.org/10.1016/S0004-3702(97)00043-X)
- Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260), 583-621. <https://doi.org/10.1080/01621459.1952.10483441>
- Ku, W. L., & Min, H. (2024). Evaluating machine learning stability in predicting depression and anxiety amidst subjective response errors. *Healthcare*, 12(6), 625. <https://doi.org/10.3390/healthcare12060625>
- Kumar, Y., Koul, A., Singla, R., & Ijaz, M. F. (2023). Artificial intelligence in disease diagnosis: A systematic literature review, synthesizing framework and future research agenda. *Journal of Ambient Intelligence and Humanized Computing*, 14(7), 8459-8486. <https://doi.org/10.1007/s12652-021-03612-z>
- López Steinmetz, L. C., Sison, M., Zhumagambetov, R., Godoy, J. C., & Haufe, S. (2024). Machine learning models predict the emergence of depression in Argentinean college students during periods of COVID-19 quarantine. *Frontiers in Psychiatry*, 15, 1376784. <https://doi.org/10.3389/fpsy.2024.1376784>
- Lytra, G. (2021). Hyperparameter optimization on supervised learning models [Artwork Size: 56 Pages: 56]. [https://doi.org/10.26267/UNIPI\\_DIONE/880](https://doi.org/10.26267/UNIPI_DIONE/880)
- Mahalingam, M., Jammal, M., Hoteit, R., Ayna, D., Romani, M., Hijazi, S., Bou-Hamad, I., & El Morr, C. (2023, 29 de junio). A Machine Learning Study to Predict Anxiety on Campuses in Lebanon. En J. Mantas, P. Gallos, E. Zoulias, A. Hasman, M. S. Househ, M. Charalampidou & A. Magdaliniou (Eds.), *Studies in Health Technology and Informatics*. IOS Press. <https://doi.org/10.3233/SHTI230430>
- Maitre, J., Bergeron-Leclerc, C., Maltais, D., & Gaboury, S. (2023). Investigating anxiety levels in the quebec university community during the COVID-19 pandemic using machine learning and data exploration techniques. *Multimedia Tools and Applications*, 82(29), 46109-46127. <https://doi.org/10.1007/s11042-023-16096-0>
- Marey, M., Salem, D., Rashidy, N. E., & ELBakry, H. (2024). Enhancing the Diagnosis of Depression and Anxiety Through Explainable Machine Learning Methods [Publisher: The Science and Information Organization]. *International Journal of Advanced Computer Science and Applications*, 15(4). <https://doi.org/10.14569/IJACSA.2024.0150481>
- Matsunaga, M. (2010). How to factor-analyze your data right: Do's, don'ts, and how-to's. *International Journal of Psychological Research*, 3(1), 97-110.
- McHugh, M. L. (2013). The Chi-square test of independence. *Biochemia Medica*, 143-149. <https://doi.org/10.11613/BM.2013.018>
- Meyer, T., Miller, M., Metzger, R., & Borkovec, T. D. (1990). Development and validation of the penn state worry questionnaire. *Behaviour Research and Therapy*, 28(6), 487-495. [https://doi.org/10.1016/0005-7967\(90\)90135-6](https://doi.org/10.1016/0005-7967(90)90135-6)
- Misra, P., & Yadav, A. S. (2020). Improving the classification accuracy using recursive feature elimination with cross-validation. 11(3), 659-665.

- Monica & Agrawal, P. (2024). A Survey on Hyperparameter Optimization of Machine Learning Models. *2024 2nd International Conference on Disruptive Technologies (ICDT)*, 11-15. <https://doi.org/10.1109/ICDT61202.2024.10489732>
- Mutalib, S., & Al, E. (2021). Mental health prediction models using machine learning in higher education institution [Number: 5]. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(5), 1782-1792. <https://doi.org/10.17762/turcomat.v12i5.2181>
- Ordonez, C. (2006). Comparing association rules and decision trees for disease prediction. *Proceedings of the international workshop on Healthcare information and knowledge management*, 17-24. <https://doi.org/10.1145/1183568.1183573>
- Orlenko, A., & Moore, J. H. (2021). A comparison of methods for interpreting random forest models of genetic association in the presence of non-additive interactions. *BioData Mining*, 14(1), 9. <https://doi.org/10.1186/s13040-021-00243-0>
- Priyatno, A. M., & Widiyaningtyas, T. (2024). A SYSTEMATIC LITERATURE REVIEW: RECURSIVE FEATURE ELIMINATION ALGORITHMS. *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, 9(2), 196-207. <https://doi.org/10.33480/jitk.v9i2.5015>
- Probst, P., Wright, M. N., & Boulesteix, A.-L. (2019). Hyperparameters and tuning strategies for random forest. *WIREs Data Mining and Knowledge Discovery*, 9(3), e1301. <https://doi.org/10.1002/widm.1301>
- Rahman, A., Ananna, A. A., Fahad, S., & Mia, M. S. (2023). Prediction of Depression and Anxiety on University Students in Bangladesh Using Machine Learning. *2023 5th International Conference on Sustainable Technologies for Industry 5.0 (STI)*, 1-6. <https://doi.org/10.1109/STI59863.2023.10465126>
- Rashidi, H. H., Albahra, S., Robertson, S., Tran, N. K., & Hu, B. (2023). Common statistical concepts in the supervised Machine Learning arena. *Frontiers in Oncology*, 13, 1130229. <https://doi.org/10.3389/fonc.2023.1130229>
- Rawat, P. (2021). Importance of Feature Selection in Model Accuracy. <https://api.semanticscholar.org/CorpusID:236907306>
- Richter, T., Fishbain, B., Fruchter, E., Richter-Levin, G., & Okon-Singer, H. (2021). Machine learning-based diagnosis support system for differentiating between clinical anxiety and depression disorders. *Journal of Psychiatric Research*, 141, 199-205. <https://doi.org/10.1016/j.jpsychires.2021.06.044>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533-536. <https://doi.org/10.1038/323533a0>
- Sandín, B., Chorot, P., Valiente, R., & Lostao, L. (2009). Validación española del cuestionario de preocupación PSWQ: Estructura factorial y propiedades psicométricas. *Revista de Psicopatología y Psicología Clínica*, 14, 107-122. <https://doi.org/10.5944/rppc.vol.14.num.2.2009.4070>
- Shah, N. P., Jeshwani, S., & Bhatt, P. (2020). An Introduction on Interpretable Machine Learning. *International Journal of Innovative Technology and Exploring Engineering*, 9(7), 107-111. <https://doi.org/10.35940/ijitee.G1023.0597S20>
- Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: from theory to algorithms*. Cambridge university press.

- Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3), 591. <https://doi.org/10.2307/2333709>
- Silveira, É. D. M., Passos, I. C., Scott, J., Bristot, G., Scotton, E., Teixeira Mendes, L. S., Umpierre Knackfuss, A. C., Gerchmann, L., Fijtman, A., Trasel, A. R., Salum, G. A., & Kauer-Sant'Anna, M. (2020). Decoding rumination: A machine learning approach to a transdiagnostic sample of outpatients with anxiety, mood and psychotic disorders. *Journal of Psychiatric Research*, 121, 207-213. <https://doi.org/10.1016/j.jpsychires.2019.12.005>
- Srivastava, S., Joshi, N., & Gaur, M. (2013). A Review Paper on Feature Selection Methodologies and Their Applications. *International Journal of Engineering Research and Development*, 7(6), 57-61. <https://www.ijerd.com/paper/vol7-issue6/H07065761.pdf>
- Stapleton, C. D. (1997). Basic Concepts and Procedures of Confirmatory Factor Analysis. <https://api.semanticscholar.org/CorpusID:58944126>
- Vos, G., van Eijk, L., Sarnyai, Z., & Azghadi, M. R. (2024). Stabilizing Machine Learning for Reproducible and Explainable Results: A Novel Validation Approach to Subject-Specific Insights [Version Number: 1]. <https://doi.org/10.48550/ARXIV.2412.16199>
- Wang, J., Black, M., Rankin, D., Wallace, J., Hughes, C. F., Hoey, L., Moore, A., Tobin, J., Zhang, M., Ng, J., Horigan, G., Carlin, P., McCarroll, K., Cunningham, C., McNulty, H., & Molloy, A. M. (2023). Analysis of Risk Factors and Diagnosis for Anxiety Disorder in Older People with the Aid of Artificial Intelligence: Observational Study. *2023 31st Irish Conference on Artificial Intelligence and Cognitive Science (AICS)*, 1-8. <https://doi.org/10.1109/AICS60730.2023.10470782>
- Watkins, M. (2020, 29 de diciembre). *A step-by-step guide to exploratory factor analysis with r and RStudio* (1.<sup>a</sup> ed.). Routledge. <https://doi.org/10.4324/9781003120001>
- WHO, W. H. O. (2023, septiembre). Trastornos de ansiedad. <https://www.who.int/es/news-room/fact-sheets/detail/anxiety-disorders>
- Wu, J., Chen, X.-Y., Zhang, H., Xiong, L.-D., Lei, H., & Deng, S.-H. (2019). Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization. *Journal of Electronic Science and Technology*, 17(1), 26-40. <https://doi.org/10.11989/JEST.1674-862X.80904120>
- Y, S., & Challa, M. (2023). A Comparative Analysis of Explainable AI Techniques for Enhanced Model Interpretability. *2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN)*, 229-234. <https://doi.org/10.1109/ICPCSN58827.2023.00043>
- Yaminfirooz, M., & Gholinia-Ahangar, H. (2014). Analysis the exploratory factor of evaluating indicators for the researchers' scientific outputs. *Caspian Journal of Scientometrics*, 1(1). <https://doi.org/10.22088/acadpub.BUMS.1.1.7>
- Yates, L. A., Aandahl, Z., Richards, S. A., & Brook, B. W. (2023). Cross validation for model selection: A review with examples from ecology. *Ecological Monographs*, 93(1), e1557. <https://doi.org/10.1002/ecm.1557>

# Apéndice A

## Consentimiento informado



INSTITUTO MEXICANO DEL SEGURO SOCIAL  
UNIDAD DE EDUCACIÓN, INVESTIGACIÓN  
Y POLÍTICAS DE SALUD  
COORDINACIÓN DE INVESTIGACIÓN EN SALUD

### Carta de consentimiento informado para participación en protocolos de investigación (adultos)

|                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|---------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Nombre del estudio:                                           | PATRONES DE LAS ESTRATEGIAS DE HISTORIA DE VIDA Y SU ASOCIACIÓN CON TRASTORNO DE ANSIEDAD GENERALIZADA EN DERECHOHABIENTES DE LA UMF NO. 20, OOAD MORELOS.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| Patrocinador externo (si aplica):                             | NA.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| Lugar y fecha:                                                | Cuernavaca, Morelos a _____ de _____ de 20____.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| Número de registro institucional:                             | En espera.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| Justificación y objetivo del estudio:                         | <p>Se estima que en México entre el 31% y el 33% de la población mexicana padece de trastorno de ansiedad generalizada. Aunque se ha avanzado en el uso de la inteligencia artificial para detectar este trastorno, no se ha investigado si las estrategias de historia de vida influyen en el desarrollo de ansiedad. Este estudio contribuirá a comprender el trastorno de ansiedad generalizada usando las estrategias de historia de vida y llevar a cabo su detección utilizando herramientas de computación.</p> <p>Objetivo general: Analizar los patrones de las estrategias de historia de vida y su asociación con trastorno de ansiedad generalizada en derechohabientes de la UMF no. 20, OOAD Morelos</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| Procedimientos:                                               | <p>A usted se le invita a participar en este estudio de investigación. Si usted de forma voluntaria acepta participar, se le solicitará que nos responda un cuestionario, el cual está dividido en cuatro secciones:</p> <ul style="list-style-type: none"> <li>a) Cuestionario sociodemográfico-académico, el cual nos permitirá obtener sus datos generales que nos serán útiles para conocerlo de mejor manera.</li> <li>b) Cuestionario de estrategias de historia de vida, que nos servirá para conocer su forma de vida y comportamiento en general.</li> <li>c) Cuestionario llamado K-SF-42, el cual nos permitirá complementar el cuestionario previo con información útil para una valoración integral.</li> <li>d) Cuestionario de preocupación de Pensilvania, este cuestionario nos será de utilidad para clasificar su nivel de ansiedad.</li> <li>e) Cuestionario llamado FACES III, el cual nos permite evaluar la funcionalidad de su familia.</li> </ul> <p>Esta encuesta será autoaplicada, es decir usted lo responderá a su ritmo y usted podrá contestar, o no, cada una de ellas, según lo que usted considere. El personal encargado de recolectar los datos estará al pendiente por si usted tiene alguna duda o requiera de apoyo adicional, sin intervenir en las respuestas que usted pudiera emitir.</p> <p>El tiempo aproximado que le tomará toda la actividad será de aproximadamente 30 minutos. Toda esta actividad la realizaremos en un aula del área de enseñanza de la clínica UMF 20.</p> |
| Posibles riesgos y molestias:                                 | <p>Este estudio es considerado como una investigación sin riesgo, es decir, que no le provocará mayor molestia, más que el tiempo de 30 minutos que le tomará contestar la encuesta. Sin embargo, si algunas preguntas en el cuestionario lo(a) hacen sentir incómodo(a), usted tiene el derecho de hacerlo saber al reclutador (a) y podría no responder a cualquier pregunta que le incomode o suspender la encuesta en cualquier momento, si así lo decide. Asimismo, puede manifestar su retiro del estudio sin explicación alguna y sin que esto represente alguna repercusión para usted en la atención a su salud en el Instituto Mexicano del Seguro Social.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| Posibles beneficios que recibirá al participar en el estudio: | <p>Usted al participar en este estudio no recibirá un pago por su participación, pero tampoco implicará gasto alguno. En caso de detectar presencia de trastorno de ansiedad generalizada, se gestionará una cita con su médico familiar para que realice una evaluación integral y en su caso determinará si requiere, o no, envío a psicología u otra especialidad.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |

Figura A.1: Consentimiento informado (parte 1)

|                                                            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Información sobre resultados y alternativas de tratamiento | Después de analizar los datos, le estaremos informando los resultados obtenidos de su nivel de ansiedad, su dinámica familiar y su estrategia de historia de vida, a través de una llamada telefónica o en su próxima consulta con su médico familiar. En caso de que usted no desee conocer sus resultados, podrá expresarlo al reclutador y se tomará en consideración para no realizarlo.                                                                   |
| Participación o retiro:                                    | Su participación será de forma voluntaria y, en caso de no querer participar o abandonar el estudio, lo podrá hacer en cualquier momento sin tener que dar explicación alguna. Tenga la total confianza de que no tendrá ninguna repercusión en sus derechos ante el Instituto Mexicano del Seguro Social.                                                                                                                                                     |
| Privacidad y confidencialidad:                             | Se mantendrá la confidencialidad de su información utilizando códigos y números de acuerdo con los lineamientos éticos vigentes para el manejo de datos. El grupo de investigación manifiesta que mantendrá en todo momento el respeto a su persona y no se divulgará su nombre en la difusión de los resultados obtenidos. Asimismo, le manifestamos que garantiremos su privacidad al realizar la encuesta en un aula del área de enseñanza de esta clínica. |

**Declaración de consentimiento:**

Después de haber leído y habiéndome explicado todas mis dudas acerca de este estudio:

☐ Si acepto participar en el estudio.

☐ No acepto participar en el estudio.

**En caso de dudas o aclaraciones relacionadas con el estudio podrá dirigirse a:**

**Investigadora o Investigador Responsable:** Dr. Ricardo Castrejón Salgado, Médico familiar con adscripción a la UMF No. 3, IMSS Morelos. Matrícula 99182952.

**Colaboradores:**

Dra. Laura Ayala Jiménez, Coordinadora Auxiliar Médica de Investigación en Salud, con adscripción a la Jefatura de Servicios de Prestaciones Médicas, IMSS Morelos. Matrícula 10202331.

Dra. Alicia Martínez Rebollar, Profesora-Investigadora del Departamento de Ciencias Computacionales con adscripción al TECN/CENIDET. Contacto: [alicia.mr@cenedet.tecnm.mx](mailto:alicia.mr@cenedet.tecnm.mx)

Ing. Ana Luisa Islas Ávila, Estudiante de primer año de la maestría en Ciencias Computacionales, con adscripción al TECN/CENIDET. Contacto: [m23ce0950@cenedet.tecnm.mx](mailto:m23ce0950@cenedet.tecnm.mx)

**En caso de dudas o aclaraciones sobre sus derechos como participante podrá dirigirse a:** Comité Local de Ética de Investigación en Salud del CNIC del IMSS: Avenida Cuauhtémoc 330 4º piso Bloque "B" de la Unidad de Congresos, Colonia Doctores, México, D.F., CP 06720. Teléfono (55) 56 27 69 00 extensión 21230, correo electrónico: [comite.eticainv@imss.gob.mx](mailto:comite.eticainv@imss.gob.mx)

---

|                                     |                                                   |
|-------------------------------------|---------------------------------------------------|
| Nombre y firma del participante     | Nombre y firma de quien obtiene el consentimiento |
| Testigo 1                           | Testigo 2                                         |
| Nombre, dirección, relación y firma | Nombre, dirección, relación y firma               |

Este formato constituye una guía que deberá completarse de acuerdo con las características propias de cada protocolo de investigación, sin omitir información relevante del estudio.

Clave: 2810-009-013

Figura A.2: Consentimiento informado (parte 2)

# Apéndice B

## Instrumentos utilizados

|                                                                                                                                                           |                                           |                             |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------|-----------------------------|
|                                                                          |                                           |                             |
| INSTITUTO MEXICANO DEL SEGURO SOCIAL                                                                                                                      |                                           |                             |
| SEGURIDAD Y SOLIDARIDAD SOCIAL                                                                                                                            |                                           |                             |
| Código: _____                                                                                                                                             | Fecha: _____                              |                             |
| <b>Datos sociodemográficos</b>                                                                                                                            |                                           |                             |
| Edad: _____                                                                                                                                               | Municipio de residencia: _____            |                             |
| <b>Sexo biológico:</b>                                                                                                                                    |                                           |                             |
| Hombre ( )                                                                                                                                                | Mujer( )                                  |                             |
| <b>Último grado de estudios concluido:</b>                                                                                                                |                                           |                             |
| Primaria ( )                                                                                                                                              | Secundaria ( )                            | Preparatoria ( )            |
| Licenciatura ( )                                                                                                                                          | Posgrado ( )                              |                             |
| <b>Estado civil:</b>                                                                                                                                      |                                           |                             |
| Soltero/a ( )                                                                                                                                             | Casado/a ( )                              | Unión libre/Concubinato ( ) |
| Divorciado/a ( )                                                                                                                                          | Viudo/a ( )                               | Separado/a ( )              |
| <b>Ocupación:</b>                                                                                                                                         |                                           |                             |
| Estudiante ( )                                                                                                                                            | Empleado/a sin estudios profesionales ( ) |                             |
| Empleado/a con estudios profesionales ( )                                                                                                                 | Negocio propio ( )                        |                             |
| Trabajador independiente ( )                                                                                                                              | Hogar ( )                                 |                             |
| Pensionado/o, Jubilado/a ( )                                                                                                                              | Desempleado/a ( )                         |                             |
| <b>Nivel socioeconómico:</b>                                                                                                                              |                                           |                             |
| Alto - acceso a educación privada, servicios de salud privados, bienes inmuebles de alto valor, sin limitaciones para gastos no esenciales. ( )           |                                           |                             |
| Medio-alto - acceso a educación privada, servicios de salud privados, casa o departamento propio, flexibilidad para asumir gastos no esenciales. ( )      |                                           |                             |
| Medio - se cubren necesidades básicas como alimentación, vivienda, y educación, pero con limitaciones para gastos no esenciales. ( )                      |                                           |                             |
| Medio-bajo - se cubren necesidades básicas con cierta estabilidad, pero se enfrentan dificultades económicas ante gastos imprevistos o no esenciales. ( ) |                                           |                             |
| Bajo - Ingresos insuficientes para satisfacer adecuadamente necesidades básicas y dependen de programas de asistencia social. ( )                         |                                           |                             |
| Pobreza - Ingresos muy bajos que no pueden cubrir las necesidades básicas de subsistencia, viviendo en condiciones de vulnerabilidad extrema. ( )         |                                           |                             |

Figura B.1: Vista parcial de cuestionario sociodemográfico

## Cuestionario Estrategias de historia de vida

A continuación se presentan una serie de enunciados asociados a las estrategias de historia de vida del modelo FSD. Favor de marcar la opción más cercana a tu sentir.

1. **Me considero una persona que prefiere decir las cosas de frente a los demás y comúnmente me dicen no tener tacto o amabilidad al comunicarme.**  

|                            |                       |                           |                       |                           |
|----------------------------|-----------------------|---------------------------|-----------------------|---------------------------|
| <input type="radio"/>      | <input type="radio"/> | <input type="radio"/>     | <input type="radio"/> | <input type="radio"/>     |
| No me describe en absoluto | Me describe muy poco  | Me describe moderadamente | Me describe bastante  | Me describe absolutamente |
  
2. **Habitualmente no me cuestiono si lo que hago, pienso y siento es bueno o malo, correcto o incorrecto.**  

|                            |                       |                           |                       |                           |
|----------------------------|-----------------------|---------------------------|-----------------------|---------------------------|
| <input type="radio"/>      | <input type="radio"/> | <input type="radio"/>     | <input type="radio"/> | <input type="radio"/>     |
| No me describe en absoluto | Me describe muy poco  | Me describe moderadamente | Me describe bastante  | Me describe absolutamente |
  
3. **Soy una persona activa y orgullosa de sí misma, sé que soy valiosa y me considero alguien importante.**  

|                            |                       |                           |                       |                           |
|----------------------------|-----------------------|---------------------------|-----------------------|---------------------------|
| <input type="radio"/>      | <input type="radio"/> | <input type="radio"/>     | <input type="radio"/> | <input type="radio"/>     |
| No me describe en absoluto | Me describe muy poco  | Me describe moderadamente | Me describe bastante  | Me describe absolutamente |
  
4. **Me gusta sentirme cómodo/a en los distintos ámbitos de la vida aunque eso me lleve a la deshonestidad en algunas ocasiones.**  

|                            |                       |                           |                       |                           |
|----------------------------|-----------------------|---------------------------|-----------------------|---------------------------|
| <input type="radio"/>      | <input type="radio"/> | <input type="radio"/>     | <input type="radio"/> | <input type="radio"/>     |
| No me describe en absoluto | Me describe muy poco  | Me describe moderadamente | Me describe bastante  | Me describe absolutamente |
  
5. **Suelo sentirme atraído/a por lo bello e imaginar distintas situaciones que me agradan.**  

|                            |                       |                           |                       |                           |
|----------------------------|-----------------------|---------------------------|-----------------------|---------------------------|
| <input type="radio"/>      | <input type="radio"/> | <input type="radio"/>     | <input type="radio"/> | <input type="radio"/>     |
| No me describe en absoluto | Me describe muy poco  | Me describe moderadamente | Me describe bastante  | Me describe absolutamente |
  
6. **Mi interés para buscar experiencias sexuales fue a una edad muy temprana.**  

|                            |                       |                           |                       |                           |
|----------------------------|-----------------------|---------------------------|-----------------------|---------------------------|
| <input type="radio"/>      | <input type="radio"/> | <input type="radio"/>     | <input type="radio"/> | <input type="radio"/>     |
| No me describe en absoluto | Me describe muy poco  | Me describe moderadamente | Me describe bastante  | Me describe absolutamente |

Figura B.2: Vista parcial del cuestionario Marcadores de estrategias de historia de vida

14. Tengo gran habilidad verbal y soy convincente cuando dialogo con los demás.

|                               |                          |                              |                          |                              |
|-------------------------------|--------------------------|------------------------------|--------------------------|------------------------------|
| <input type="checkbox"/>      | <input type="checkbox"/> | <input type="checkbox"/>     | <input type="checkbox"/> | <input type="checkbox"/>     |
| No me describe<br>en absoluto | Me describe<br>muy poco  | Me describe<br>moderadamente | Me describe<br>bastante  | Me describe<br>absolutamente |

15. No reconozco rostros u objetos con facilidad y me cuesta trabajo orientarme en un espacio.

|                               |                          |                              |                          |                              |
|-------------------------------|--------------------------|------------------------------|--------------------------|------------------------------|
| <input type="checkbox"/>      | <input type="checkbox"/> | <input type="checkbox"/>     | <input type="checkbox"/> | <input type="checkbox"/>     |
| No me describe<br>en absoluto | Me describe<br>muy poco  | Me describe<br>moderadamente | Me describe<br>bastante  | Me describe<br>absolutamente |

16. Considero que mi maduración sexual (cambios físicos, hormonales y emocionales) ha sido temprana y rápida.

|                               |                          |                              |                          |                              |
|-------------------------------|--------------------------|------------------------------|--------------------------|------------------------------|
| <input type="checkbox"/>      | <input type="checkbox"/> | <input type="checkbox"/>     | <input type="checkbox"/> | <input type="checkbox"/>     |
| No me describe<br>en absoluto | Me describe<br>muy poco  | Me describe<br>moderadamente | Me describe<br>bastante  | Me describe<br>absolutamente |

## Instrumento K-SF-42

Por favor, indica cuánto estás de acuerdo o en desacuerdo con las siguientes afirmaciones. Utiliza la escala a continuación para responder a cada afirmación, marca con una X el recuadro de la opción que consideras que se aplica a ti. Para cualquier ítem que no se aplique a ti, selecciona 'No sé/No aplica'.

### Perspectiva, Planificación y Control

1. Cuando me encuentro con una mala situación, hago lo que puedo para sacar lo mejor de ella.

|                                 |                          |                                 |                          |                           |                          |                           |
|---------------------------------|--------------------------|---------------------------------|--------------------------|---------------------------|--------------------------|---------------------------|
| <input type="checkbox"/>        | <input type="checkbox"/> | <input type="checkbox"/>        | <input type="checkbox"/> | <input type="checkbox"/>  | <input type="checkbox"/> | <input type="checkbox"/>  |
| Fuertemente<br>en<br>desacuerdo | En<br>desacuerdo         | Ligeramente<br>en<br>desacuerdo | No sé / No<br>aplica     | Ligeramente<br>de acuerdo | De acuerdo               | Fuertemente<br>de acuerdo |

2. Cuando me encuentro en problemas, no me rindo hasta que los resuelvo.

|                                 |                          |                                 |                          |                           |                          |                           |
|---------------------------------|--------------------------|---------------------------------|--------------------------|---------------------------|--------------------------|---------------------------|
| <input type="checkbox"/>        | <input type="checkbox"/> | <input type="checkbox"/>        | <input type="checkbox"/> | <input type="checkbox"/>  | <input type="checkbox"/> | <input type="checkbox"/>  |
| Fuertemente<br>en<br>desacuerdo | En<br>desacuerdo         | Ligeramente<br>en<br>desacuerdo | No sé / No<br>aplica     | Ligeramente<br>de acuerdo | De acuerdo               | Fuertemente<br>de acuerdo |

Figura B.3: Vista parcial de cuestionario K-SF-42 (parte 1)

*Nota.* Versión en español desarrollada por Figueredo y Gaxiola, 2007

3. Me doy cuenta que generalmente aprendo algo valioso de una situación difícil.

|                       |                       |                       |                       |                       |                       |                       |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Fuertemente           | En                    | Ligeramente           | No sé / No            | Ligeramente           | De acuerdo            | Fuertemente           |
| en                    | desacuerdo            | en                    | aplica                | de acuerdo            |                       | de acuerdo            |
| desacuerdo            |                       | desacuerdo            |                       |                       |                       |                       |

4. Cuando me enfrento con una mala situación, esto me ayuda a encontrar una manera diferente de ver las cosas.

|                       |                       |                       |                       |                       |                       |                       |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Fuertemente           | En                    | Ligeramente           | No sé / No            | Ligeramente           | De acuerdo            | Fuertemente           |
| en                    | desacuerdo            | en                    | aplica                | de acuerdo            |                       | de acuerdo            |
| desacuerdo            |                       | desacuerdo            |                       |                       |                       |                       |

5. Aunque todas las cosas parecen ir mal, puedo usualmente encontrar un lado bueno.

|                       |                       |                       |                       |                       |                       |                       |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Fuertemente           | En                    | Ligeramente           | No sé / No            | Ligeramente           | De acuerdo            | Fuertemente           |
| en                    | desacuerdo            | en                    | aplica                | de acuerdo            |                       | de acuerdo            |
| desacuerdo            |                       | desacuerdo            |                       |                       |                       |                       |

6. Puedo encontrar algo positivo hasta en las peores situaciones.

|                       |                       |                       |                       |                       |                       |                       |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Fuertemente           | En                    | Ligeramente           | No sé / No            | Ligeramente           | De acuerdo            | Fuertemente           |
| en                    | desacuerdo            | en                    | aplica                | de acuerdo            |                       | de acuerdo            |
| desacuerdo            |                       | desacuerdo            |                       |                       |                       |                       |

#### Altruismo General

7. Paso una gran cantidad de tiempo por mes dando apoyo emocional informal a mis parientes consanguíneos (no políticos).

|                       |                       |                       |                       |                       |                       |                       |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Fuertemente           | En                    | Ligeramente           | No sé / No            | Ligeramente           | De acuerdo            | Fuertemente           |
| en                    | desacuerdo            | en                    | aplica                | de acuerdo            |                       | de acuerdo            |
| desacuerdo            |                       | desacuerdo            |                       |                       |                       |                       |

8. Contribuyo con una gran cantidad de recursos para el bienestar económico y general de mis parientes consanguíneos (no políticos).

|                       |                       |                       |                       |                       |                       |                       |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Fuertemente           | En                    | Ligeramente           | No sé / No            | Ligeramente           | De acuerdo            | Fuertemente           |
| en                    | desacuerdo            | en                    | aplica                | de acuerdo            |                       | de acuerdo            |
| desacuerdo            |                       | desacuerdo            |                       |                       |                       |                       |

9. Paso una gran cantidad de tiempo por mes dando apoyo emocional informal a mis conocidos (tales como vecinos o gente de la iglesia).

|                       |                       |                       |                       |                       |                       |                       |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Fuertemente           | En                    | Ligeramente           | No sé / No            | Ligeramente           | De acuerdo            | Fuertemente           |
| en                    | desacuerdo            | en                    | aplica                | de acuerdo            |                       | de acuerdo            |
| desacuerdo            |                       | desacuerdo            |                       |                       |                       |                       |

Figura B.4: Vista parcial de cuestionario K-SF-42 (parte 2)

*Nota.* Versión en español desarrollada por Figueredo y Gaxiola, 2007

39. Durante el último mes, ¿qué tan frecuente te han dicho tus amigos que les gustaba tu forma de ser?

☐  
Nunca

☐  
Muy poco

☐  
Poco

☐  
Mucho

40. Durante el último mes, ¿qué tan frecuente tus amigos te han demostrado afecto?

☐  
Nunca

☐  
Muy poco

☐  
Poco

☐  
Mucho

41. Durante el último mes, ¿qué tan frecuente tus amigos te han escuchado cuando hablabas de tus sentimientos?

☐  
Nunca

☐  
Muy poco

☐  
Poco

☐  
Mucho

42. Durante el último mes, ¿qué tan frecuente tus amigos han demostrado interés y preocupación por tu bienestar?

☐  
Nunca

☐  
Muy poco

☐  
Poco

☐  
Mucho

## Instrumento PSWQ

A continuación encontrará una lista de 16 enunciados relacionados con la manera de sentir que tiene la gente. Lee detenidamente cada uno de ellos y, aplicándolo a ti, seleccionando la opción que mejor se ajuste a tu forma de sentir habitual. No pienses demasiado en cada frase. Marca tu opción con una X que a primera vista creas que es lo que mejor te define.

1. Me preocupo si no tengo tiempo suficiente para hacerlo todo.

☐  
No me describe  
en absoluto

☐  
Me describe  
muy poco

☐  
Me describe  
moderadamente

☐  
Me describe  
bastante

☐  
Me describe  
absolutamente

2. Me agobian mis preocupaciones.

☐  
No me describe  
en absoluto

☐  
Me describe  
muy poco

☐  
Me describe  
moderadamente

☐  
Me describe  
bastante

☐  
Me describe  
absolutamente

3. Suelo preocuparme por las cosas.

☐  
No me describe  
en absoluto

☐  
Me describe  
muy poco

☐  
Me describe  
moderadamente

☐  
Me describe  
bastante

☐  
Me describe  
absolutamente

Figura B.5: Vista parcial de cuestionario PSWQ

*Nota.* Versión PSWQ-16 en español traducida y validada por Sandín et al., 2009