

Comprendiendo la depresión a través de *Explainable Artificial Intelligence*

Tamara Perez Perez

Grado de Ingeniería Informática
Inteligencia Artificial

Consultor: Ferran Diego Andilla

Profesor responsable: Susana Acedo Nadal

31/01/2024



Esta obra está sujeta a una licencia de
Reconocimiento-NoComercial-
SinObraDerivada [3.0 España de Creative
Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

FICHA DEL TRABAJO FINAL

Título del trabajo:	Comprendiendo la depresión a través de <i>Explainable Artificial Intelligence</i>
Nombre del autor:	Tamara Perez Perez
Nombre del consultor/a:	Ferran Diego Andilla
Nombre del PRA:	Susana Acedo Nadal
Fecha de entrega (mm/aaaa):	01/2024
Titulación:	Grado en Ingeniería Informática
Área del Trabajo Final:	Inteligencia Artificial
Idioma del trabajo:	Castellano
Número de créditos:	12
Palabras clave	depresión, NHANES, XAI

Resumen del Trabajo (máximo 250 palabras): *Con la finalidad, contexto de aplicación, metodología, resultados i conclusiones del trabajo.*

Este proyecto se enfoca en comprender los factores vinculados a la depresión mediante la aplicación de técnicas de inteligencia artificial y *Explainable AI* (XAI). Utilizando el conjunto de datos del *National Health and Nutrition Examination Survey* (NHANES), se aplicaron modelos como *Decision Tree* y *Explainable Boosting* para explorar las relaciones entre variables. La implementación de técnicas XAI, como SHAP y LIME, sobre un modelo *Random Forest*, permitió una comprensión detallada y extensible de los resultados.

Entre los hallazgos destacados, los trastornos del sueño se identificaron como un indicador significativo de la depresión. Además, se observaron patrones demográficos y biomarcadores en sangre relacionados con la prevalencia de la depresión. Factores de estilo de vida, como la actividad física y la nutrición, también mostraron impacto en la salud mental.

Sin embargo, la desigualdad en el conjunto de datos y la posible reticencia de los individuos a compartir información sobre su salud mental presentaron desafíos significativos. Se sugiere una colaboración más estrecha con profesionales de la salud mental para mejorar la validez de las mediciones.

Como línea de trabajo futura, se propone un estudio exclusivo sobre la relación entre nutrición y depresión, aprovechando la riqueza de datos detallados disponibles en NHANES. Este proyecto destaca la importancia de abordar la complejidad de la depresión desde diversas perspectivas, integrando avances en inteligencia artificial con un enfoque comprensible y explicativo.

Abstract (in English, 250 words or less):

This project focuses on understanding factors linked to depression through the application of artificial intelligence (AI) and Explainable AI (XAI) techniques. Using the dataset from the National Health and Nutrition Examination Survey (NHANES), models such as Decision Tree and Explainable Boosting were employed to explore relationships between variables. The implementation of XAI techniques, such as SHAP and LIME, on a Random Forest model, allowed for a detailed and extensible understanding of the results.

Among the notable findings, sleep disorders were identified as a significant indicator of depression. Additionally, demographic patterns and blood biomarkers related to the prevalence of depression were observed. Lifestyle factors, including physical activity and nutrition, also demonstrated an impact on mental health.

However, the inequality in the dataset and the potential neglect by individuals to share information about their mental health arose significant challenges. A closer collaboration with mental health professionals is suggested to improve measurement validity.

As a future line of work, an exclusive study on the relationship between nutrition and depression is proposed, taking advantage of the detailed data available in NHANES. This project highlights the importance of addressing the complexity of depression from various perspectives, integrating advancements in artificial intelligence with an understandable and explanatory approach.

Índice

1	Introducción	1
1.1	Contexto y justificación del Trabajo	1
1.2	Objetivos del Trabajo	2
1.2.1	Objetivo general	2
1.2.2	Objetivos específicos	2
1.3	Enfoque y método seguido	3
1.4	Planificación del Trabajo	4
1.5	Breve resumen de productos obtenidos	5
1.6	Breve descripción de los otros capítulos de la memoria	6
2	Estado del arte	7
2.1	Introducción a la Depresión	7
2.2	NHANES como fuente de datos	8
2.3	<i>Explainable Artificial Intelligence (XAI)</i>	10
3	Metodología	13
3.1	Selección y tratamiento de datos	13
3.2	Exploración de datos	18
3.3	Modelos ML y técnicas XAI	23
3.3.1	Decision Tree Classifier	24
3.3.2	<i>Explainable Boosting Classifier (EBM)</i>	29
3.3.3	<i>Random Forest</i>	34
3.3.4	LIME	35
3.3.5	SHAP	37
4	Discusión	46
5	Conclusiones	48
6	Glosario	50
7	Bibliografía	51
8	Anexos	53

Lista de Figuras

Figura 1 - Planificación temporal.....	4
Figura 2 - Diagrama de Gantt.....	4
Figura 3 - Distribución variables.....	18
Figura 4 - Distribución variables numéricas	19
Figura 5 - Relación entre variables numéricas continuas y depresión (Scatterplot).....	20
Figura 6 - Relación entre variables categóricas y Depresión (countplot)	21
Figura 7 – Informe de clasificación y matriz de confusión DT (antes de aplicar técnica Oversample) ...	24
Figura 8 - Informe de clasificación y matriz de confusión DT (antes de aplicar técnica Oversample)	25
Figura 9 - Árbol de clasificación Instancia 2 (explain_local)	26
Figura 10- Árbol de clasificación Instancia 108 (explain_local)	27
Figura 11- Árbol de clasificación global (comportamiento variable age)	28
Figura 12 - Informe de clasificación y matriz de confusión EBM	29
Figura 13 - EBM. Explicación local Instancias 2 y 108	30
Figura 14 - EBM. Explicación global (explain_global)	31
Figura 15 - EBM. Comportamiento variables trouble_sleeping_history y vitamin_score.....	32
Figura 16 - Informe de clasificación y matriz de confusión (antes y despues de aplicar GridSearchCV).....	34
Figura 17- LIME. Explicaciones para instancias 2 y 108	36
Figura 18 - SHAP Force Plot instancia 2.....	38
Figura 19 - Gráfico sumario SHAP para instancia 2.....	39
Figura 20- - SHAP Force Plot instancias de 0 a 19 (explicación global)	40
Figura 21 - SHAP Force Plot. Comportamiento variable total_nutrient_score.....	40
Figura 22 - Contribución variable total_nutrient_score variable objetivo	41
Figura 23 – SHAP. Gráfica sumario conjunto total.....	42
Figura 24 - SHAP. Gráfica sumario subconjuntos	43
Figura 25 - Análisis de Factores Relevantes en Subgrupos Específicos	44

1 Introducció

1.1 Contexto y justificación del Trabajo

La depresión es un trastorno de salud mental que afecta a millones de personas en todo el mundo. Según la Organización Mundial de la Salud (OMS) [\[1\]](#), se estima que aproximadamente el 5% de la población adulta a nivel mundial sufre de depresión. El impacto de este trastorno se extiende mucho más allá de la esfera individual, afectando a familias, amigos y la sociedad en su conjunto. Lamentablemente, existe una falta de profesionales especializados en salud mental capaces de satisfacer estas necesidades, lo que crea un desafío importante para la salud pública en todo el mundo.

El interés y la motivación detrás de este trabajo de investigación nacen de una experiencia personal. Hace algunos años fui diagnosticada de esclerosis múltiple (EM), un trastorno autoinmune que afecta al sistema nervioso central. Esta experiencia me impulsó a profundizar sobre la relación que hay entre enfermedades y otros factores, como la alimentación, la microbiota, el estilo de vida, el bienestar emocional, entre otros. Además, la depresión es un trastorno que afecta a varias personas de mi entorno cercano, y se ha observado que es hasta tres veces más común en personas con EM [\[2\]](#).

Para abordar esta complejidad, este trabajo se apoyará en las herramientas de *Intelligence Artificial Explainable* (XAI). La XAI es una rama de la inteligencia artificial que se centra en hacer que los sistemas de IA sean transparentes y comprensibles para las personas. Esto resulta esencial en el contexto de la salud mental, ya que permite que las decisiones y conclusiones que los modelos de aprendizaje automático generan sean comprendidas.

Para llevar a cabo esta investigación, se utilizará el conjunto de datos NHANES (*National Health and Nutrition Examination Survey*) [\[3\]](#) que se encuentra en la plataforma *Kaggle* [\[4\]](#). NHANES es un repositorio de datos relacionado con salud y nutrición de la población de Estados Unidos, que abarca desde 2013 a 2014. Este repositorio incluye datos demográficos, exámenes médicos, información sobre nutrición, datos de laboratorio y cuestionarios sobre salud mental, diabetes, consumo de alcohol, entre otros.

En resumen, este trabajo aspira a aportar conocimiento e innovación en la comprensión de la depresión y su impacto en la sociedad. La combinación de datos obtenidos del conjunto NHANES junto con las herramientas de XAI, permitirá una visión completa y transparente de los factores que influyen en la depresión. Esta comprensión puede contribuir en mejorar la exactitud del diagnóstico, la comprensión de la depresión y reducir la creciente carga que esta problemática supone para la salud pública.

1.2 Objetivos del Trabajo

1.2.1 Objetivo general

Obtener un conocimiento profundo de los factores que pueden influir en el desarrollo del trastorno de la depresión y fomentar una mayor conciencia sobre esta afección y su impacto en la sociedad.

1.2.2 Objetivos específicos

1. Revisar literatura médica relacionada con la depresión y salud mental.
2. Actualizar y ampliar conocimientos en minería de datos, aprendizaje automático y otros temas relacionados.
3. Adquirir conocimientos en *Explainable Artificial Intelligence* (XAI)
4. Identificar y recopilar información relevante a partir del conjunto de datos NHANES, asegurando que se seleccionan las variables adecuadas y se realice una limpieza apropiada.
5. Realizar un análisis exploratorio preliminar
6. Aplicar modelos *Machine Learning* (ML) y técnicas de XAI: SHAP y LIME.
7. Evaluar e interpretar resultados

1.3 Enfoque y método seguido

Este Trabajo de Fin de Grado (TFG) tiene como objetivo principal analizar la relación entre los datos de salud de la Encuesta Nacional sobre Salud y Nutrición (NHANES) y la depresión. Para lograr una comprensión más profunda y transparente de los resultados, se emplearán técnicas de *Explainable Artificial Intelligence* (XAI), seleccionadas específicamente por su capacidad para mejorar la transparencia en la interpretación de los resultados.

Todo el análisis se realizará utilizando el lenguaje de programación Python en un entorno *Jupyter Notebook* proporcionado por Anaconda.

1. Recopilación y preprocesamiento de Datos

La fuente principal de datos es el conjunto de datos NHANES (2013-2014), disponible en *Kaggle*. Este conjunto abarca aspectos demográficos, exámenes médicos, nutrición, datos de laboratorio y un cuestionario sobre hábitos de vida.

En esta etapa se identifican y seleccionan las variables relevantes para el análisis de la depresión, y se introducen nuevas mediante cálculos específicos. La limpieza y transformación se lleva a cabo para abordar valores atípicos, nulos y garantizar la coherencia.

2. Exploración y visualización de Datos

Se llevará a cabo una exploración exhaustiva de los datos recopilados. Se aplicarán técnicas de visualización, como gráficos y diagramas, para obtener una visión inicial de las relaciones y patrones presentes en el conjunto de datos.

3. Aplicación de modelos *Machine Learning* y Técnicas de *Explainable Artificial Intelligence* (XAI)

Se procederá a la selección, entrenamiento y validación de modelos ML. Posteriormente se aplicarán técnicas XAI para mejorar la comprensión e interpretación de los resultados obtenidos, contribuyendo así a una evaluación más clara de la relación entre los datos de salud y nutrición de NHANES y la presencia de la depresión. Entre las técnicas seleccionadas se incluyen SHAP y LIME, que serán detalladas en la sección 2.3 - *Explainable Artificial Intelligence* (XAI).

4. Evaluación e interpretación de Datos

A medida que se avanza en el proceso de modelado y aplicación de técnicas XAI, se llevará a cabo una evaluación continua e interpretación de los resultados. Este enfoque permitirá analizar y entender progresivamente la relación entre los datos de salud y nutrición de NHANES y la presencia de la depresión, de manera paralela al desarrollo de los modelos y la aplicación de técnicas XAI.

1.4 Planificación del Trabajo

La planificación del trabajo se estructura considerando los plazos asociados a las entregas de las diversas PEC a lo largo de la ejecución del proyecto. Se presenta a continuación una representación visual de esta programación mediante un diagrama de Gantt.

Comprendiendo la depresión a través de XAI

Tarea	Nombre	Fecha de inicio	Fecha de fin
TFG. Comprendiendo la depresión a través de Explainable AI	Fase 0. Definición de los contenidos del trabajo	27/9/23	31/1/24
	Definición del tema del TFG	27/9/23	9/10/23
	Búsqueda y estudio de documentación relacionada	27/9/23	28/9/23
	Elaboración documento PEC0	28/9/23	4/10/23
	Entrega PEC0	6/10/23	7/10/23
	Entrega PEC0	9/10/23	9/10/23
Fase 1. Plan de trabajo	Plan de trabajo	10/10/23	24/10/23
	Lectura de la estructura de la memoria del TFG	10/10/23	12/10/23
	Planificación temporal de cada fase del proyecto	10/10/23	17/10/23
	Definición de objetivos generales y específicos	13/10/23	20/10/23
	Entrega PEC1	24/10/23	24/10/23
Fase 2. Desarrollo del trabajo	Estado del arte	25/10/23	29/11/23
	Selección y tratamiento de datos	25/10/23	8/11/23
	Entrega PEC2	31/10/23	29/11/23
	Entrega PEC2	29/11/23	29/11/23
	Entrega PEC2	29/11/23	29/11/23
Fase 3. Desarrollo del trabajo	Exploración de datos	30/11/23	25/12/23
	Entrenamiento de modelos de XAI	30/11/23	4/12/23
	Evaluación e interpretación de modelos obtenidos	5/12/23	18/12/23
	Entrega PEC3	19/12/23	25/12/23
	Entrega PEC3	25/12/23	25/12/23
Fase 4. Redacción de la memoria	Redacción de la memoria	26/12/23	16/1/24
	Entrega PEC4	26/12/23	8/1/24
	Entrega PEC4	16/1/24	16/1/24
	Entrega PEC4	17/1/24	31/1/24
	Entrega PEC4	17/1/24	21/1/24
Fase 5. Elaboración de la presentación	Realización de la presentación	17/1/24	21/1/24
	Entrega PEC5a	17/1/24	21/1/24
	Entrega PEC5a	21/1/24	21/1/24
	Entrega PEC5a	21/1/24	31/1/24
	Entrega PEC5a	21/1/24	31/1/24

Figura 1 - Planificación temporal

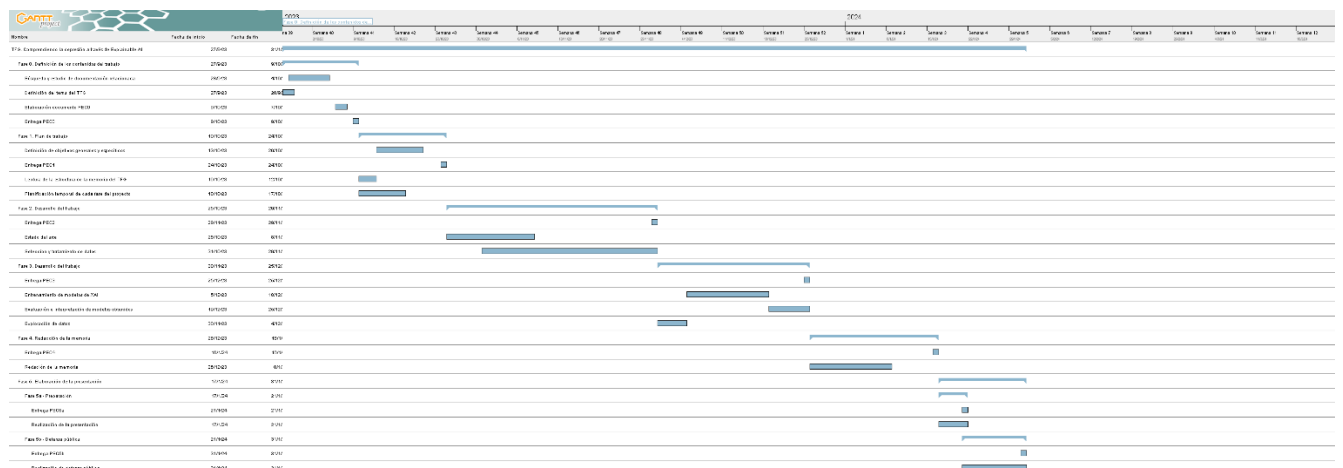


Figura 2 - Diagrama de Gantt

1.5 Breve resumen de productos obtenidos

Los productos generados por esta investigación se encuentran detalladamente documentados en esta memoria. Se aborda la formulación del problema, el análisis exhaustivo de la base de datos NHANES, la implementación de modelos de aprendizaje automático y técnicas de Inteligencia Artificial Explicable (XAI), así como las conclusiones derivadas de los análisis.

El código utilizado en la investigación está disponible en el repositorio de GitHub mencionado en el capítulo de Anexos. Además, se preparará una presentación visual que destaca los aspectos clave del trabajo y un video de exposición oral que proporciona una explicación detallada y dinámica de la investigación realizada.

1.6 Breve descripción de los otros capítulos de la memoria

Capítulo 2: Estado del Arte: En este capítulo, se llevará a cabo una revisión exhaustiva del estado del arte. Primero, se proporcionará una introducción detallada a la depresión, explorando definiciones, teorías y factores influyentes. Seguidamente, se analizará el *National Health and Nutrition Examination Survey* (NHANES) como fuente de datos clave. Por último, se abordará la *Explainable Artificial Intelligence* (XAI) y su relevancia para la comprensión de datos, particularmente en la relación de diversos factores con la depresión.

Capítulo 3: Metodología: Se detalla la metodología empleada en el estudio. En primer lugar, se aborda el tratamiento de datos, incluyendo procesos de limpieza, imputación y transformación. En segundo lugar, se explora la información contenida en los datos tratados mediante técnicas visuales. Finalmente, se describe la aplicación de modelos de ML y técnicas XAI, así como los resultados obtenidos.

Capítulo 4: Discusión: En este capítulo, se exponen y analizan los resultados obtenidos a través de modelos ML y técnicas XAI, estableciendo conexiones con estudios previos para contextualizar y validar los patrones identificados y las relaciones entre variables en el ámbito de la depresión.

Capítulo 5: Conclusiones: Se extraen las conclusiones fundamentales del trabajo. Se discuten los resultados obtenidos, se reflexiona sobre los logros alcanzados y se ofrecen sugerencias para investigaciones futuras.

Capítulo 6: Glosario: En este apartado se recopilan y definen los términos técnicos y acrónimos utilizados a lo largo del trabajo.

Capítulo 7: Bibliografía: Este capítulo contiene la lista completa de las fuentes bibliográficas, estudios y documentos consultados durante la investigación, siguiendo la Guía Normas APA 7ª edición.

Capítulo 8: Anexos: Incluye el enlace al código fuente utilizado en esta investigación, disponible en GitHub.

2 Estado del arte

2.1 Introducción a la Depresión

La depresión, conocida también como trastorno depresivo mayor, afecta a personas de todas las edades, razas y géneros, siendo un trastorno del estado de ánimo común pero serio. Se manifiesta con síntomas severos que impactan en la forma en que una persona siente, piensa y lleva a cabo sus actividades diarias. Entre los distintos tipos de depresión se encuentran el trastorno depresivo mayor, el trastorno distímico, la depresión perinatal, el trastorno afectivo estacional, la depresión con síntomas psicóticos y la asociada al trastorno bipolar.

Aunque la depresión puede afectar a cualquier persona, hay factores de riesgo como antecedentes personales o familiares de depresión, cambios de vida significativos y eventos traumáticos. Los síntomas de este trastorno incluyen un estado de ánimo persistente de tristeza, irritabilidad, cambios en el sueño y apetito, así como pensamientos de muerte o suicidio. El tratamiento implica medicamentos, psicoterapia, terapias de estimulación cerebral y, en algunos casos, terapias alternativas [\[5\]](#).

Investigaciones actuales establecen conexiones entre la depresión y diversos factores, como la actividad física [\[6\]](#), los trastornos del sueño [\[7\]](#), el estrés crónico [\[8\]](#), y un estilo de vida saludable [\[9\]](#). La Dra. Eva Selhub, experta en medicina cuerpo-mente, destaca la importancia de estos factores, señalando que el cerebro, siempre activo, requiere un suministro constante de "combustible" que proviene de los alimentos que ingerimos, y que la calidad de este combustible influye directamente en la estructura y función cerebral, y en última instancia, en el estado de ánimo [\[10\]](#).

Para este estudio, la fuente de análisis de datos será NHANES, permitiendo una exploración detallada de los vínculos entre la depresión y diversos aspectos de la salud.

2.2 NHANES como fuente de datos

La comprensión completa de la depresión requiere abordarla desde diversas perspectivas, desde sus manifestaciones clínicas hasta los factores que la desencadenan. En este contexto NHANES proporciona una amplia fuente de información que abarca diversos aspectos relacionados con la salud de la población.

Este conjunto de datos abarca categorías como datos demográficos, exámenes médicos, nutrición, medicación, datos de laboratorio y respuestas detalladas de los participantes a cuestionarios específicos. Estos componentes proporcionan una visión integral de la salud de los individuos, lo que resulta particularmente relevante al abordar la complejidad de la depresión.

Es relevante señalar que el conjunto de datos de medicaciones no ha sido incluido en el presente estudio. Esta decisión se tomó con el objetivo de focalizarnos en el análisis de otros atributos relevantes para la investigación sobre la relación entre la depresión y diversos factores de estilo de vida.

- **Datos demográficos:** Explorar variables como edad, género, raza y nivel educativo no solo ayuda a descubrir posibles patrones en la prevalencia de la depresión, sino que también proporciona una comprensión más profunda de cómo estos factores sociodemográficos pueden afectar la salud mental.
- **Exámenes médicos:** Desde la medición de la presión arterial hasta la evaluación de la salud dental y el índice de masa corporal (BMI), estos exámenes ofrecen una visión completa de la salud física estableciendo conexiones salud mental y física.
- **Datos de laboratorio:** Incorporar mediciones detalladas como colesterol, niveles de hierro y calcio abre la posibilidad de descubrir vínculos biomédicos entre condiciones médicas específicas y la depresión.
- **Dieta:** Proporcionar datos sobre la ingesta de nutrientes, calorías y dietas destaca la conexión innegable entre lo que comemos y nuestra salud mental.
- **Cuestionario:** Explorar aspectos como el consumo de alcohol, tabaco y actividad física añade una capa adicional de comprensión sobre el estilo de vida y su impacto en la salud mental.

La exploración detallada del conjunto de datos NHANES ha sido ampliamente utilizado en diversas investigaciones científicas; uso de modelos de aprendizaje automático para predecir riesgos de enfermedades cardiovasculares [\[11\]](#), clasificar individuos con pérdida de audición [\[12\]](#), y prever el riesgo de diabetes en jóvenes

[13]. Además, se han utilizado enfoques como redes neuronales artificiales para identificar casos de hipertensión [14] y algoritmos de aprendizaje profundo para detectar la depresión en veteranos de los Estados Unidos [15], entre muchos otros estudios. Estos ejemplos ilustran la versatilidad de NHANES como fuente de datos y su impacto en una amplia gama de disciplinas de la salud.

2.3 Explainable Artificial Intelligence (XAI)

La Inteligencia Artificial (IA) ha ganado una relevancia destacada, marcando su presencia en las rutinas cotidianas y desempeñando un papel crucial en la toma de decisiones en distintos ámbitos. Desde las recomendaciones de productos en plataformas como Amazon hasta la aplicación de avanzadas predicciones médicas en el sector de la salud, los algoritmos de la IA desempeñan un papel central en la sociedad actual.

En el ámbito de la atención médica, la IA se extiende a la predicción de eventos críticos, como, por ejemplo, la probabilidad de sufrir un infarto. Sin embargo, la opacidad inherente de muchos modelos de IA plantea la pregunta fundamental: ¿por qué se ha tomado una decisión específica? Para abordar esta cuestión, surge la Inteligencia Artificial Explicable (XAI), un conjunto de técnicas que busca transparentar y clarificar el razonamiento detrás de las decisiones automatizadas, facilitando la interpretación y comprensión por parte de los seres humanos.

En el contexto del aprendizaje automático (ML), se distinguen diversos modelos, denominados como *glass box* (caja transparente) y *black box* (caja negra). Entre los modelos transparentes se encuentran ejemplos como el Árbol de Decisión (*Decision Tree*) y la Regresión Logística (*Logistic Regression*), donde las decisiones que toma el modelo son comprensibles para los seres humanos. En contraste, los modelos de caja negra, como el Bosque Aleatorio (*Random Forest*), Vecinos Más Cercanos (*K-Nearest Neighbour*), Máquinas de Soporte Vectorial (SVM), Redes Neuronales (*Neural Network*), y Redes Neuronales Convolucionales (*Convolutional Neural Networks*), son opacos, ya que las decisiones no son visualizables ni comprensibles por humanos.

Es importante destacar que los modelos de caja negra suelen ofrecer una precisión superior en términos de predicciones en comparación con los modelos transparentes. En este contexto la Inteligencia Artificial Explicable (XAI) se convierte en una herramienta indispensable, permitiendo que las decisiones de los modelos de caja negra sean comprensibles. Esto es fundamental para lograr que los seres humanos entiendan el razonamiento detrás de las predicciones, facilitando así la aceptación de estas decisiones y promoviendo la toma de decisiones informadas.

En el ámbito de la inteligencia artificial, la explicabilidad desempeña un papel importante para diversos actores, desde el usuario final hasta los desarrolladores y auditores. En el contexto de una predicción de IA que indica que un paciente tiene una probabilidad del 90% de sufrir un infarto, XAI se manifiesta mediante la capacidad de ofrecer explicaciones detalladas. Por ejemplo, no solo indica que la probabilidad de un infarto es elevada, sino que contextualiza esta predicción al considerar factores, como, por ejemplo, edad superior a 65 años, un índice de masa corporal (BMI) superior a 35 y la presencia de hipertensión. Además, también ofrece

explicaciones contrafactuales, permitiendo explorar cómo cambios hipotéticos pueden influir en las predicciones futuras.

El papel de XAI se expande más allá de los usuarios finales, abarcando a desarrolladores y auditores en el ámbito de la inteligencia artificial. Para los desarrolladores, XAI les permite identificar áreas donde los modelos pueden no estar funcionando de manera óptima, y así poder realizar ajustes y diseñar modelos más efectivos. En el ámbito regulatorio, la importancia de XAI se destaca mediante la conformidad con regulaciones como el Reglamento General de Protección de Datos (GDPR). Este reglamento establece el derecho de los individuos a obtener explicaciones significativas sobre la lógica involucrada en decisiones automatizadas [\[16\]](#). Este requisito es especialmente crucial en situaciones de toma de decisiones automatizada, asegurando que los reguladores tengan las herramientas necesarias para verificar la conformidad con las normativas establecidas.

Para abordar la complejidad de la explicabilidad en los modelos de inteligencia artificial, han surgido diversas técnicas, entre las cuales destacan SHAP, LIME y WHAT-IF Tool de Google.

- **SHAP (*SHapley Additive exPlanations*)**: Basada en la teoría de juegos, SHAP asigna valores de contribución a cada característica en una predicción. Esta técnica es particularmente valiosa para modelos de caja negra, ya que desglosa las contribuciones de cada variable a nivel individual y global. SHAP ofrece explicaciones detalladas, permitiendo entender cómo cada característica influye en la salida del modelo.

Sus ventajas incluyen una fundamentación sólida en la teoría de juegos y la capacidad de proporcionar interpretaciones intuitivas. No obstante, se debe tener en cuenta que su implementación puede ser computacionalmente costosa, especialmente en conjuntos de datos grandes, y podría enfrentar desafíos en situaciones de alta dimensionalidad.

- **LIME (*Local Interpretable Model-agnostic Explanations*)**: En contraste, LIME adopta un enfoque local al generar explicaciones interpretables. Crea modelos más simples que aproximan el comportamiento del modelo complejo en una región específica del espacio de características. Esta aproximación es útil para comprender el modelo a nivel de instancia, proporcionando interpretaciones específicas para casos individuales.

Aunque es adaptable a cualquier modelo, su sensibilidad a los hiperparámetros y la posible dificultad al explicar modelos con datos one-hot encoding son desventajas que considerar. La ventaja de su enfoque local es contrarrestada por la necesidad de ajustar cuidadosamente los hiperparámetros para obtener resultados significativos en cada contexto de aplicación.

- **What-If tool de Google:** Esta herramienta proporciona una interfaz interactiva que facilita el análisis de cada punto de datos seleccionado. Permite explorar y comprender las predicciones del modelo al ajustar variables y observar el impacto en tiempo real. Al seleccionar un punto de datos específico, el WHAT-IF Tool ofrece detalles detallados sobre la información relacionada con ese punto, brindando una visión más clara y contextualizada de las decisiones del modelo.

En este proyecto, se adoptará una aproximación de caja blanca y mixta, utilizando métodos como *Decision Tree* y *Explainable Boosting Machine*, aprovechando funciones de la biblioteca InterpretML (*explain_local* y *explain_global*) para obtener explicaciones. Como método de caja negra, se modelará un *Random Forest*, empleando SHAP y LIME para las explicaciones.

Estas técnicas, como SHAP, LIME, y las funciones de InterpretML, se han convertido en herramientas fundamentales para mejorar la comprensión de los modelos de inteligencia artificial, ofreciendo enfoques diversos pero complementarios que permiten analizar y explicar las decisiones de manera transparente y accesible.

3 Metodología

3.1 Selección y tratamiento de datos

El conjunto de datos del presente estudio está compuesto por cinco archivos en formato CSV: *demographic*, *diet*, *examination*, *labs* y *questionnaire*, con un total de 9813 entradas. Para comprender cada variable incluida en estos conjuntos, se ha llevado a cabo una revisión continua de la información oficial disponible en la página web del NHANES correspondiente a los años 2013-2014 [\[17\]](#). Este recurso proporciona una explicación detallada de cada variable, así como del tipo de código o valor utilizado en cada caso.

En el proceso de tratamiento de datos, se ha empleado la función *detect_outliers_and_replace_with_nan()* específicamente en datos numéricos. Esta función detecta valores atípicos y los reemplaza con valores nulos (NaN) para evitar alterar las métricas estadísticas. Posteriormente, estos valores nulos se sustituyeron por la media de la columna correspondiente. Este enfoque asegura la integridad y coherencia de los datos, abordando los valores atípicos sin afectar significativamente las estadísticas descriptivas del conjunto de datos.

- **Objetivo del estudio (*Target*):**

El ámbito de este proyecto se enfoca en la comprensión de la depresión, lo que requiere la obtención de información para determinar si una persona está experimentando esta condición. Con este propósito, se seleccionó el cuestionario PHQ-9 [\[18\]](#), una herramienta de 9 preguntas ampliamente utilizada en Estados Unidos para evaluar la presencia y gravedad de la depresión. Las respuestas se puntúan en un rango de 0 a 27, donde un total de 10 o más sugiere la posible presencia de depresión.

El archivo *questionnaires.csv* contiene las respuestas a las preguntas codificadas desde DPQ010 hasta DPQ090. Se detectaron entre 4777 y 4786 valores nulos en estas respuestas, los cuales fueron excluidos para asegurar la integridad del estudio. Esta elección resulta fundamental, dado que la obtención completa de respuestas es esencial para llevar a cabo un análisis completo y preciso sobre la presencia de depresión en la muestra estudiada.

Con un total de 5372 entradas, de las cuales 4861 están etiquetadas como *not_depressed* y 511 como *depressed*, este conjunto de datos servirá como base para el análisis de las variables restantes. Se observa que el resultado de la variable objetivo no está balanceado, con un número reducido de muestras etiquetadas como *depressed* en comparación con *not_depressed*. En apartados posteriores, se abordará esta problemática mediante el empleo de la técnica *oversampling*.

- **Datos demográficos**

En el conjunto de datos *demographic.csv* se encuentra información individual y familiar. A continuación, se presenta una descripción general de las variables seleccionadas:

- *Gender*
- *Age*
- *Race*
- *Education Level*
- *Marital Status*
- *People in the household*
- *Annual household income*
- *Citizenship status*

Los valores de todas las variables son de tipo entero, donde cada número representa una categoría específica. Por ejemplo, la variable *Gender* tiene los valores 1 y 2, que corresponden a *male* y *female*, respectivamente, en ese orden. Los valores de estas variables se han sustituido por descripciones escritas para facilitar la comprensión, convirtiéndolos en variables categóricas. Los valores nulos, respuestas desconocidas o rechazadas, se han agrupado bajo la categoría *unknown* para mantener la coherencia en el conjunto de datos.

- **Datos de laboratorio**

El conjunto de datos "labs.csv" contiene información sobre resultados de laboratorio, incluyendo variables como colesterol, triglicéridos, calcio, entre otros. Las 32 variables seleccionadas se presentan en formato de valores numéricos de tipo flotante. Durante la exploración inicial de los datos, se observó la presencia de valores atípicos, los cuales fueron identificados y reemplazados por la media de cada variable correspondiente.

Este enfoque se implementó para mejorar la integridad de los datos y garantizar que las métricas estadísticas reflejen de manera más precisa la distribución de los resultados de laboratorio. Adicionalmente, se identificaron valores iguales a 0 en algunas variables, los cuales no son representativos ni válidos para datos de laboratorio. Estos valores también fueron sustituidos por la media respectiva de cada variable.

- **Datos de exámenes**

En el conjunto de datos *examination.csv* se encuentra datos de examen como pulso, presión arterial, entre otros. Se han seleccionado 5 variables, todas con tipo de datos flotante. Del mismo modo al que se realizó en los datos de laboratorio, se ha procedido a reemplazar los valores nulos y atípicos por la media de cada variable.

- **Datos de cuestionario**

En el conjunto de datos *questionnaire.csv* se han seleccionado aquellas variables consideradas importantes para el estudio y se han creado nuevas donde ha sido necesario realizar cálculos.

Consumo de alcohol: En el cuestionario, se incluyen preguntas relacionadas con las unidades de alcohol ingeridas en el último año y los hábitos de consumo, con respuestas afirmativas o negativas. Este apartado incluirá tres variables: el número de bebidas ingeridas el año pasado (*number_drinks_last_year*), para el cual se han realizado los cálculos necesarios en función de las respuestas obtenidas; alcohol alguna vez en la vida (*alcohol_lifetime*) y alcohol abusivo alguna vez en la vida (*binge_drinking_lifetime*), ambas con respuestas afirmativas o negativas. En total, no hay valores nulos, sin embargo, hay 11 valores codificados como desconocidos y se ha decidido codificarlos como una respuesta negativa (0).

Consumo de tabaco: En el cuestionario, se encuentran preguntas relacionadas con el consumo de tabaco. Este apartado comprenderá tres variables: edad a la que se empezó a fumar (*age_started_smoking*); el número de cigarrillos consumidos en el último mes (*number_cigarettes_last_month*), para el cual se han realizado los cálculos necesarios en función de las respuestas obtenidas y consumo de tabaco alguna vez en la vida (*smoke_lifetime*) con respuestas afirmativas o negativas.

La variable correspondiente a la edad a la que se empezó a fumar presenta un número considerable de valores nulos, concretamente, 3232 instancias. Dada la magnitud de esta ausencia de datos, finalmente se ha decidido codificarlo con el valor 0 para no desvirtuar el resultado. Además, se encontraron valores atípicos los cuales fueron sustituidos por la media de cada variable.

Análisis del sueño: En el cuestionario, se encuentran preguntas relacionadas con los hábitos de sueño. Este apartado incluirá dos variables: horas de sueño (*hours_sleep*) e historial de problemas de sueño (*trouble_sleeping_history*) con respuestas afirmativas o negativas. Se encuentra un número bajo de valores nulos y desconocidos. Para la variable correspondiente a las horas de sueño se ha optado por sustituirlos por la media y la correspondiente al historial de problemas de sueño el único valor desconocido se ha codificado como negativo.

Actividad física: El cuestionario aborda aspectos vinculados a las actividades diarias, actividades de ocio y comportamientos sedentarios. Se han identificado variables que reflejan la frecuencia de actividades laborales moderadas e intensas en una semana, así como actividades recreativas de similar naturaleza. A partir de estas variables, se ha generado una nueva clasificación que categoriza cada instancia en tres grupos distintos:

sedentario (*sedentary*), moderadamente activo (*moderately_active*) y activo (*active*). Además, se ha incorporado una variable que indica la duración diaria en minutos de actividad sedentaria. Se han detectado valores atípicos en la variable *minutes_sedentary_activity* que han sido sustituidos por la media de la variable.

- **Datos de dieta**

En el conjunto de datos *diet.csv*, se recopilan datos relacionados con la ingesta total de nutrientes en un día de una persona, abarcando aspectos como calorías, macronutrientes, minerales, vitaminas, entre otros. Aunque actualmente existen varios estudios que emplean diversas métricas para evaluar la calidad de la dieta, como *Nutrient-Rich Foods* (NRF), *Food Quality Score* (FQS), *Nutrient for Calorie* (NFC), *Calorie for Nutrient* (CFN), o *Nutrient Density Score* (NDS) [19], que se centran en porciones de comida o grupos alimenticios, estos datos específicos no están presentes en el conjunto actual.

Es importante señalar que la base de datos original en la página oficial de NHANES contiene información más detallada, aunque su extensión y complejidad podrían requerir un análisis único. Dado que en este estudio se evaluarán otros factores además de la nutrición, nos concentraremos en los datos disponibles, que incluyen la cantidad total de calorías, macronutrientes, vitaminas y minerales.

Para realizar la evaluación, se siguieron las pautas establecidas por la *Dietary Guidelines for Americans 2015-2020* [20]. Este marco establece las cantidades recomendadas de calorías, macronutrientes, minerales y vitaminas que un individuo debería consumir diariamente, teniendo en cuenta factores como el género, edad y el nivel de actividad física. Se realizó una comparación con los datos reales de consumo, asignando un punto por cada dato que se encontrara dentro de los márgenes recomendados.

En el caso de los macronutrientes, las recomendaciones presentan un rango aceptable, mientras que, para los nutrientes, se proporcionan valores mínimos. Se consideró adecuada una desviación absoluta de hasta un 10% para los nutrientes, permitiendo así una flexibilidad en la evaluación.

En el proceso, se han desarrollado funciones específicas para calcular las calorías recomendadas y para evaluar tanto los macronutrientes como los nutrientes consumidos. Se asignó el valor 0 a las entradas nulas, resultando en la creación de cuatro nuevas columnas: *macro_score*, *mineral_score*, *vitamin_score* y *total_nutrient_score*. Estas columnas proporcionan una puntuación que refleja la adherencia del consumo a las recomendaciones establecidas, facilitando así una evaluación de la calidad nutricional de la muestra.

Finalmente se ha obtenido un juego de datos con 5372 entradas y 59 columnas de la siguiente forma:

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 5372 entries, 73557 to 83727
Data columns (total 59 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   gender                                     5372 non-null   object
1   age                                       5372 non-null   int64
2   race                                     5372 non-null   object
3   education_level                         5372 non-null   object
4   marital_status                         5372 non-null   object
5   people_in_the_household                5372 non-null   int64
6   annual_household_income                5372 non-null   int64
7   citizenship_status                     5372 non-null   object
8   total_cholesterol                      5372 non-null   float64
9   HDL-cholesterol                       5372 non-null   float64
10  LDL-cholesterol                        5372 non-null   float64
11  triglyceride                           5372 non-null   float64
12  lymphocyte_percent                     5372 non-null   float64
13  monocyte_percent                       5372 non-null   float64
14  neutrophils_percent                   5372 non-null   float64
15  eosinophils_percent                   5372 non-null   float64
16  basophils_percent                     5372 non-null   float64
17  hematocrit_percent                    5372 non-null   float64
18  vitamin_B12                           5372 non-null   float64
19  albumin                               5372 non-null   float64
20  AST                                    5372 non-null   float64
21  ALT                                    5372 non-null   float64
22  blood_urea_nitrogen                   5372 non-null   float64
23  bicarbonate                           5372 non-null   float64
24  total_calcium                         5372 non-null   float64
25  CPK                                    5372 non-null   float64
26  chloride                              5372 non-null   float64
27  creatinine                            5372 non-null   float64
28  globulin                              5372 non-null   float64
29  glucose                               5372 non-null   float64
30  gamma_glutamyl_transferase            5372 non-null   float64
31  iron                                   5372 non-null   float64
32  potassium                             5372 non-null   float64
33  lactate_dehydrogenase                 5372 non-null   float64
34  sodium                                5372 non-null   float64
35  osmolality                            5372 non-null   float64
36  phosphorus                            5372 non-null   float64
37  total_bilirubin                       5372 non-null   float64
38  total_protein                         5372 non-null   float64
39  uric_acid                             5372 non-null   float64
40  pulse                                 5372 non-null   float64
41  systolic_blood_pressure                5372 non-null   float64
42  diastoli_blood_pressure                5372 non-null   float64
43  BMI                                    5372 non-null   float64
44  alcohol_lifetime                       5372 non-null   object
45  binge_drinking_lifetime                5372 non-null   object
46  number_drinks_last_year                5372 non-null   int64
47  age_started_smoking                   5372 non-null   int64
48  smoke_lifetime                         5372 non-null   object
49  number_cigarettes_last_month           5372 non-null   float64
50  hours_sleep                            5372 non-null   int64
51  trouble_sleeping_history               5372 non-null   object
52  minutes_sedentary_activity             5372 non-null   int64
53  physical_activity_level                5372 non-null   object
54  macro_score                           5372 non-null   int64
55  mineral_score                         5372 non-null   int64
56  vitamin_score                         5372 non-null   int64
57  total_nutrient_score                   5372 non-null   int64
58  depression_status                     5372 non-null   object
dtypes: float64(37), int64(11), object(11)
memory usage: 2.5+ MB
```


3.2 Exploración de datos

En esta sección, se llevará a cabo la exploración de los datos obtenidos en la fase de tratamiento con el objetivo de comprender la distribución y relaciones entre las diferentes variables en el conjunto de datos. A continuación, se presentan algunos aspectos destacados de la exploración:

Se examinó la distribución de cada variable en el conjunto de datos. A continuación, se puede ver una muestra de las variables más destacadas:

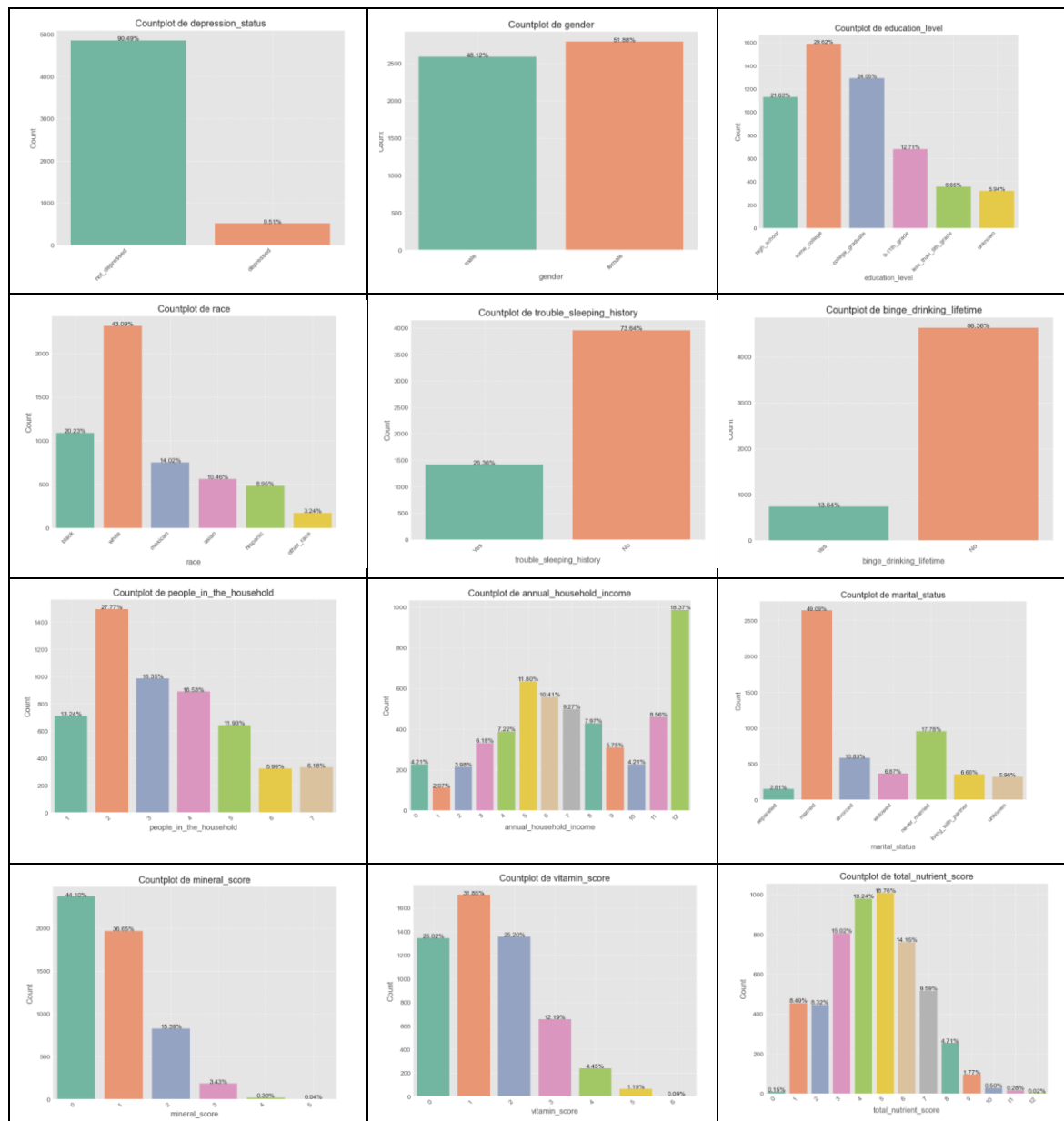


Figura 3 - Distribución variables

Según la Figura 3, se advierte un marcado desequilibrio en la variable *depression_status* ya que solo el 9.51% de las instancias están etiquetadas como casos de depresión. En relación con el género, se observa una distribución bastante equilibrada entre hombres y mujeres. Respecto al nivel educativo, las categorías con mayor incidencia son aquellas que cuentan con alguna titulación o graduación. Sin embargo, la suma de individuos sin estudios superiores supera a los graduados.

En cuanto a la variable étnica, la categoría predominante es la raza blanca, aunque se registra una presencia considerable de individuos de raza negra e hispánica. El 27% de la muestra ha experimentado algún trastorno del sueño, un porcentaje que podría considerarse elevado. Además, casi el 14% ha reportado abuso de alcohol en algún momento de su vida.

En relación con la composición familiar, la mayoría de las instancias corresponden a hogares con dos personas, aunque se destaca un porcentaje importante de familias numerosas o aquellas que tienen más de cinco miembros. Respecto a los ingresos anuales, se observa una concentración significativa en el grupo de ingresos más elevados (grupo 12), aunque existen porcentajes considerables en grupos inferiores.

La mayoría de las instancias corresponden a individuos casados. En el ámbito de la nutrición, se evidencia que la mayoría de las personas obtienen puntuaciones bajas, siendo importante señalar que a menor puntuación se asocia una dieta de menor calidad.

Siguiendo con el análisis exploratorio, se examinará la distribución de algunas de las variables numéricas continuas presentes en el conjunto de datos:

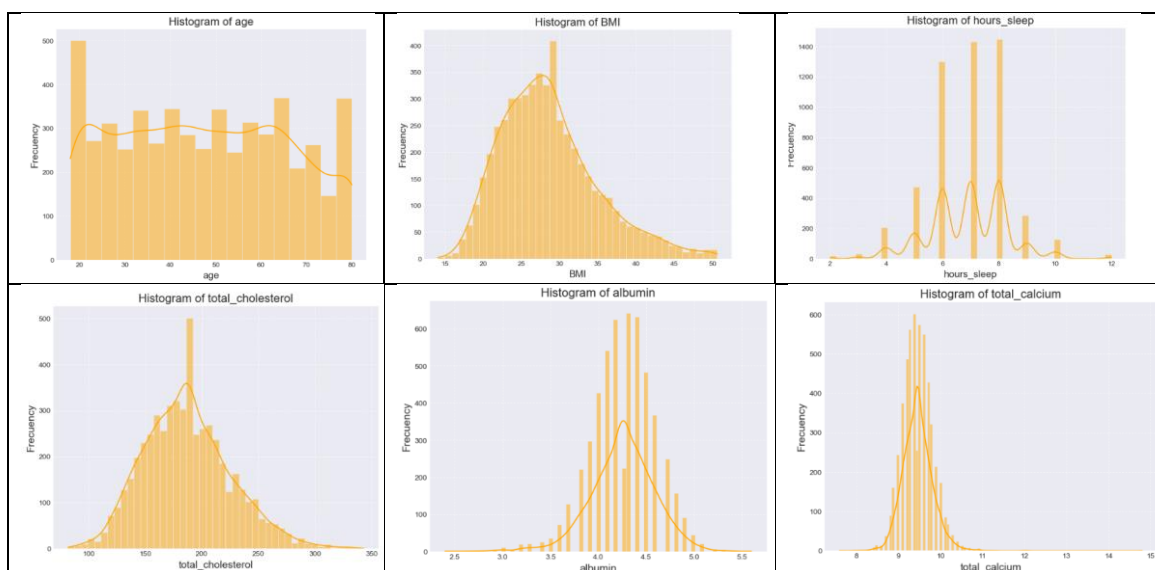


Figura 4 - Distribución variables numéricas

Las imágenes en la Figura 4 proporcionan una visión de la distribución de estas variables. En relación con la edad, se observa que la mayoría de los datos corresponden a individuos en los primeros años de sus 20, aunque la distribución abarca un rango amplio de edades en todo el conjunto.

La mayoría de las variables exhiben una distribución que se asemeja a la normal, con la excepción de algunas, como el índice de masa corporal (BMI), que muestra una asimetría hacia la derecha. Esta asimetría sugiere que hay casos con valores más altos de BMI, indicando la presencia de individuos con sobrepeso u obesidad en el estudio. En cuanto a las horas de sueño, la mayoría de los participantes informan dormir entre 6 y 8 horas, lo que concuerda con las recomendaciones generales de salud.

Lo siguiente en nuestra exploración es analizar detalladamente la relación entre las variables seleccionadas y la variable *depression_status*. Para visualizar estas relaciones, hemos utilizado gráficos de dispersión, gráficos de barras y categorías, que nos permiten identificar posibles patrones y correlaciones. A continuación, presentamos algunas de las gráficas más interesantes:

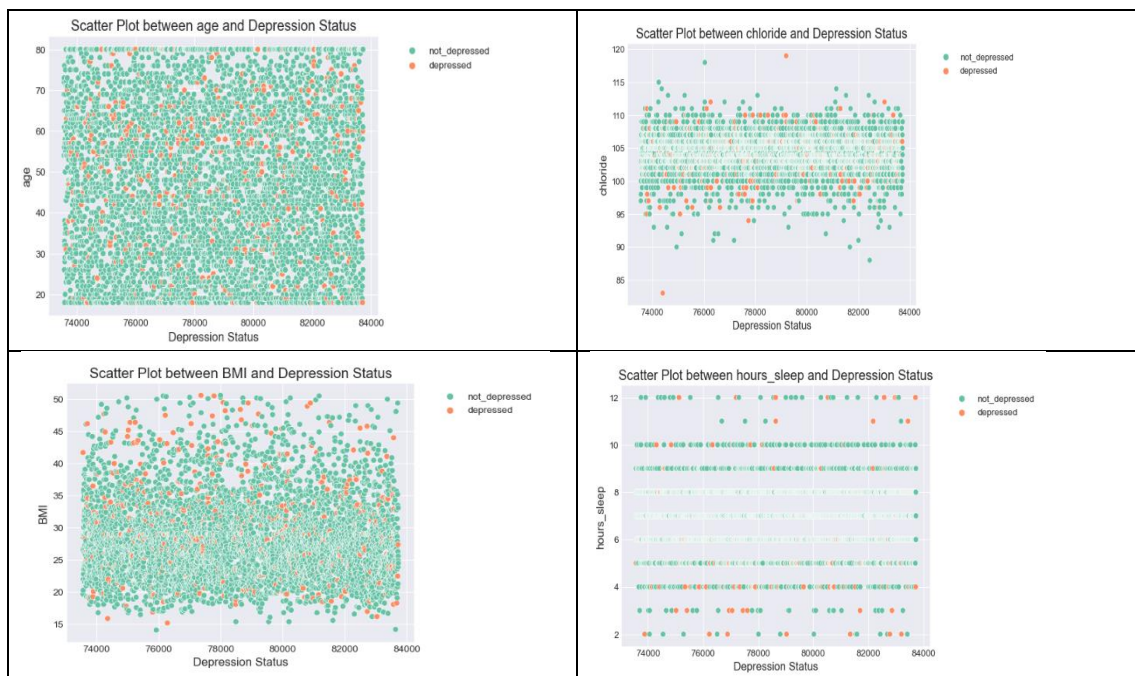


Figura 5 - Relación entre variables numéricas continuas y depresión (Scatterplot)

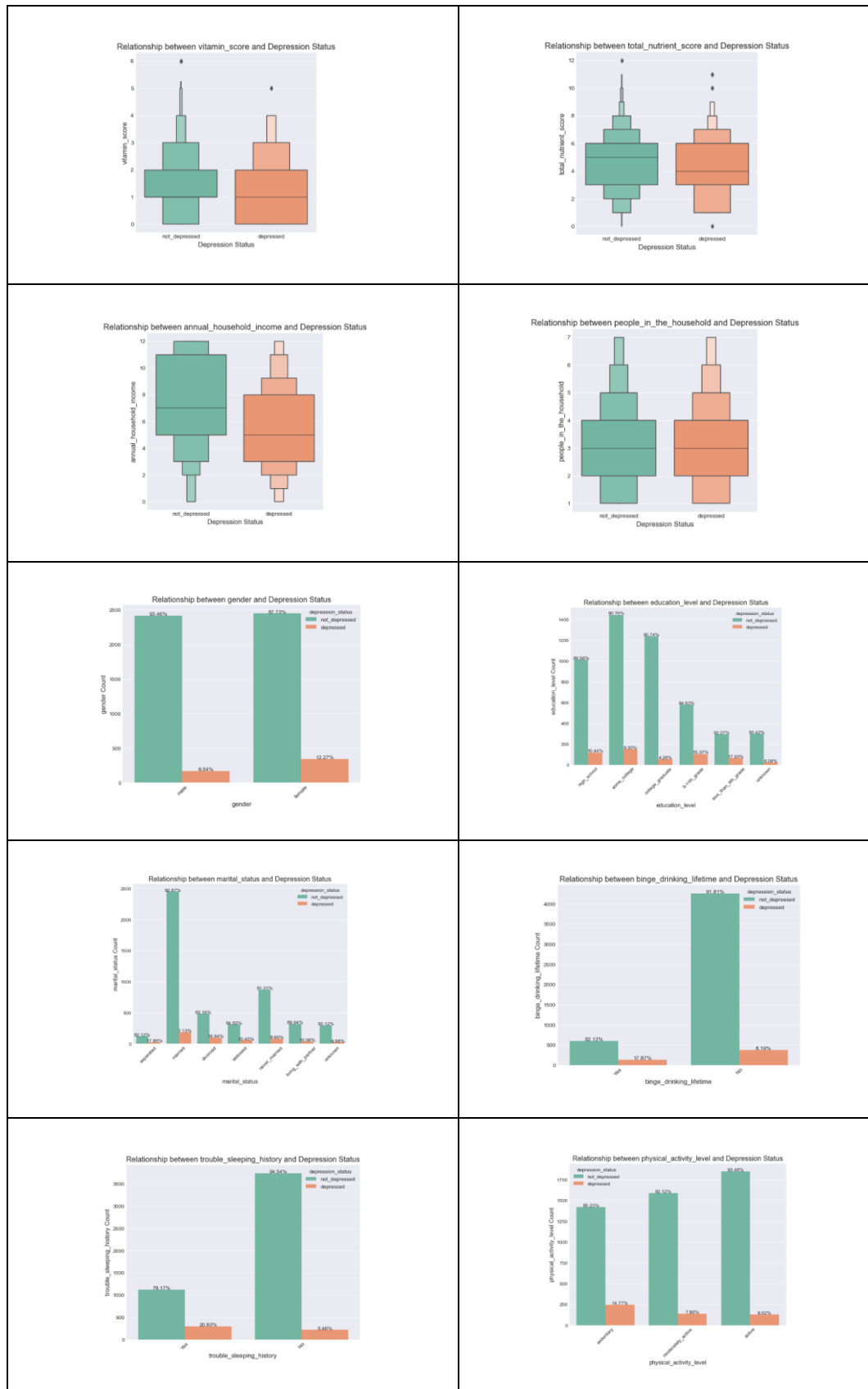


Figura 6 - Relación entre variables categóricas y Depresión (countplot)

A partir del análisis de las imágenes de las Figura 5 y 6, se observa una dispersión considerable en los casos de depresión en relación con la edad, aunque se identifica una concentración apreciable entre los 50 y 70 años. En cuanto al nivel de cloruro en sangre (*chloride*), se evidencian más casos de depresión tanto en niveles bajos como altos de este marcador. Respecto al Índice de Masa Corporal (BMI), se observa una concentración mayor de casos en valores altos, aunque también se registran casos en valores bajos. En lo que respecta a las horas de sueño, entre 6 y 9 horas se asocia con escasos casos de depresión, mientras que la mayoría se encuentra en aquellos que duermen menos de 6 horas o más de 10 horas.

En el análisis de las puntuaciones de vitaminas y la nutrición total en la dieta, se observa que aquellos con puntuaciones bajas presentan una mayor prevalencia de depresión. Con respecto a los ingresos anuales, se percibe que a ingresos más bajos corresponde una mayor proporción de depresión, y viceversa. En cuanto al número de personas que viven en el hogar, no se aprecian grandes diferencias en la posible incidencia.

En términos de género, se constata que las mujeres muestran una mayor prevalencia de depresión (12.27%) en comparación con los hombres (6.54%). El abuso de alcohol también se asocia con una mayor proporción. Los problemas de sueño también indican una mayor incidencia, y en cuanto a la actividad física, aquellos con un estilo de vida sedentario exhiben un porcentaje más alto, mientras que los más activos presentan menos casos.

3.3 Modelos *Machine Learning* y técnicas de Explainable Artificial Intelligence (XAI)

El análisis se inicia con el conjunto obtenido en la fase de tratamiento de datos que incluye información demográfica, resultados de laboratorio, exámenes, respuestas a cuestionarios y puntuaciones nutricionales. Con el objetivo de preparar los datos para el modelado, se realizaron transformaciones específicas, como la conversión de respuestas *yes*, *no* en valores binarios (0,1), y la aplicación de *one-hot -encoding* a las variables categóricas. Esto resultó en un *Dataframe* homogéneo con valores numéricos.

El juego de datos resultante se dividió en conjuntos de entrenamiento y test, siendo el 20% destinado al conjunto de prueba. Este conjunto fue utilizado para el estudio de todos los modelos, aunque fue necesario realizar modificaciones específicas para el *Decision Tree Classifier* y el *Explainable Boosting Classifier*, como se detallará en las siguientes secciones.

3.3.1 Decision Tree Classifier

Se procedió a entrenar el *Decision Tree Classifier* con el conjunto de entrenamiento original, pero los resultados indicaron una baja eficacia en la predicción de la clase de depresión. Al realizar una observación más detallada, se identificó un desbalance significativo entre las clases, especialmente con una presencia limitada de casos de depresión (clase 1).

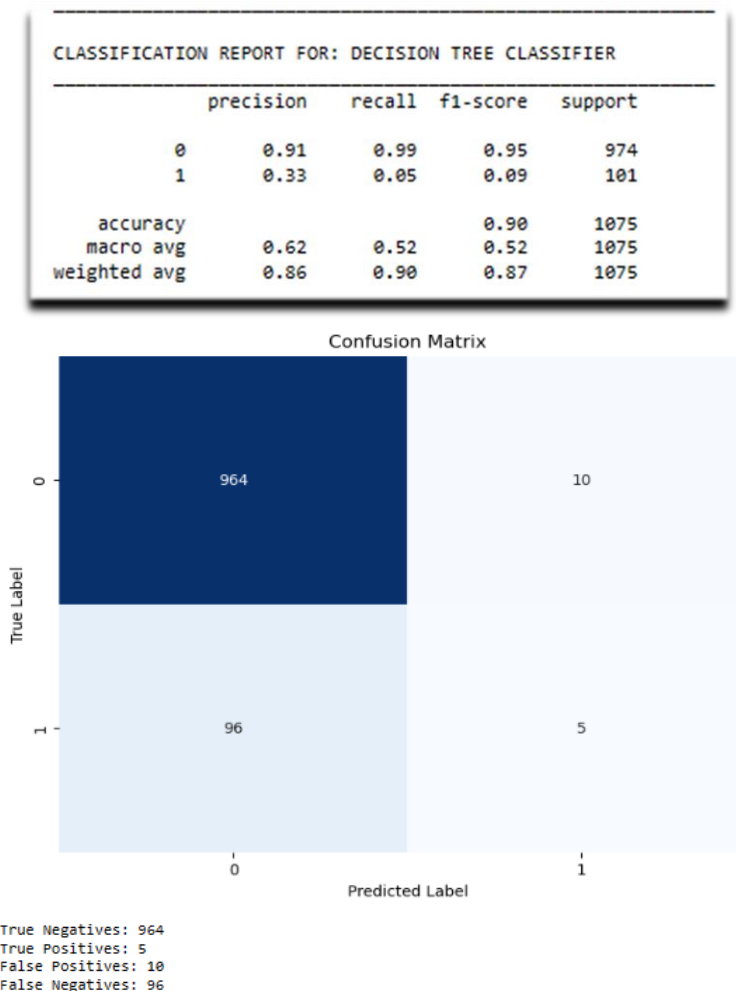


Figura 7 – Informe de clasificación y matriz de confusión DT (antes de aplicar técnica Oversample)

Para abordar este problema, se aplicó la técnica de *oversampling* en la clase minoritaria, lo que resultó en una mejora leve en las métricas del modelo.

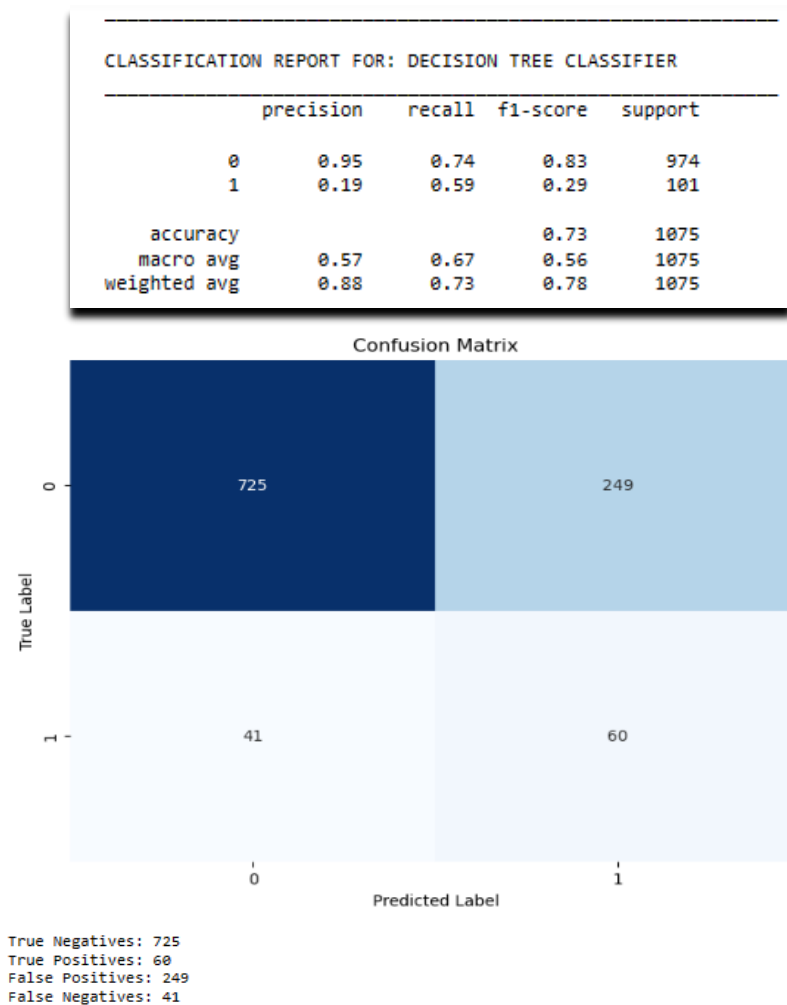


Figura 8 - Informe de clasificación y matriz de confusión DT (antes de aplicar técnica Oversample)

Se introdujo el paquete *InterpretML* que soporta tanto modelos de caja blanca, como modelos de caja gris y negra. Para trabajar con *InterpretML*, se importaron los paquetes necesarios:

```
from interpret.glassbox import (ClassificationTree, ExplainableBoostingClassifier)
from interpret import show
```

Para el análisis detallado del árbol de decisión se emplearon métodos como `explain_local` y `explain_global`. Estos métodos proporcionaron una visión profunda de la estructura del árbol, permitiendo un análisis granular del comportamiento de instancias individuales. En el caso del método `explain_local`, que proporciona explicaciones a nivel de instancia, se identificaron casos donde la ausencia de problemas de sueño y unos ingresos anuales elevados resultaron relevantes para predecir la no depresión. En contraste, factores como problemas de sueño o dificultades económicas fueron determinantes para predecir la depresión.

Ejemplo 1: En la instancia 2 (Figura 9), un individuo sin depresión con una precisión del 86%. La ausencia de problemas de sueño, ingresos anuales superiores al nivel 10,5 (>87.000 dólares) y un nivel de alanina aminotransferasa (ALT) inferior a 58 fueron clave para determinar esta predicción.

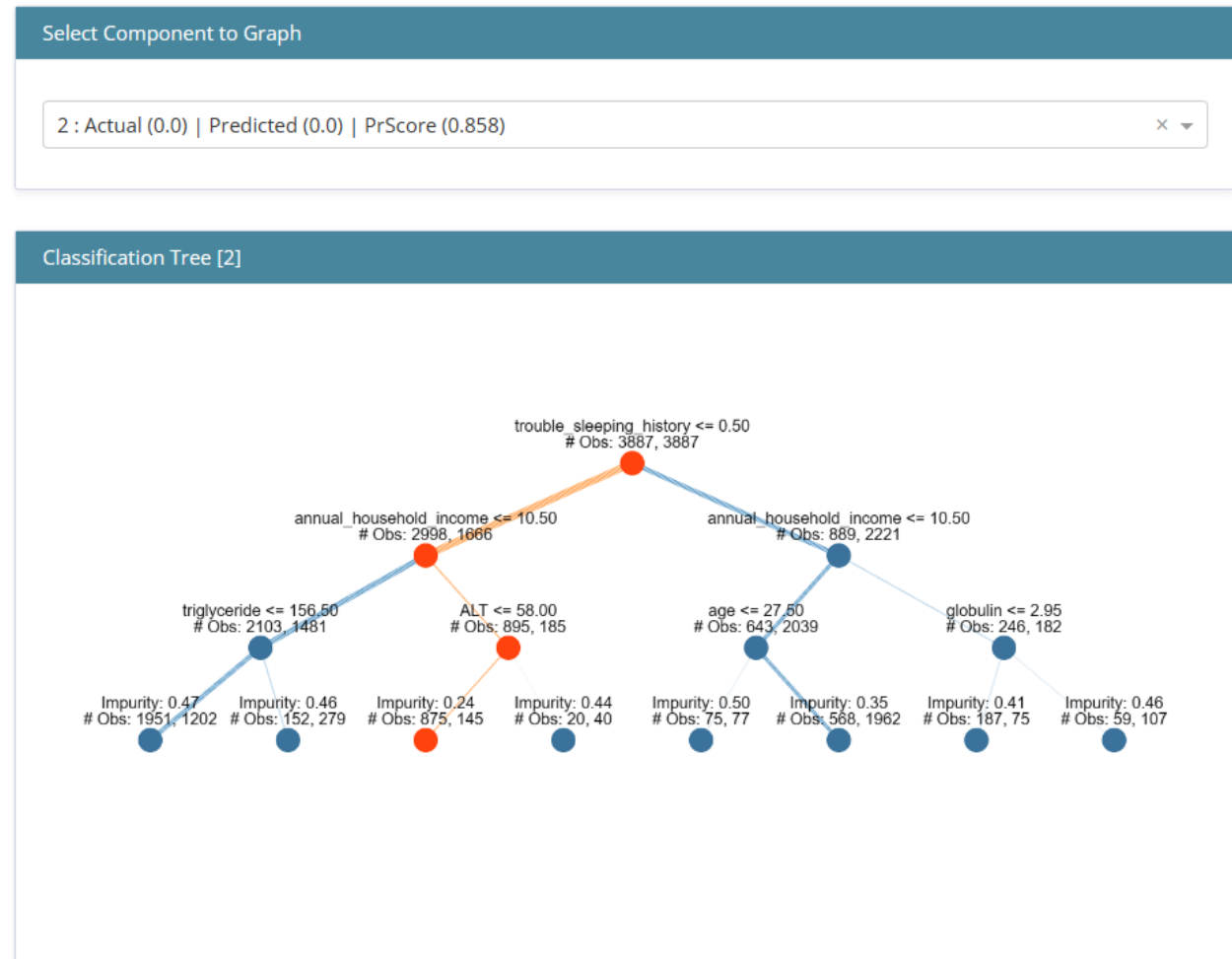


Figura 9 - Árbol de clasificación Instancia 2 (explain_local)

Ejemplo 2: En la instancia 108 (Figura 10), un individuo con depresión con una precisión del 78%. La presencia de problemas de sueño, dificultades económicas y una edad superior a 27,5 años fueron factores decisivos en esta predicción.

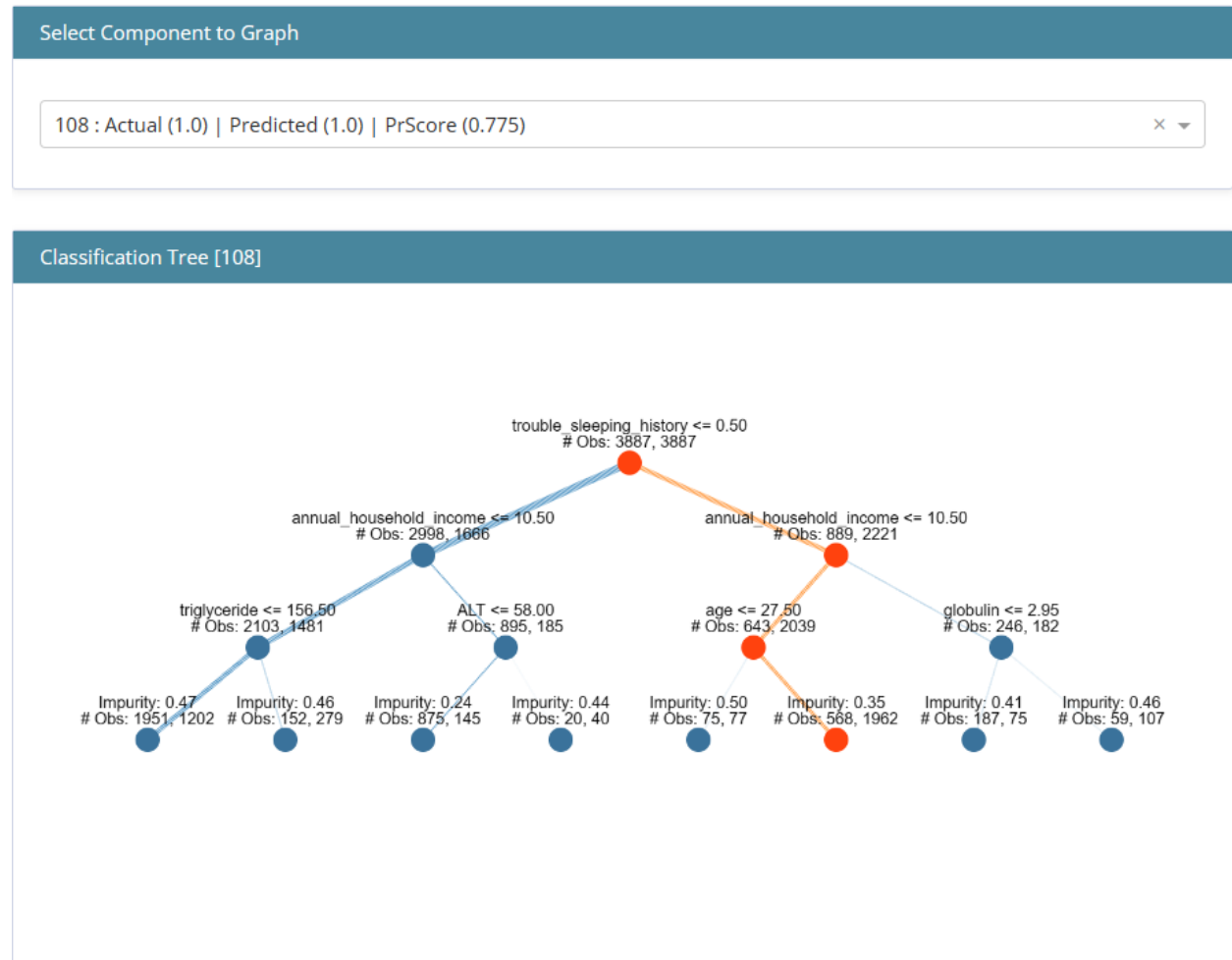


Figura 10- Árbol de clasificación Instancia 108 (explain_local)

En el caso del método `explain_global`, se exploraron las variables una a una, situándolas dentro del contexto del árbol de decisión (Figura 11). Este análisis global permitió entender cómo estas variables interactúan y afectan las decisiones del modelo en el ámbito más amplio del conjunto de datos.

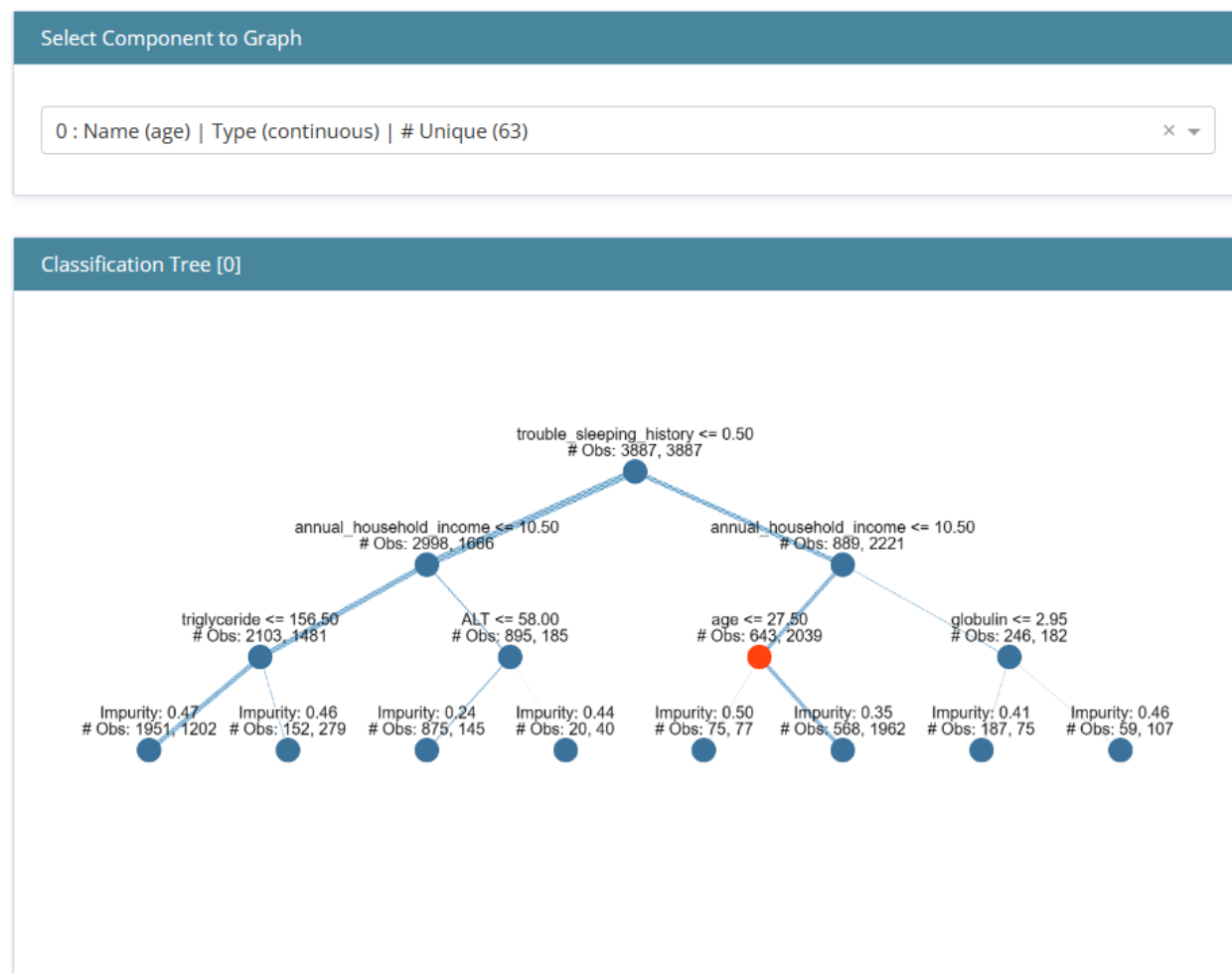


Figura 11- Árbol de clasificación global (comportamiento variable age)

3.3.2 Explainable Boosting Classifier (EBM)

Para el modelo EBM se utilizó el conjunto de datos balanceado. Aunque las métricas para la clase de depresión no alcanzaron niveles óptimos (Figura 12), la utilización de *InterpretML* posibilitó un análisis detallado de las variables influyentes.

	precision	recall	f1-score	support
0	0.92	0.90	0.91	974
1	0.19	0.22	0.20	101
accuracy			0.84	1075
macro avg	0.55	0.56	0.56	1075
weighted avg	0.85	0.84	0.84	1075

AUC-PR: 0.84

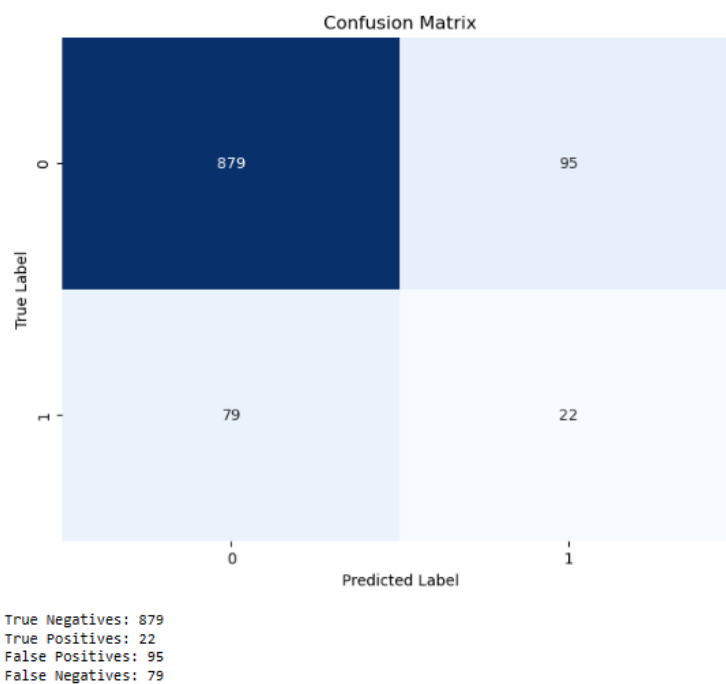


Figura 12 - Informe de clasificación y matriz de confusión EBM

Al explorar el modelo de forma local, se identificaron las variables más destacadas en la predicción de la depresión. En este contexto, las variables que contribuyen a la predicción de "No Depresión" se representan a la izquierda en color azul, mientras que las que influyen en la predicción de "Depresión" se sitúan a la derecha en color naranja. Al analizar las mismas instancias (2 y 108) que se examinaron en el árbol de decisión, se obtuvieron los siguientes datos:

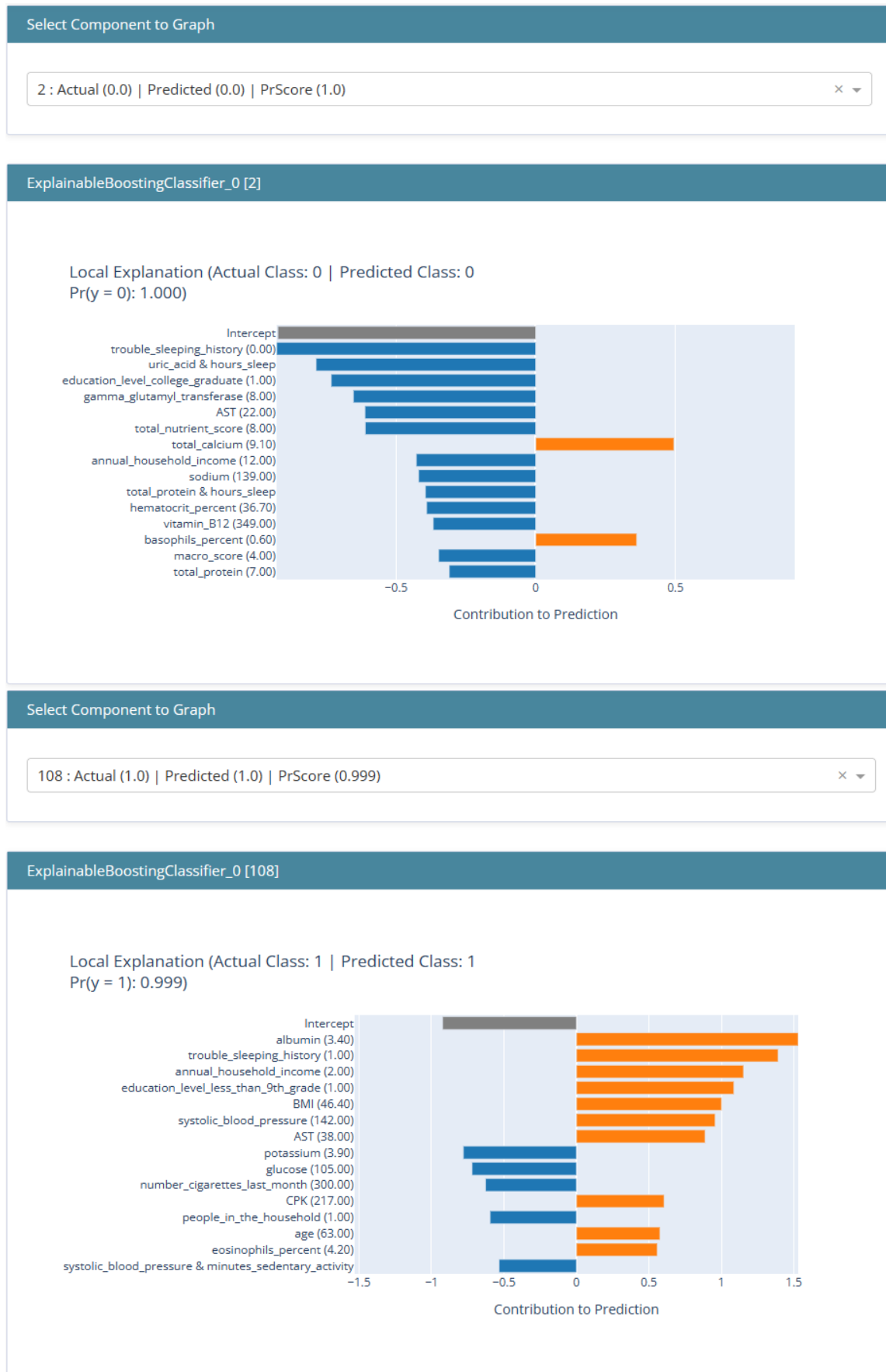


Figura 13 - EBM. Explicación local Instancias 2 y 108

En relación con las explicaciones globales, se destaca notablemente que el problema del sueño es la variable más significativa relacionada con la predicción de la depresión. Otras variables importantes incluyen los ingresos anuales en el hogar, la edad, la edad, el pulso, la puntuación de consumo de vitaminas en un día, entre otros (Figura 14).

Además, se observó que hay varios factores vinculados a los niveles en sangre. Resulta de interés profundizar en el análisis de cada uno de estos factores y explorar si existen estudios previos que hayan investigado su posible relación con la depresión. En caso de no existir, podría ser valioso considerar estos aspectos para investigaciones futuras.

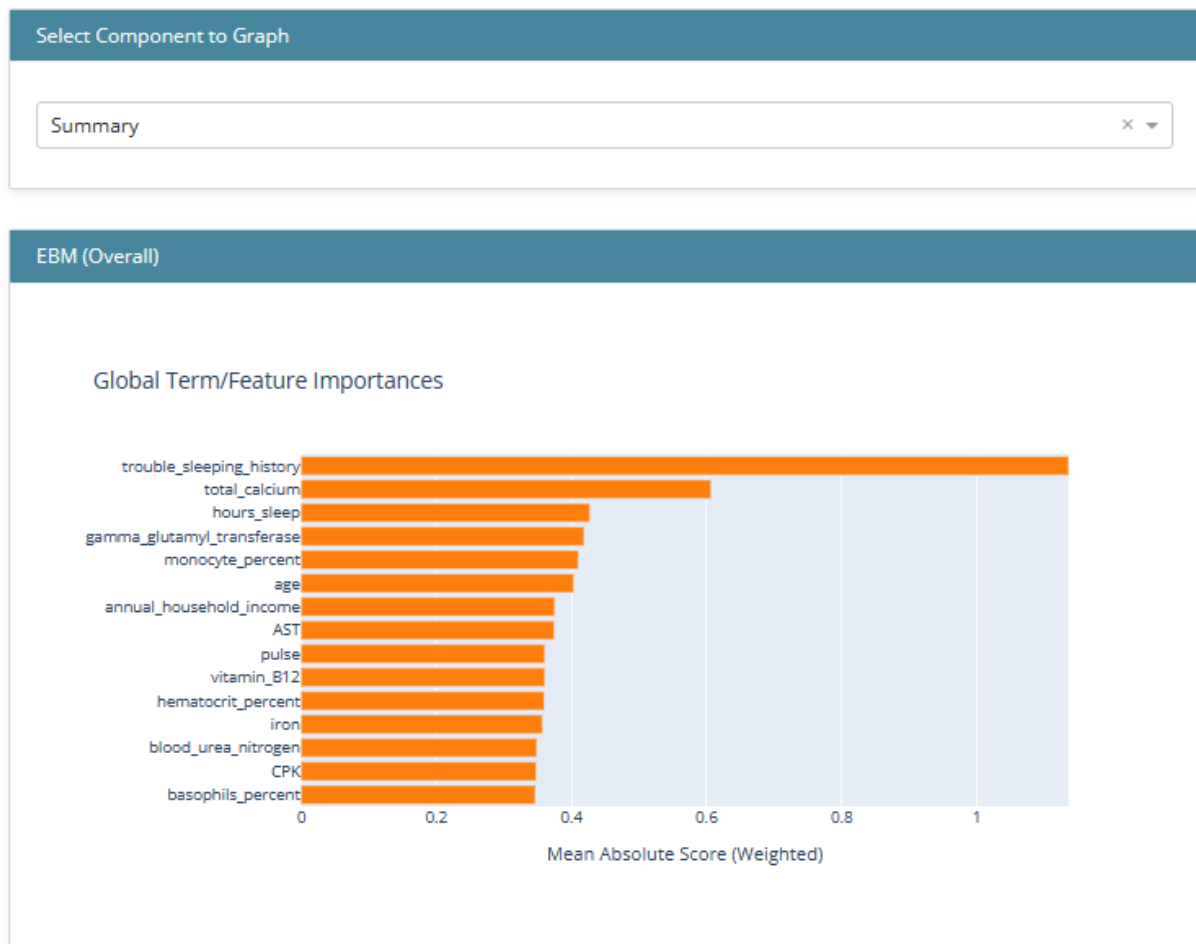


Figura 14 - EBM. Explicación global (*explain_global*)

Una funcionalidad adicional del método `explain_global` en un *Explainable Boosting Classifier* (EBM) es la posibilidad de analizar cada una de las variables y examinar su influencia en la variable objetivo. Por ejemplo, se puede explorar la relación entre las variables que indican la presencia de problemas del sueño o la puntuación en consumo de vitaminas en la dieta en el contexto de la predicción de la depresión.



Figura 15 - EBM. Comportamiento variables trouble_sleeping_history y vitamin_score

En las gráficas correspondientes a la Figura 15, se evidencia claramente la fuerte correlación entre los problemas de sueño y la presencia de depresión. Asimismo, en relación con el consumo de vitaminas en la dieta, se observa de manera destacada que, a mayor puntuación, es decir, una mejor calidad en la ingesta de vitaminas se asocia con una menor incidencia de depresión.

El análisis explicativo proporcionado por el paquete *InterpretML* sobre un modelo EBM resulta considerablemente más detallado y concluyente en comparación con un árbol de decisión. Aunque el árbol ofrece una visión más global y es visualmente impactante en un primer vistazo, puede pasar por alto detalles importantes. En cambio, el EBM permite un análisis mucho más detallado y preciso.

Por ejemplo, al comparar la instancia 108 en ambos modelos, el árbol indicaba que los ingresos eran inferiores al índice 10.5, correspondiente a ingresos menores a 87.000 dólares. Sin embargo, el EBM proporcionó un dato más preciso, mostrando que los ingresos pertenecían al índice 2, lo que se traduce en un rango de entre 5.000 y 9.999 dólares anuales en el hogar, evidenciando ingresos muy bajos.

Es importante señalar que, con el EBM, la precisión en ambas instancias mejoró al 99% en los dos casos. Esta mejora resulta particularmente interesante en el contexto de la predicción de la variable de depresión.

A pesar de estos beneficios, se destaca que el EBM es más lento en comparación con el árbol de decisión, lo que podría considerarse como una limitación. Sin embargo, la riqueza de detalles y la precisión en las explicaciones ofrecidas por el EBM pueden compensar esta desventaja en determinados contextos de análisis.

3.3.3 Random Forest

La siguiente etapa del análisis se enfoca en el modelo de caja negra *Random Forest*. Los resultados obtenidos resultaron desfavorables, lo que motivó ajustes exhaustivos de los parámetros para mejorar el rendimiento. En este escenario, no se aplicó la técnica de *oversampling*, ya que el modelo incorpora un parámetro dedicado a este propósito. La búsqueda de los mejores parámetros se llevó a cabo mediante el uso de *GridSearchCV*. A pesar de que se logró mejorar el rendimiento en comparación con el árbol de decisión y el EBM, los resultados obtenidos aún no alcanzan niveles óptimos.

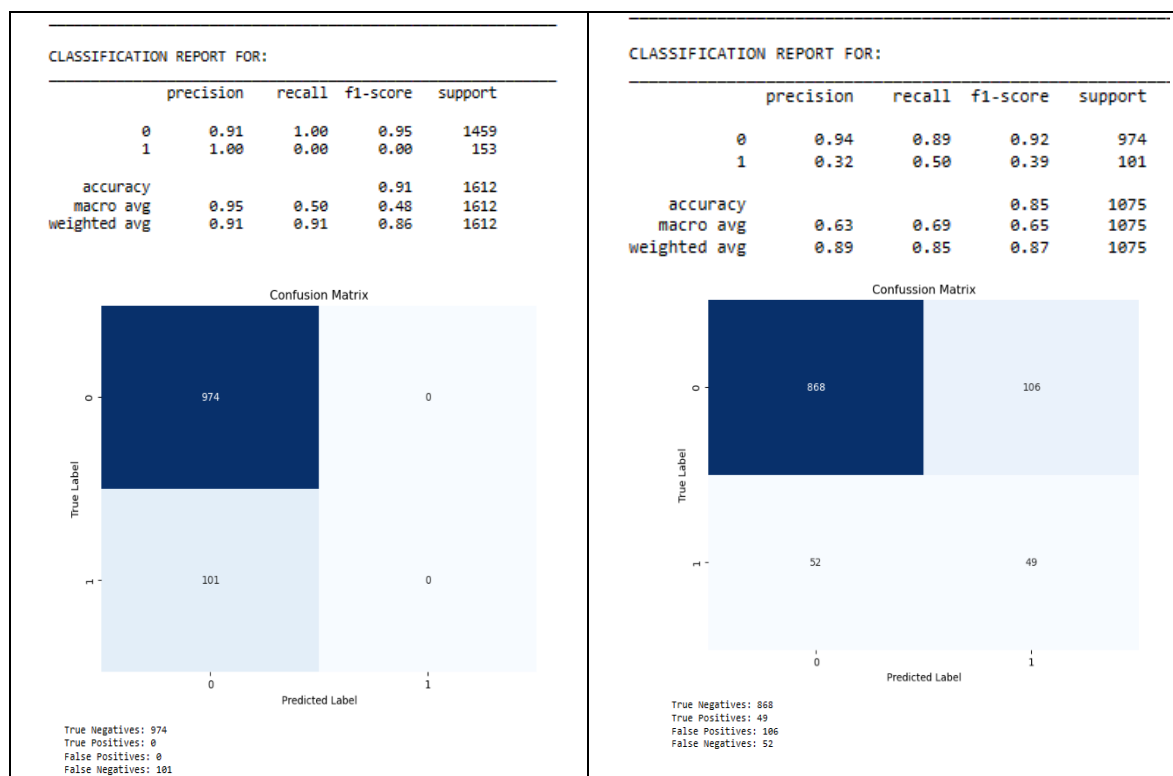


Figura 16 - Informe de clasificación y matriz de confusión (antes y despues de aplicar *GridSearchCV*)

Conforme al enfoque presentado en el capítulo 2.3, donde se abordaron las técnicas de Explicabilidad, estas surgieron con el propósito de comprender y explicar las decisiones de los modelos de caja negra. En consonancia con este propósito, se ha empleado SHAP y LIME sobre el modelo entrenado de *Random Forest*. Estas técnicas han proporcionado interpretaciones más detalladas y comprensibles sobre las predicciones del modelo, contribuyendo así a la transparencia y explicabilidad de sus resultados.

3.3.4 LIME

La introducción a las técnicas de explicabilidad XAI se inicia con LIME, aplicada específicamente al modelo entrenado *Random Forest*. Como se expuso en el capítulo sobre técnicas de explicabilidad en la inteligencia artificial (XAI), LIME constituye una metodología diseñada para proporcionar explicaciones a modelos de caja negra. Para su implementación, fue necesario instalar la biblioteca correspondiente mediante el comando `pip install lime` e importar las librerías necesarias:

```
from interpret.blackbox import LimeTabular
```

En este estudio, también se recurrió al paquete *InterpretML* para generar explicaciones interactivas. En primer lugar, se creó un objeto `LimeTabularExplainer`, al cual se le pasaron las predicciones generadas por el modelo RF en el conjunto de entrenamiento:

```
# Create a LIME explainer
explainer = lime.lime_tabular.LimeTabularExplainer(X_train.values,
                                                    class_names=['Not Depressed', 'Depressed'],
                                                    feature_names=X_train.columns.tolist())
```

A continuación, se escoge la instancia deseada para generar explicaciones, se generan y visualizan:

```
# Choose the specific instance for explanation
instance_to_explain = X_test.iloc[instance_to_explain]

# Generate the explanation
explanation = explainer.explain_instance(instance_to_explain.values,
                                       model.predict_proba, num_features=num_features)

# Visualize the explanation
explanation.show_in_notebook(show_all=False)
```

En este contexto, se han elaborado explicaciones detalladas para las instancias 2 y 108, las cuales se presentan de manera visual en la Figura 17.

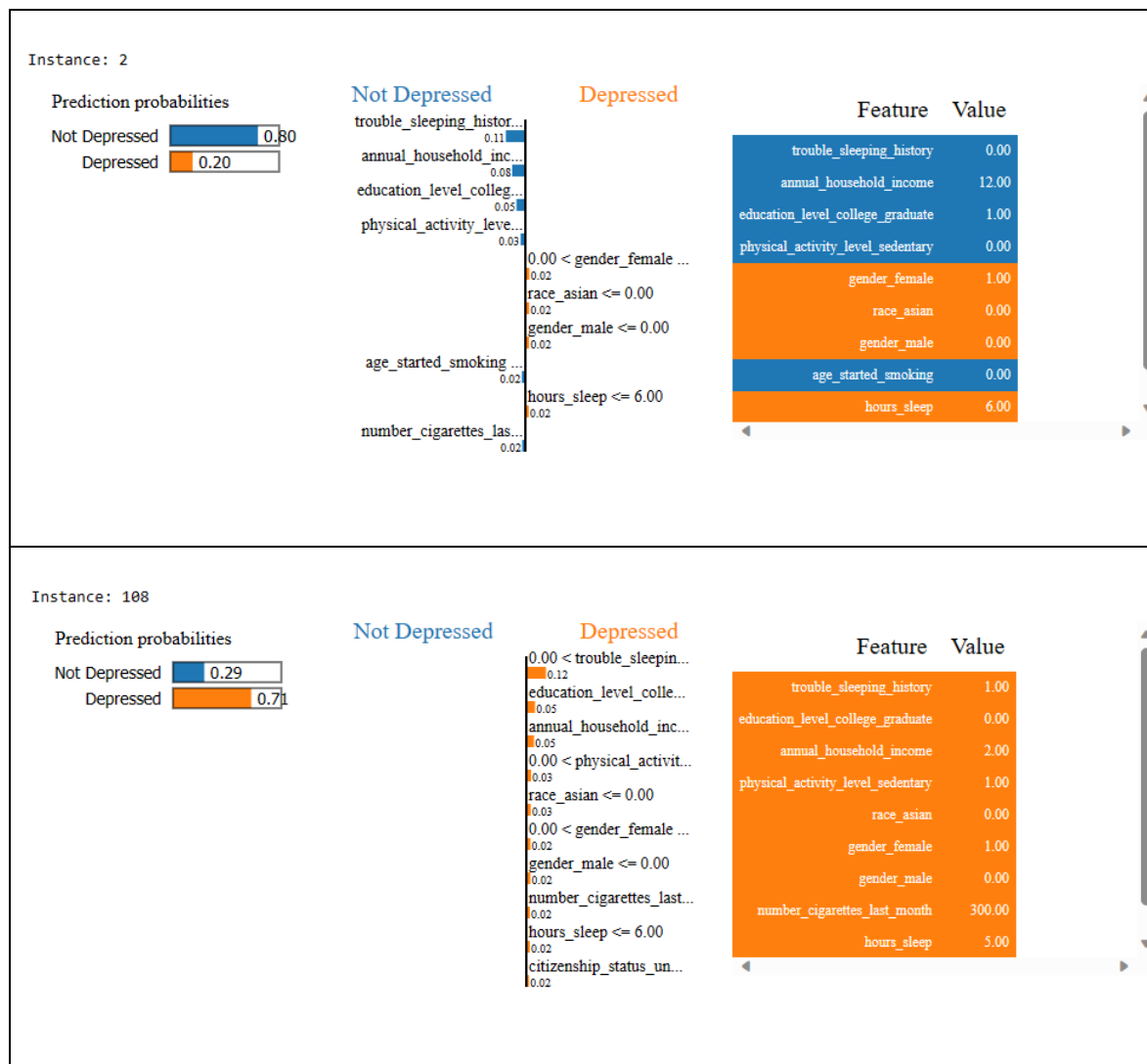


Figura 17- LIME. Explicaciones para instancias 2 y 108

Este gráfico resulta de fácil comprensión, permitiendo visualizar la contribución de cada variable a la clase objetivo. En la instancia visualizada se destaca nuevamente la importancia de variables como los trastornos del sueño, nivel educativo, género, actividad física, entre otros.

3.3.5 SHAP

La segunda técnica XAI explorada es SHAP. Para implementar esta técnica Como segunda técnica XAI fue necesario instalar la biblioteca correspondiente mediante el comando `pip install shap`. A continuación, se creó un objeto denominado `explainer` y se invocó el método `shap_values` sobre este objeto para calcular los valores SHAP de todos los atributos, empleando el modelo entrenado *Random Forest*.

```
start_index = 4
end_index = 5
explainer = shap.TreeExplainer(model)
shap_values = explainer.shap_values(X_test.iloc[start_index:end_index])
```

Es importante destacar que SHAP es un modelo local, lo que implica que su alcance se limita a una instancia específica. No obstante, es posible crear un modelo global combinando y agregando dos o más instancias. En este contexto, se procedió a analizar la instancia 4 como ejemplo, y los valores SHAP generados para esta instancia son los siguientes:

```
(1, 79)
[array([[ 6.82752343e-03,  1.06306602e-03,  5.72689387e-02,
  1.70967475e-03,  5.05453380e-03,  1.40151092e-03,
  3.37147795e-03,  3.69721795e-03,  1.08439096e-03,
  4.15960997e-03,  6.37547751e-04,  5.58835503e-04,
 -3.87273313e-03,  5.61173270e-03,  5.73795787e-04,
  2.23202655e-03, -1.80046257e-03,  1.53675175e-03,
  2.24446953e-03, -1.04108143e-03,  7.36505257e-05,
  1.83718641e-03,  8.52074974e-04,  1.01556022e-03,
  2.68039721e-03,  1.37106382e-02, -1.82877821e-03,
 -2.50568657e-03,  4.16850644e-05,  9.61995033e-04,
  2.14587828e-03,  4.28490432e-03,  8.90709975e-04,
  1.95317247e-03, -3.53350505e-03,  1.53323578e-03,
  5.81156356e-03,  2.71694963e-03,  1.06874337e-02,
  1.11640969e-04,  6.01077858e-03,  1.35502864e-03,
  1.28036957e-02,  8.61766948e-03,  7.43339909e-03,
  4.11834856e-03,  4.72723314e-02,  2.46516675e-04,
  2.70479971e-03,  4.23453274e-04,  2.66679669e-03,
  4.67515429e-03,  6.10334704e-05,  5.04617455e-05,
  6.63819336e-06,  2.39696127e-03,  4.54253904e-02,
  5.52203983e-05,  1.63754983e-03, -1.07163845e-05,
  4.85522354e-04, -8.44151492e-03, -6.47434936e-03,
  1.41393124e-03,  3.02518736e-05,  1.07680366e-02,
 -5.31995841e-05,  2.08722128e-04,  1.30560876e-04,
  7.71069905e-04,  3.20044006e-03, -1.01348256e-03,
  1.42018145e-02, -6.62253595e-04,  8.64810991e-05,
  2.94368111e-04,  5.82276742e-04,  1.54155707e-04,
  3.40073562e-05]])], array([[-6.82752343e-03, -1.06306602e-03, -5.72689387e-02,
 -1.70967475e-03, -5.05453380e-03, -1.40151092e-03,
 -3.37147795e-03, -3.69721795e-03, -1.08439096e-03,
 -4.15960997e-03, -6.37547751e-04, -5.58835503e-04,
 -3.87273313e-03, -5.61173270e-03, -5.73795787e-04,
 -2.23202655e-03,  1.80046257e-03, -1.53675175e-03,
 -2.24446953e-03,  1.04108143e-03, -7.36505257e-05,
 -1.83718641e-03, -8.52074974e-04, -1.01556022e-03,
 -2.68039721e-03, -1.37106382e-02,  1.82877821e-03,
  2.50568657e-03, -4.16850644e-05, -9.61995033e-04,
  2.14587828e-03, -4.28490432e-03, -8.90709975e-04,
  1.95317247e-03,  3.53350505e-03, -1.53323578e-03,
  5.81156356e-03, -2.71694963e-03, -1.06874337e-02,
 -1.11640969e-04, -6.01077858e-03, -1.35502864e-03,
 -1.28036957e-02, -8.61766948e-03, -7.43339909e-03,
```

```
-4.11834856e-03, -4.72723314e-02, -2.46516675e-04,
-2.70479971e-03, -4.23453274e-04, -2.66679669e-03,
-4.67515429e-03, -6.10334704e-05, -5.04617455e-05,
-6.63819336e-06, -2.39696127e-03, -4.54253904e-02,
-5.52203983e-05, -1.63754983e-03, 1.07163845e-05,
-4.85522354e-04, 8.44151492e-03, 6.47434936e-03,
-1.41393124e-03, -3.02518736e-05, -1.07680366e-02,
5.31995841e-05, -2.08722128e-04, -1.30560876e-04,
-7.71069905e-04, -3.20044006e-03, 1.01348256e-03,
-1.42018145e-02, 6.62253595e-04, -8.64810991e-05,
-2.94368111e-04, -5.82276742e-04, -1.54155707e-04,
-3.40073562e-05]]]
```

Se está analizando una instancia con 79 variables (1,79). Estos valores se almacenan en un array, donde están agrupados en dos conjuntos. El primer conjunto son los valores SHAP clasificados como clase 0 (color azul), mientras que el segundo corresponde a los agrupados en la clase 1 (color rojo).

Para hacer que estos datos sean informativos y explicables, se procedió a generar gráficos. A continuación, se presenta el código para mostrar un gráfico SHAP para la instancia 2:

```
# Display the SHAP force plot
prediction = model.predict(X_test[start_index:end_index])[0]
print(f"Prediction: {prediction}")
display(shap.force_plot(explainer.expected_value[1], shap_values[1],
                        X_test.iloc[2:3]))
```



Figura 18 - SHAP Force Plot instancia 2

El gráfico de fuerza SHAP presenta diversos valores, con el valor base ubicado en el centro. En el rango de 0.5 a 1, los valores se dirigen hacia la clase 1, que indica depresión, mientras que de 0.5 a 0 se asocian con la clase 0, es decir, no depresión. Esto permite visualizar la contribución de cada atributo a la clasificación de depresión o no depresión. En este ejemplo, la ausencia de problemas relacionados con el trastorno del sueño, ingresos altos, educación superior y no tener un estilo de vida sedentario son los principales factores que contribuyen a clasificarlo como "No depresión".

Se utilizó el siguiente código para generar un resumen gráfico de SHAP por variables, mostrando de mayor a menor importancia que contribuyen a cada clase:

```
display(shap.summary_plot(shap_values, X_test))
```

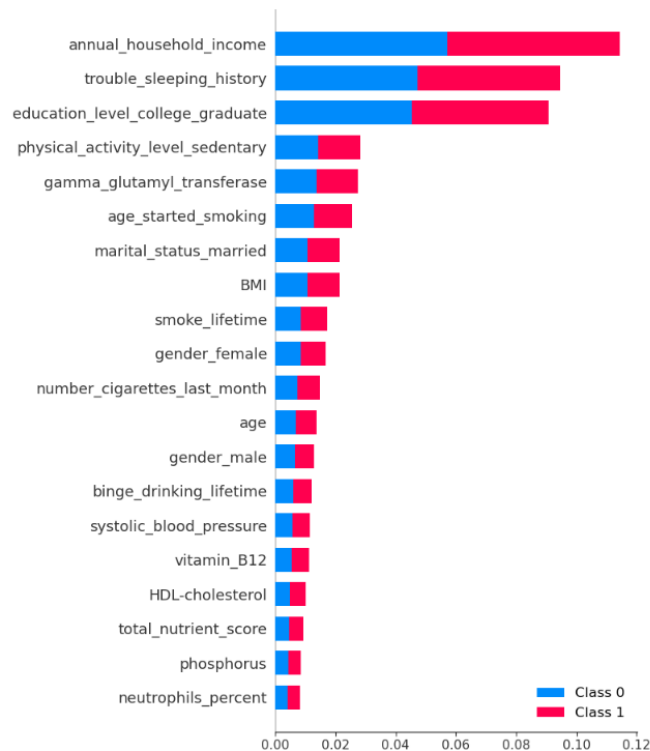


Figura 19 - Gráfico sumario SHAP para instancia 2

Este gráfico proporciona una visión general de la contribución de cada variable al modelo, ordenadas por importancia. Puede observarse cómo cada variable afecta las predicciones para ambas clases, permitiendo una comprensión más clara de los factores influyentes en el modelo.

Para generar explicaciones globales, se seleccionó un rango de índices, en este caso 20 instancias. La captura de imagen resultante (Figura 20) muestra en el eje X las instancias del 0 al 20. En el eje Y, el valor base es 0.5, indicando por encima de este valor la clase 0 (no depresión) y por debajo la clase 1 (depresión). Este gráfico proporciona explicaciones tanto locales como globales. Al desplazarse por la gráfica, se pueden observar los detalles de la instancia seleccionada, ofreciendo explicaciones más detalladas. En términos generales, el gráfico destaca la variable de problemas del sueño como un indicador significativo para la depresión, evidenciado por el área más extensa tanto en el lado positivo como en el negativo del eje Y (azul: no depresión, rojo: depresión).

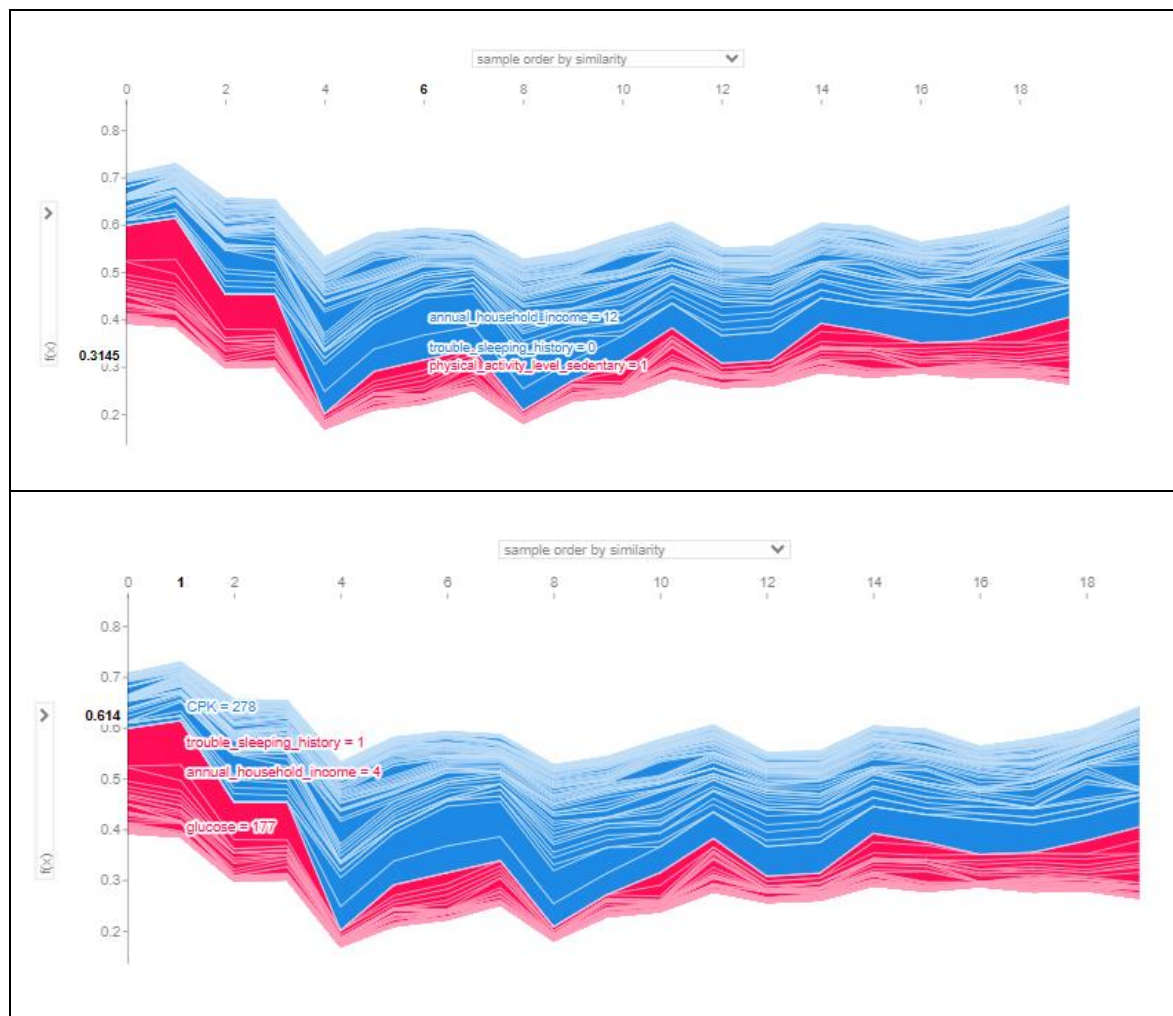


Figura 20- - SHAP Force Plot instancias de 0 a 19 (explicación global)

Además, se puede seleccionar un atributo específico para observar su comportamiento con respecto a la salida, por ejemplo, *total_nutrient_score* (Figura 21). Claramente se aprecia cómo a una puntuación más alta se asocian menos instancias con depresión.

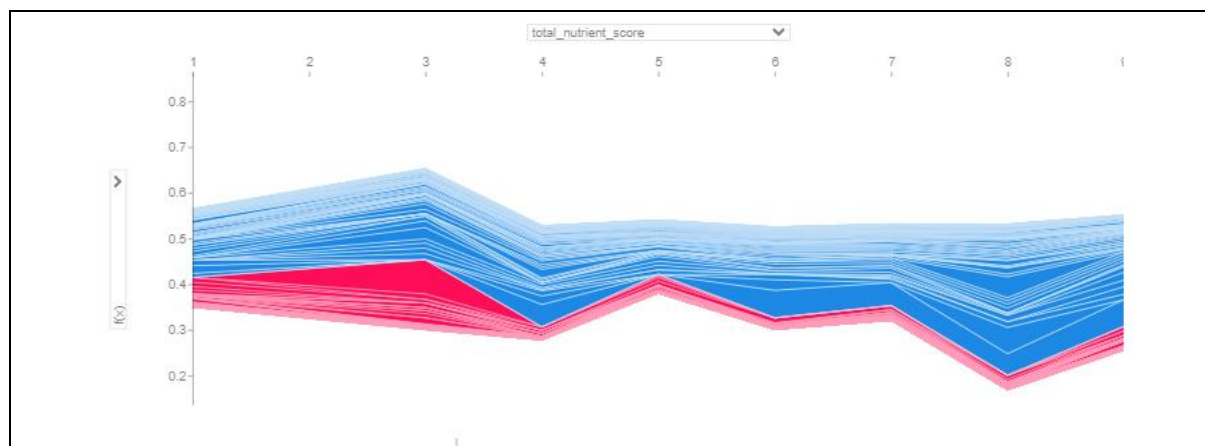


Figura 21 - SHAP Force Plot. Comportamiento variable *total_nutrient_score*

En el gráfico mencionado, destaca otra característica interesante. Al seleccionar la variable *total_nutrient_score* en el eje Y (Figura 22), se observa claramente que, para valores bajos de puntuación en nutrientes, las probabilidades de tener depresión son significativamente altas, mientras que los valores altos indican lo contrario. Este análisis se puede extender a las demás variables para evaluar su contribución a la variable objetivo.

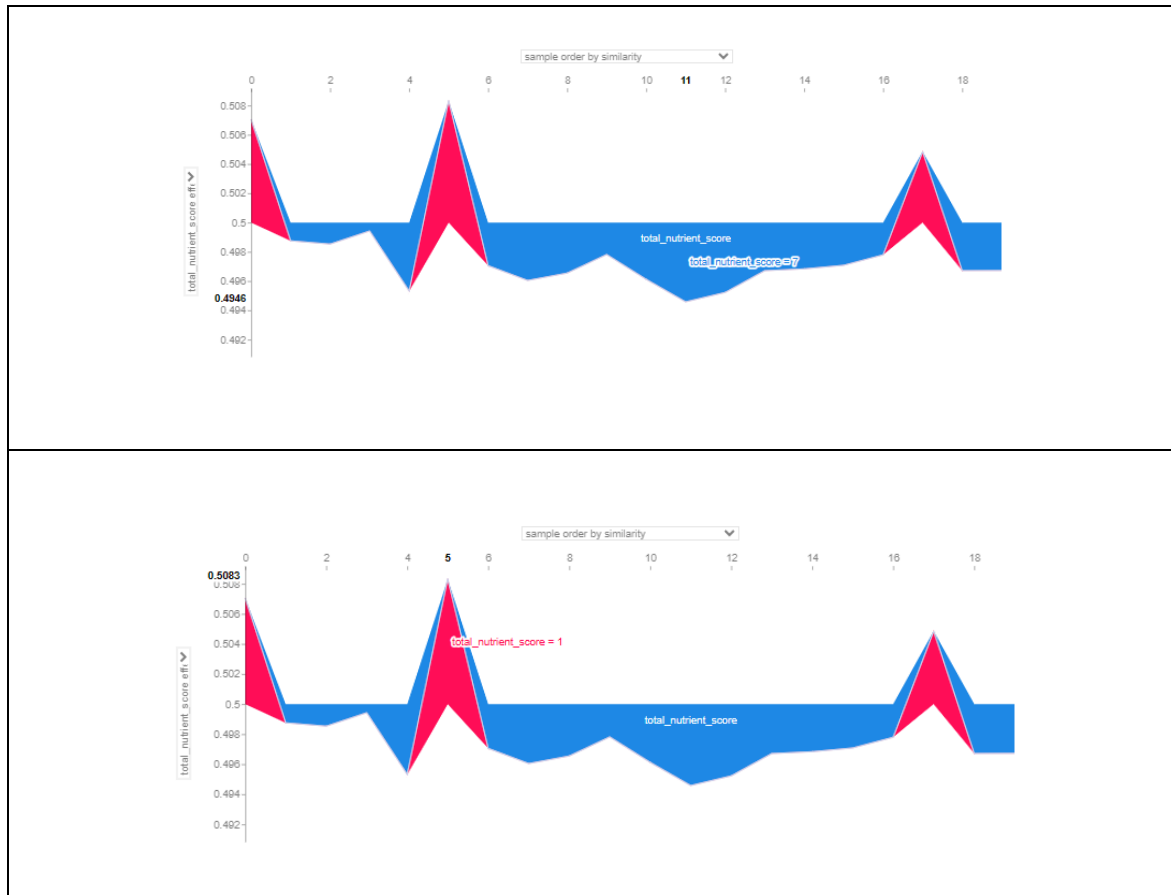


Figura 22 - Contribución variable *total_nutrient_score* variable objetivo

Al analizar los datos en el conjunto completo, se percibió la posibilidad de perder información debido a la gran cantidad de variables. Por lo tanto, se llevó a cabo el mismo estudio para distintos subgrupos del conjunto de datos, como demográficos, cuestionario, y combinaciones de varios, como demográfico-dieta, demográfico-cuestionario, y laboratorio-examen.

En líneas generales, el rendimiento de los modelos ha sido similar, sin lograr un rendimiento alto para la variable "depresión" en ninguno de los casos. En todos los modelos, se ha seguido la misma estrategia utilizada en el *Dataframe* total, como se ha explicado en esta sección. Para el árbol de decisión y el EBM, fue necesario emplear la técnica de *oversampling* para equilibrar los datos, mientras que para el *Random Forest* se realizó una búsqueda óptima de hiper parámetros mediante *GridSearchCV*.

En la realización de la comparativa, se llevó a cabo la ejecución de una gráfica de SHAP utilizando todas las instancias del conjunto de prueba con el objetivo de obtener el listado de las variables más representativas a nivel global. A continuación, se presentan algunos de los datos más representativos para cada subconjunto:

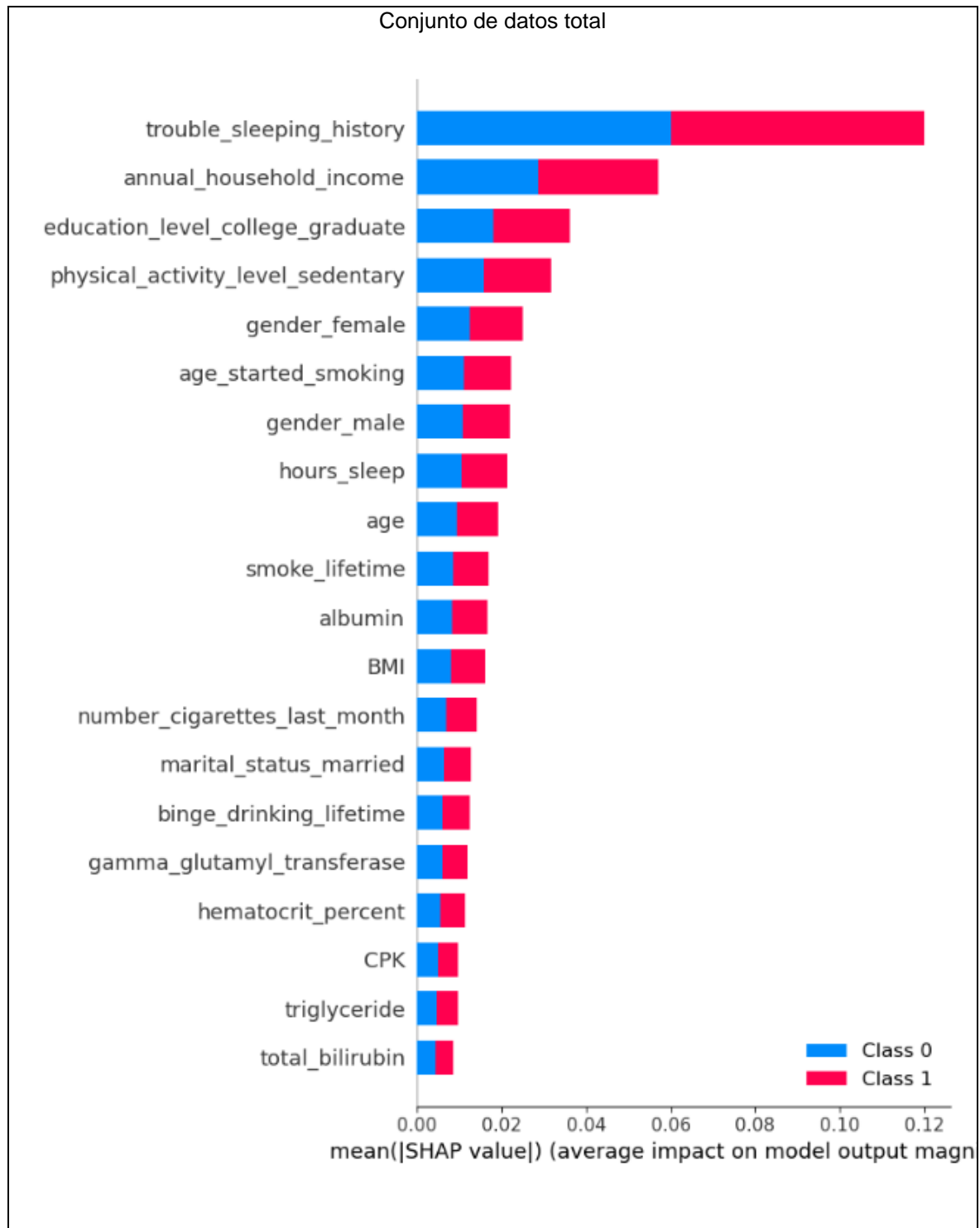


Figura 23 – SHAP. Gráfica sumario conjunto total

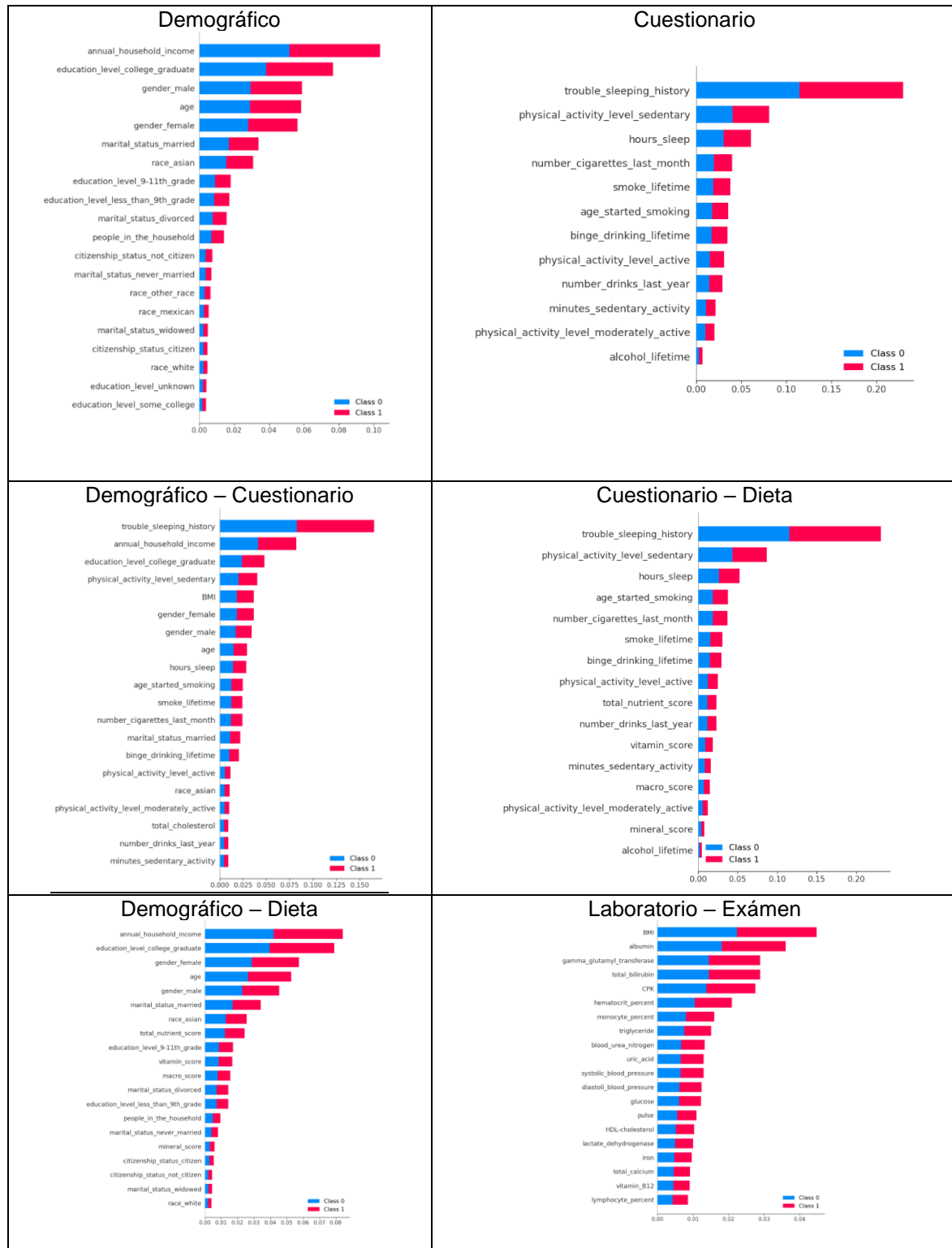


Figura 24 - SHAP. Gráfica sumario subconjuntos

Al haber analizado de forma separada distintos subgrupos y combinados del conjunto total, se ha examinado detalladamente instancias específicas, en este caso, las instancias 2 y 108, para comprender qué factores influyen en cada subgrupo. A continuación, se presenta un cuadro (Figura 25) que resume los datos más relevantes para predecir la presencia de depresión o no en cada subconjunto. La información que contiene la imagen se ha obtenido a partir del modelo de *Random Forest* con explicaciones de SHAP.

Variables más relevantes Random Forest (SHAP)	Instancia 2 (Sin Depresión)	Instancia 108 (Depresión)
Datos de Cuestionario		
Historial de problemas de sueño	No	Sí
Sedentarismo	No	Sí
Fumador	No	Sí
Horas de sueño	-	5
Actividad física	Sí	-
Datos de Cuestionario y Dieta		
Historial de problemas de sueño	No	Sí
Sedentarismo	-	Sí
Actividad física	Sí	-
Fumador	No	-
Cigarrillos último mes	-	300
Horas de sueño	-	8
Puntuación Nutrición	Alta	Baja
Datos Demográficos		
Ingresos anuales	Altos (+100.000\$)	Bajos (5.000-9.000\$)
Estudios	Superiores	Sin estudios
Estado civil	Casado	-
Edad	45	63
Género	-	Mujer
Datos Laboratorio y examen		
Nivel Gamma-glutamyl transferasa (GGT)	8	23.24
IMC	22.3	46.4
Presión arterial	108/60	-
Triglicéridos	-	182
Bilirrubina	-	0.4
Vitamina B12	349	-
Datos Demográficos y Dieta		
Ingresos anuales	>100.000\$	5.000-9.000\$
Estudios	Superiores	Sin estudios
Estado civil	Casado	-
Edad	45	63
Puntuación Nutrición	-	2 (baja)
Datos Demográficos y Cuestionario		
Estudios	Superiores	-
IMC	22.3	46.4
Fumador	No	-
Sedentarismo	-	Sí
Horas de sueño	-	5

Figura 25 - Análisis de Factores Relevantes en Subgrupos Específicos

En el análisis detallado de subgrupos específicos, se ha identificado información clave que podría haberse pasado por alto en el conjunto total. Detalles como el consumo de tabaco, presión arterial, niveles de triglicéridos, edad, género y horas de sueño se vuelven más evidentes y proporcionan una comprensión más profunda de los factores que influyen en la depresión o su ausencia. Este enfoque detallado en subgrupos nos permite capturar matices importantes y obtener una visión más completa de las relaciones entre variables, enriqueciendo así el entendimiento de los determinantes de la depresión.

4 Discussión

El análisis de técnicas de inteligencia artificial ha sido crucial para identificar factores clave vinculados a la depresión. La implementación de modelos como *Decision Tree* y *Explainable Boosting Models* proporcionó una visión general, siendo EBM más preciso y comprensible gracias al uso de la librería Interpret ML. Sin embargo, al aplicar técnicas de XAI sobre un modelo *Random Forest*, como SHAP y LIME, se lograron explicaciones más claras.

SHAP demostró ser la herramienta más práctica al permitir explorar no solo a nivel local, sino también global, seleccionando diversas instancias. Su utilidad se extendió a la exploración del comportamiento de variables en relación con el objetivo. Este análisis, integrando resultados de diferentes modelos y técnicas XAI, ha revelado hallazgos interesantes para comprender los factores determinantes de la depresión.

El análisis de los datos revela que, entre los diversos factores de estilo de vida considerados, los trastornos del sueño surgen como uno de los indicadores más significativos en relación con la depresión. Este hallazgo respalda la creciente evidencia en la literatura científica que subraya la importancia crítica del sueño en la salud mental [\[21\]](#). La conexión entre la calidad del sueño y la salud mental ha sido objeto de numerosos estudios, y los resultados obtenidos en el presente proyecto refuerzan la noción de que abordar los problemas del sueño puede ser fundamental en la prevención y gestión de la depresión.

El análisis de variables demográficas ha destacado patrones de prevalencia de la depresión en ciertos grupos, como mujeres y aquellos con ingresos más bajos o sin educación formal. Estos resultados están en consonancia con la literatura existente sobre desigualdades socioeconómicas y de género en la salud mental [\[22\]](#).

La exploración de biomarcadores en sangre, como la albúmina y la *gamma-glutamyl transferasa* (GGT), ha añadido una dimensión biomédica al estudio de la depresión. Estos hallazgos concuerdan con investigaciones previas que sugieren la asociación entre niveles más altos de albúmina y una menor prevalencia de depresión [\[23\]](#). La relación entre niveles más bajos de albúmina y mayores niveles de GGT en personas con depresión también ha sido documentada [\[24\]](#). Estos resultados sugieren la importancia de explorar más a fondo las interconexiones entre la salud mental y los marcadores bioquímicos, abriendo vías para futuras investigaciones.

En relación con el estilo de vida, la actividad física, el alcohol y el tabaco aparecen como indicadores significativos en la prevalencia de la depresión, respaldando la importancia de considerar factores de estilo de vida en la evaluación y prevención de la depresión [\[25\]](#).

Aunque la nutrición se sintetizó en una puntuación para este estudio, la base de datos NHANES ofrece detalles específicos sobre la ingesta de alimentos. Un análisis más profundo de estos datos podría revelar la contribución específica de cada alimento a la depresión, siendo un área de investigación valiosa. No obstante, se ha evidenciado en el estudio que a mayor puntuación de consumo de vitaminas y minerales menor prevalencia en depresión.

Este hallazgo sugiere la importancia de considerar no solo la cantidad total de alimentos ingeridos, sino también la calidad de la dieta en relación con la salud mental. La relación entre la nutrición y la depresión ha sido objeto de interés en la investigación científica, y la observación en el presente estudio respalda la noción de que una dieta equilibrada y rica en nutrientes puede desempeñar un papel protector contra la depresión [\[26\]](#). Sin embargo, es necesario reconocer que la síntesis de la nutrición en una puntuación en este estudio es una simplificación, y se requiere una investigación más detallada para comprender completamente cómo diferentes alimentos y nutrientes pueden influir en la salud mental.

En conclusión, la aplicación de modelos de inteligencia artificial y técnicas XAI ha proporcionado una comprensión más profunda de la complejidad de los factores que influyen en la depresión, brindando conocimientos fundamentales para futuras investigaciones y estrategias de intervención.

5 Conclusiones

Este trabajo ha alcanzado con éxito los objetivos iniciales de contribuir al conocimiento e innovación en la comprensión de la depresión y su impacto en la sociedad. Sin embargo, durante su desarrollo, se identificó un aspecto crítico relacionado con la desigualdad en el conjunto de datos, específicamente en la variable de depresión. Con un total de 10.175 entradas en el conjunto inicial, apenas 5372 contenían respuestas relacionadas con el test para clasificar la presencia de depresión. Esta disparidad en la distribución de datos planteó un desafío durante el proceso de modelado.

La desproporción en la cantidad de instancias clasificadas en depresión introduce un sesgo significativo en los modelos generados, afectando la capacidad de generalización y precisión. La escasez de casos positivos de depresión podría limitar la capacidad de los modelos para capturar la complejidad de los factores que contribuyen a esta condición.

Adicionalmente, se plantea la posibilidad de reticencia por parte de los individuos a compartir información sobre su salud mental. La estigmatización asociada con la depresión puede influir en la sinceridad de las respuestas, subrayando la importancia de garantizar la confidencialidad y el anonimato en la recopilación de datos. En términos de estrategias futuras, se sugiere una colaboración más estrecha con profesionales de la salud mental para mejorar la validez de las mediciones y la interpretación de los resultados, considerando la posibilidad de reticencia como un factor a abordar.

También, se ha observado que en los datos de laboratorio había numerosos valores fuera de rango. Este aspecto deberá tenerse en cuenta en futuras investigaciones para garantizar que los datos recopilados contengan información válida. La calidad de los datos de laboratorio es crucial para la integridad y la fiabilidad de los resultados, por lo que se sugiere una atención especial a la recolección de datos en investigaciones subsiguientes.

Una dirección prometedora para investigaciones futuras sería la realización de un estudio exclusivo centrado en la relación entre nutrición y depresión. La base de datos completa disponible incluye una detallada entrevista que abarca todos los alimentos individuales ingeridos durante dos días, así como la información sobre suplementos. Este enfoque permitiría un análisis más profundo, considerando aspectos como el lugar de consumo, la procedencia de los alimentos, las cantidades, entre otros.

Aunque en el presente estudio se sintetizó la información en una puntuación, se reconoce que esta simplificación no captura completamente la riqueza de los datos alimentarios. La distinción entre el porcentaje de hidratos obtenidos de un cereal integral versus uno no integral, por ejemplo, podría tener implicaciones significativas. A pesar de las limitaciones de este estudio para abordar esta complejidad, se ha observado que una mayor puntuación en nutrientes se relaciona con una menor prevalencia de depresión. Explorar específicamente esta

información detallada en futuras investigaciones podría proporcionar una comprensión más completa de la influencia de la nutrición en la salud mental.

Los resultados revelan que factores como los problemas de sueño, la actividad física y el nivel de estudios desempeñan un papel significativo tanto en el análisis del conjunto total como en el examen de subgrupos específicos. Estos hallazgos sugieren la necesidad de una investigación más profunda sobre estas variables para entender mejor su relación con la depresión. Aunque la dirección causal de estas variables en relación con la depresión no está clara, su importancia destaca la relevancia de fomentar patrones de sueño saludables, la práctica regular de actividad física y la búsqueda de estudios motivadores. Este énfasis en promover hábitos de vida positivos podría tener implicaciones significativas en la prevención y gestión de la depresión.

En cuanto al proceso de seguimiento de la planificación, este ha sido en su mayoría consistente, aunque surgieron ajustes notables durante el desarrollo del trabajo. Inicialmente, la planificación presentó deficiencias que llevaron a una reestructuración en la segunda entrega. A pesar de estos ajustes iniciales, la ejecución del proyecto avanzó siguiendo la planificación revisada. No obstante, se encontraron ocasiones en las que fue necesario abordar errores en entregas anteriores para garantizar la coherencia y continuidad del trabajo planificado. Estos contratiempos, aunque demandaron esfuerzos adicionales, no comprometieron significativamente la ejecución global del proyecto, y se logró mantener la trayectoria prevista hacia la consecución de los objetivos.

En resumen, este proyecto ha sido un significativo punto de partida hacia la comprensión de la depresión y su impacto en la sociedad. Los desafíos identificados, como la posible reticencia de los participantes a contestar preguntas sobre salud mental, subrayan la complejidad inherente a estos temas. La sugerencia de una mayor prevalencia de depresión en casos de nutrición deficiente plantea la posibilidad de investigaciones más exhaustivas. Explorar minuciosamente la relación entre nutrición y depresión mediante la totalidad de la información disponible en NHANES, y examinar cómo los biomarcadores en sangre pueden detectar o estar vinculados con la depresión, abre nuevas puertas para obtener una comprensión más profunda. Este proyecto, lejos de ser un punto final, se presenta como punto de partida que insta a futuras investigaciones a desentrañar los misterios que rodean la conexión entre estos factores y la depresión.

6 Glosario

OMS: Organización Mundial de la Salud

EM: Esclerosis Múltiple

NHANES: *National Health and Nutrition Examination Survey*

TFG: Trabajo Final de Grado

IA: Inteligencia Artificial

XAI: *Explainable Artificial Intelligence*

ML: *Machine Learning*

SVM: *Support Vector Machine*

LR: *Logistic Regression*

RF: *Random Forest*

DT: *Decision Tree*

EBM: *Explainable Boosting Classifier*

SHAP: *Shapley Additive Explanations*

LIME: *Local Interpretable Model-agnostic Explanations*

GDPR: Reglamento General de Protección de Datos

BMI: *Body mass index*

ALT: alanina aminotransferasa

GGT: *Gamma Glutamyl Transferasa*

NaN: *Not a number*

7 Bibliografía

- [1] World Health Organization: WHO & World Health Organization: WHO. (2023, 31 marzo). Depresión. <https://www.who.int/es/news-room/fact-sheets/detail/depression>
- [2] Gaem. (2020, 3 agosto). La depresión es hasta tres veces más común en personas con EM. GAEM (Grupo de Afectados de Esclerosis Múltiple). <https://fundaciongaem.org/depresion-comun-esclerosis-multiple/#:~:text=Concretamente%2C%20las%20personas%20con%20esclerosis,del%20diagn%C3%B3stico%20de%20la%20enfermedad.>
- [3] NHANES - National Health and Nutrition Examination Survey homepage. (s. f.). <https://www.cdc.gov/nchs/nhanes/index.htm>
- [4] National Health and Nutrition Examination Survey. (2017, 26 enero). Kaggle. <https://www.kaggle.com/datasets/cdc/national-health-and-nutrition-examination-survey/data>
- [5] Depression. (s. f.). National Institute of Mental Health (NIMH). <https://www.nimh.nih.gov/health/topics/depression#:~:text=Research%20suggests%20that%20genetic%2C%20biological,it%20often%20begins%20in%20adulthood.>
- [6] Sharpe, D. K., Collins-McNeil, J., Jones, J. W., & Bailey, R. K. (2016). Understanding the correlation between physical activity and clinical depression in Women: A Review of the literature. Journal of depression & anxiety, 5(4). <https://doi.org/10.4172/2167-1044.1000249>
- [7] Depression and sleep: Understanding the connection. (2021, 21 octubre). Johns Hopkins Medicine. <https://www.hopkinsmedicine.org/health/wellness-and-prevention/depression-and-sleep-understanding-the-connection>
- [8] Pti. (2023, 30 enero). Chronic stress has a correlation with depression: study. The Economic Times. <https://economictimes.indiatimes.com/magazines/panache/chronic-stress-has-a-correlation-with-depression-study/articleshow/97452481.cms>
- [9] Healthy lifestyle can help prevent depression – and new research may. (2023, 11 septiembre). University of Cambridge. <https://www.cam.ac.uk/research/news/healthy-lifestyle-can-help-prevent-depression-and-new-research-may-explain-why>
- [10] Selhub, E., MD. (2022, 18 septiembre). Nutritional Psychiatry: Your brain on food. Harvard Health. <https://www.health.harvard.edu/blog/nutritional-psychiatry-your-brain-on-food-201511168626>
- [11] Klados, G. A., Politof, K., Bei, E. S., Moirgiorgou, K., Anousakis-Vlachochristou, N., Μαρσόπουλος, Γ. K., & Zervakis, M. (2021). Machine Learning model for Predicting CVD risk on NHANES data. 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). <https://doi.org/10.1109/embc46164.2021.9630119>
- [12] Ellis, G. M., & Souza, P. E. (2021). Using machine learning and the National Health and Nutrition Examination Survey to classify individuals with hearing loss. Frontiers in digital health, 3. <https://doi.org/10.3389/fdgth.2021.723533>
- [13] Vangeepuram, N., Liu, B., Chiu, P., Wang, L., & Pandey, G. (2021). Predicting youth diabetes risk using NHANES data and machine learning. Scientific Reports, 11(1). <https://doi.org/10.1038/s41598-021-90406-0>

- [14] López-Martínez, F., Núñez-Valdéz, E. R., Crespo, R. G., & García-Díaz, V. (2020b). An artificial neural network approach for predicting hypertension using NHANES data. *Scientific Reports*, 10(1). <https://doi.org/10.1038/s41598-020-67640-z>
- [15] Qu, Z., Wang, Y., Guo, D., He, G., Sui, C., Duan, Y., Zhang, X., Lan, L., Meng, H., Wang, Y., & Liu, X. (2023). Identifying depression in the United States Veterans using deep learning Algorithms, NHANES 2005–2018. *BMC Psychiatry*, 23(1). <https://doi.org/10.1186/s12888-023-05109-9>
- [16] Chapter 3 – Rights of the data subject - General Data Protection Regulation (GDPR). (2018, 5 octubre). General Data Protection Regulation (GDPR). <https://gdpr-info.eu/chapter-3/>
- [17] NHANES questionnaires, datasets, and related documentation. (s. f.). <https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2013>
- [18] colaboradores de Wikipedia. (2023, 9 octubre). PHQ-9. Wikipedia, la enciclopedia libre. <https://es.wikipedia.org/wiki/PHQ-9>
- [19] Rampersaud, G. (2012, 1 enero). Calculation and comparison of nutrient density/quality scores for commonly consumed fresh fruit/Estimativa e comparacao de varios indices utilizados para calculo de densidade nutricional/qualidade de frutas. Document - Gale OneFile: Health and Medicine. <https://go.gale.com/ps/i.do?p=HRCA&u=googlescholar&id=GALE|A339848715&v=2.1&it=r&sid=HRCA&asid=49ffbase>
- [20] : U.S. Department of Health and Human Services and U.S. Department of Agriculture. (2015). 2015–2020 Dietary Guidelines for Americans. 8th Edition. HHS Publication #: HHS-ODPHP-2015-2020-01-DGA-A. USDA Publication #: Home and Garden Bulletin No. 232. https://health.gov/sites/default/files/2019-09/2015-2020_Dietary_Guidelines.pdf
- [21] Plante, D. T. (2021). The evolving nexus of sleep and depression. *American Journal of Psychiatry*, 178(10), 896-902. <https://doi.org/10.1176/appi.ajp.2021.21080821>
- [22] Jo SJ, Yim HW, Bang MH, Lee MO, Jun TY, Choi JS, Lee MS, Lee WC, Park YM (2011). The Association between Economic Status and Depressive Symptoms: An Individual and Community Level Approach. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3182383/#:~:text=The%20studies%20on%20the%20association,a%20lower%20prevalence%20of%20depression.&text=Such%20relationship%20can%20be%20explained%20by%20the%20stress%20theory>.
- [23] Al-Marwani, S., Batieha, A., Khader, Y., El-Khateeb, M., Jaddou, H., & Ajlouni, K. (2023). Association between albumin and depression: a population-based study. *BMC Psychiatry*, 23(1). <https://doi.org/10.1186/s12888-023-05174-0>
- [24] Li X, Mao Y, Zhu S, Ma J, Gao S, Jin X, Wei Z, Geng Y. (2023). Relationship between depressive disorders and biochemical indicators in adult men and women. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9847124/#:~:text=As%20compared%20to%20the%20control,odum%20than%20the%20control%20group>.
- [25] Sarris, J., Thomson, R., Hargraves, F., Eaton, M., De Manincor, M. J., Veronese, N., Solmi, M., Stubbs, B., Yung, A. R., & Firth, J. (2020). Multiple lifestyle factors and depressed mood: A cross-sectional and longitudinal analysis of the UK Biobank (N = 84,860). *BMC Medicine*, 18(1). <https://doi.org/10.1186/s12916-020-01813-5>
- [26] Ekinici, G. N., & Şanlıer, N. (2023). The Relationship between nutrition and Depression in the life Process: a mini-review. *Experimental Gerontology*, 172, 112072. <https://doi.org/10.1016/j.exger.2022.112072>

8 Anexos

Se adjunta el código fuente utilizado en esta investigación. Para acceder y revisar el código, visite el repositorio de GitHub en el siguiente enlace: [tperez85/NHANES_depression_XAI \(github.com\)](https://github.com/tperez85/NHANES_depression_XAI)