

# ¿QUÉ HAY DETRÁS DEL DESEMPEÑO?

Análisis del acceso a la salud mental como motor de  
rendimiento en las empresas

*Modelo predictivo de acceso a tratamiento de salud mental en contextos  
laborales*

## **Autores:**

Ana María Ramírez Acevedo  
Johanna Gonzalez Franco  
Santiago Quintero Ramírez

## **Ejecutores**

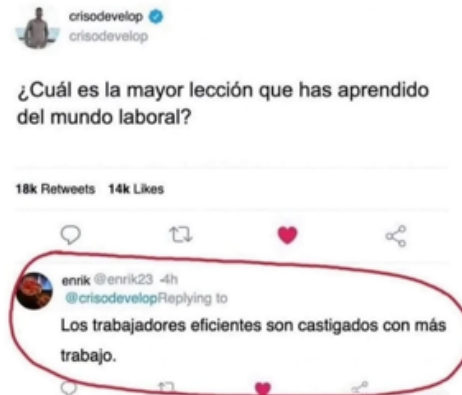
Natalia Betancur Herrera  
Frank Yesid Zapata Castaño  
Margarita Maria Orozco  
Andres Sánchez

Universidad de Antioquia, Universidad de Caldas  
Talento TECH  
BOOTCAMP Inteligencia Artificial

**Mayo 2025**

## 1. Introducción

*“El trabajo es la condición básica y fundamental de toda la vida humana. Y lo es en tal grado que, hasta cierto punto, debemos decir que el trabajo ha creado al propio hombre.”*<sup>1</sup> (Engels, 1876) Siguiendo a Engels, si el trabajo nos define, ¿qué sucede con los demás aspectos que nos acompañan en el camino de la existencia? ¿En qué momento cobra relevancia nuestra salud mental? ¿Al ser el trabajo el eje fundamental, están dispuestas las empresas a entender a sus trabajadores desde una visión integral?



Lo que pareciera ser un meme<sup>2</sup> cualquiera o una historia mal contada, termina mostrando lo que muchos hemos pensado en algún momento cuando nos movemos en las dinámicas del entorno laboral.

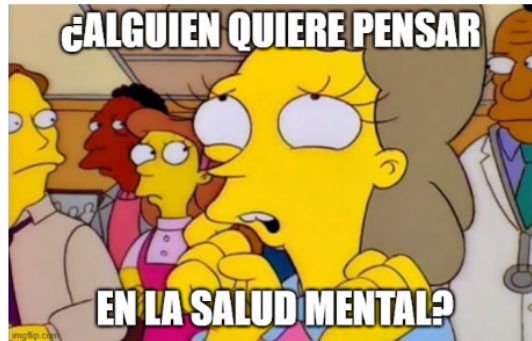
Trabajar resulta siendo así, una necesidad básica que ha formado parte de la historia del ser humano hasta convertirse en un elemento constitutivo de su qué hacer en el mundo y, es bajo esta dinámica que termina siendo aceptado como algo natural e inevitable. Teniendo en cuenta esto, genera inquietud si al ser humano realmente le gusta trabajar o no, si le gusta estar desempleado o no, ¿estará siempre atrapado en un vaivén entre trabajar y no trabajar, o existen formas alternativas para vivir?

Partiendo de esta reflexión y teniendo en cuenta que, si bien el trabajo es un factor importante en la vida del ser humano, la salud mental en los entornos laborales ha emergido como un componente crítico del rendimiento organizacional, especialmente en contextos donde las exigencias laborales, los cambios en la modalidad de trabajo y los desafíos sociales confluyen para impactar la productividad de las personas. En este contexto, el presente informe de investigación se centra en el desarrollo de un modelo predictivo de acceso a tratamiento de salud mental en contextos laborales, con el objetivo de identificar los factores que influyen en la interferencia de los problemas de salud mental

<sup>1</sup> Engels, F. (1876). *El papel del trabajo en la transformación del mono en hombre*. En *Dialéctica de la naturaleza*. Recuperado de <https://www.marxists.org/espanol/m-e/1870s/1876trab.htm>

<sup>2</sup> “Los memes de internet son elementos digitales que se comparten, imitan y transforman en línea con un reconocimiento mutuo.” (Shifman, 2014, p. 41)

sobre el desempeño laboral, cómo frente al conocimiento y acceso a tratamientos adecuados, puede convertirse en un impulso para mejorar no solo el desempeño, sino también el bienestar en los entornos productivos.



Hemos construido un modelo apoyado en técnicas de Machine Learning (ML) que está orientado en predecir la afectación que tienen los colaboradores<sup>3</sup> de una empresa en su rendimiento laboral por temas de salud mental. Este modelo tiene en cuenta distintas variables sociodemográficas que pueden ser aplicadas en distintos entornos que busquen obtener una visión integral de los desafíos que se enfrentan al hablar de salud mental del ser humano, el acceso a ella y su impacto.

## 2. Justificación:

En un mundo cada vez más conectado y exigente, la salud mental se ha convertido en un reto global que impacta directamente la productividad y el bienestar de los equipos de trabajo. Este proyecto busca ir más allá de los números: nos propone contar la historia humana detrás de los datos, revelando patrones ocultos y necesidades reales en el acceso a servicios de salud mental dentro de las empresas. Nuestra justificación radica en que, al comprender y atender estas historias, podemos generar recomendaciones prácticas que ayuden a las organizaciones a fortalecer el bienestar de sus colaboradores, optimizar su rendimiento y construir ambientes laborales más sanos y productivos.

## 3. Objetivo general:

- Desarrollar un modelo predictivo utilizando técnicas de aprendizaje automático que determine los factores socio demográficos, ocupacionales y organizacionales que afectan el rendimiento laboral debido a problemas de salud mental, en un esfuerzo por mejorar el acceso a la atención, el tratamiento y el bienestar en el trabajo.

## Objetivos específicos:

<sup>3</sup> En entornos laborales, "colaborador" se refiere a una persona que contribuye activamente al logro de los objetivos organizacionales, destacando su rol participativo y compromiso, a diferencia del término "trabajador", que se centra en la relación contractual o laboral. Chiavenato, I. (2009). *Gestión del talento humano* (3.ª ed.). McGraw-Hill.

**3.1** Examinar los factores sociodemográficos, las percepciones organizacionales y sus características ocupacionales asociadas en relación con la accesibilidad de los servicios de salud mental en el lugar de trabajo.

**3.2** Diseñar e implementar una solución de aprendizaje automático para estimar la probabilidad del deterioro del rendimiento laboral debido a un tratamiento psicológico insuficiente o inadecuado de una enfermedad. Entrenar el modelo para asegurar una mayor precisión en la predicción.

**3.3** Evaluar los resultados del modelo predictivo y formular políticas organizacionales que inviten a la participación activa hacia la optimización de la salud psicológica de los empleados, así como mejorar la productividad a través de intervenciones reflexivas.

#### **4. Alcance del proyecto:**

El proyecto se basa en el análisis y modelado predictivo utilizando la base de datos “Mental Health in Tech Survey” disponible en Kaggle<sup>4</sup>. Esta base recoge, a nivel internacional, información detallada sobre factores sociodemográficos, percepciones en el trabajo, modalidades laborales y acceso a tratamientos de salud mental en el sector tecnológico. Gracias a esta riqueza de variables, podemos explorar cómo el bienestar emocional de las personas se relaciona con su desempeño laboral, permitiendo desarrollar un modelo de Machine Learning (AdaBoost Classifier) que ayude a predecir cuándo las condiciones psicológicas podrían afectar el rendimiento en el trabajo.

El alcance del proyecto busca que el modelo sea útil no solo en el sector tecnológico, sino también en otros entornos laborales con perfiles y condiciones organizacionales similares. Sin embargo, es importante reconocer que los datos provienen principalmente de empleados del sector tecnológico y de países desarrollados, lo que significa que, para extender los resultados a otros sectores o regiones, sería necesario realizar ajustes y validaciones adicionales que garanticen la pertinencia y representatividad de las recomendaciones.

Para asegurar que el modelo predictivo sea útil y relevante en Colombia, se han considerado datos secundarios y estudios locales sobre salud mental y productividad, así como las particularidades del mercado laboral nacional. Aunque la base de datos original no incluye información específica de Colombia, el proyecto utiliza variables universales como edad, género, modalidad de trabajo y acceso a tratamientos que resultan relevantes para la población trabajadora del país.

---

<sup>4</sup> Open Sourcing Mental Illness. (2016). *Mental health in tech survey* [Dataset]. Kaggle.  
<https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey>

El alcance en Colombia contempla:

- **Estimación del impacto financiero:**

Se analizan cifras locales sobre costos asociados al ausentismo, la rotación y la baja productividad relacionadas con trastornos mentales, para dimensionar el valor de invertir en bienestar emocional dentro de las empresas colombianas.

- **Adaptación del modelo:**

El modelo se ajusta para identificar los grupos poblacionales y las modalidades de trabajo más comunes en Colombia (presencial, remoto e híbrido), facilitando el diseño de intervenciones específicas y efectivas.

- **Recomendaciones prácticas:**

Se generan recomendaciones para las empresas colombianas, comparando la inversión en salud mental con los costos de reemplazo de personal, utilizando tanto los patrones predichos por el modelo como datos financieros nacionales.

Sin embargo, se reconoce una limitación importante: la falta de datos locales directamente disponibles para entrenar el modelo. Por ello, se establece la necesidad de realizar validaciones y ajustes futuros con información colombiana, con el fin de mejorar la precisión y la contextualización del modelo, para este caso, se realizó una encuesta en el territorio para tener un primer acercamiento a la información en el contexto actual con el fin de orientar las métricas del modelo y búsqueda de criterios de éxito.

Este alcance representa un avance significativo para sensibilizar y gestionar la salud mental en el ámbito laboral colombiano, abriendo la puerta a una mejora continua y a la adaptación del modelo a las realidades y necesidades específicas del país

## **2. Metodología: CRISP-DM**

### **2.1. Comprensión del Negocio:**

En el entorno empresarial actual, la salud mental de los colaboradores es un factor clave que influye directamente en su rendimiento y, por lo tanto, en los resultados de la empresa. Problemas como el ausentismo, la rotación de personal y la baja productividad, relacionados con el bienestar emocional, pueden generar pérdidas importantes tanto para la organización como para los propios colaboradores<sup>5</sup>

Este proyecto busca responder preguntas relevantes para la empresa: ¿cuánto le cuesta a una organización perder a un colaborador por motivos asociados a la salud mental? ¿Es más rentable invertir en la salud mental de los empleados que en procesos de

---

<sup>5</sup> El Economista. (s. f.). ¿Cuánto le cuesta a tu empresa el ausentismo?  
PRAX. (s. f.). ¿Cuánto le cuesta a las empresas el trabajo no saludable?

contratación y formación? ¿Cómo puede un modelo predictivo ayudar a anticipar estos riesgos y tomar decisiones más estratégicas?<sup>6</sup>

El modelo desarrollado permite identificar patrones que predicen afectaciones en el rendimiento laboral, facilitando la implementación de acciones preventivas. Esto no solo mejora el bienestar de los colaboradores, sino que también representa un ahorro económico al evitar los costos asociados al reemplazo de talento y la disminución de la productividad.

Por eso, poner la salud mental en el centro de la estrategia organizacional y apostar por la prevención no solo es una decisión inteligente y sostenible desde el punto de vista económico, sino que también construye una cultura más humana, productiva y capaz de adaptarse a los desafíos del entorno.

### 2.1.1 Objetivos del negocio:

Desarrollar e implementar un modelo predictivo apoyado en técnicas de Machine Learning que ayude a las organizaciones a entender y anticipar cómo el bienestar emocional de sus colaboradores influye en el día a día del trabajo.

Este modelo permitirá detectar señales tempranas de riesgo para la salud mental, facilitando la toma de decisiones estratégicas centradas en las personas. Así, las empresas podrán promover el bienestar, reducir el impacto del ausentismo y la rotación, y mejorar la eficiencia operativa, todo a través de acciones preventivas que cuidan tanto a los equipos como a los resultados del negocio

### 2.1.2 Criterios de éxito:

Para asegurar que el modelo y el proyecto realmente marquen una diferencia en la vida de las personas y en la organización, se han definido los siguientes criterios de éxito:

- **Impacto claro y relevante:**

El modelo debe mostrar, con datos reales y métricas de desempeño, cuáles son los factores de salud mental que más afectan la productividad y el bienestar de los colaboradores, ayudando a identificar dónde enfocar los esfuerzos.

Estos factores psicosociales y organizacionales generan un impacto considerable en la salud mental de los trabajadores, aumentando el riesgo de enfermedades mentales y físicas, afectando el rendimiento laboral y elevando el absentismo y la rotación en las empresas<sup>7</sup>

**Presión laboral y sobrecarga de trabajo:** La presión de tiempo y la sobrecarga de tareas aumentan significativamente el riesgo de trastornos mentales como

---

<sup>6</sup> Expansión. (2023). La calculadora que mide el costo de ignorar la salud mental de los trabajadores.

<sup>7</sup> Agencia Europea para la Seguridad y la Salud en el Trabajo (EU-OSHA). (s.f.). *Riesgos psicosociales y salud mental en el trabajo*. <https://osha.europa.eu/es/themes/psychosocial-risks-and-mental-health>

estrés, ansiedad y depresión, con una prevalencia ocho veces mayor entre quienes están expuestos a estas condiciones. El exceso de trabajo puede generar ansiedad, depresión, estrés crónico y agotamiento emocional (burnout), afectando la concentración, el sueño y la salud física. Factores que agravan este impacto son la falta de control sobre las tareas, ambientes laborales hostiles y expectativas poco realistas.<sup>8</sup>

**Falta de reconocimiento y discriminación:** La ausencia de reconocimiento y la discriminación en el trabajo afectan la motivación, el compromiso y el bienestar emocional de los empleados. Aproximadamente el 40% de los trabajadores a nivel mundial se sienten poco reconocidos, lo que impacta negativamente en su salud mental<sup>9</sup>

**Monotonía y trabajo repetitivo:** Las tareas monótonas y repetitivas pueden desencadenar trastornos músculo-esqueléticos y trastornos mentales como la depresión, debido a la falta de estímulo y la repetición constante de operaciones durante la jornada laboral.

- **Segmentación generacional:**

Es clave que el modelo permita reconocer si algunas generaciones, como Millennials, Generación X o Baby Boomers, están más expuestas a riesgos de salud mental, para así diseñar estrategias de apoyo adaptadas a sus necesidades y realidades.

---

<sup>8</sup> Martín Díaz, O., & Rosado Luna, E. (2023, 6 de septiembre). *¿Qué factores impactan en la salud mental en el contexto laboral?* Infocop.

<https://www.infocop.es/que-factores-impactan-en-la-salud-mental-en-el-contexto-laboral/>

<sup>9</sup> Infocop. (2023, 6 de septiembre). *¿Qué factores impactan en la salud mental en el contexto laboral?*

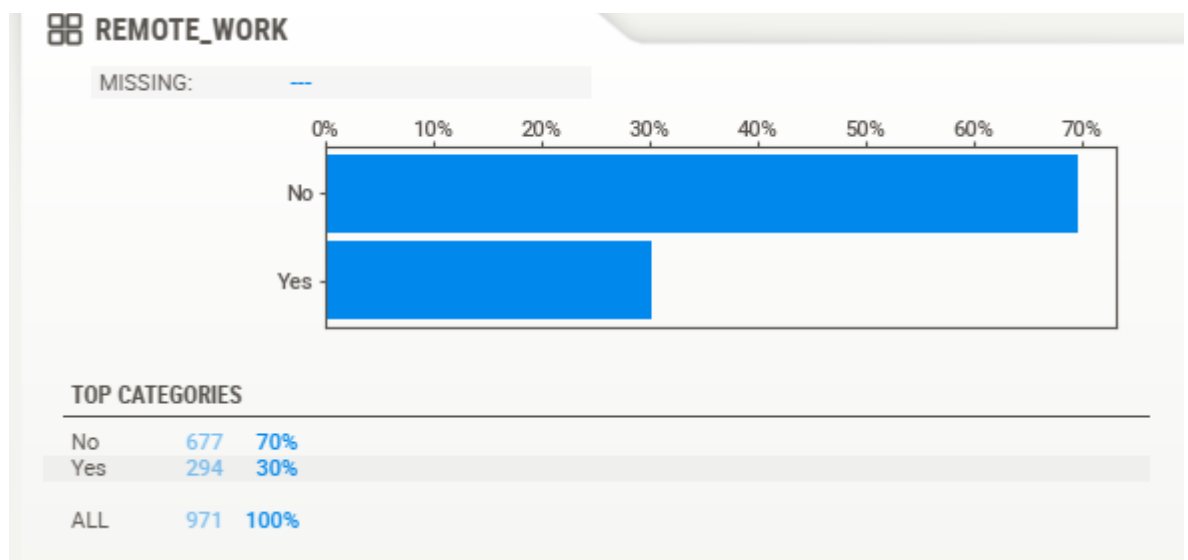
Infocop. <https://www.infocop.es/que-factores-impactan-en-la-salud-mental-en-el-contexto-laboral/>



Generación	Edad Aproximada	Factores de Riesgo	Datos Clave	Estrategias de Apoyo
<b>Generación Z</b>	13–28 años	Alta presión por rendimiento, baja tolerancia a entornos tóxicos, fatiga digital, poca experiencia laboral	60% reporta ansiedad laboral alta y mayor demanda de apoyo emocional (McKinsey, 2023)	Mentoría, espacios psicológicos seguros, flexibilidad y reconocimiento frecuente
<b>Millennials</b>	29–44 años	Burnout por exceso de roles, expectativas de propósito, transición generacional	76% afirma que el estrés afecta su desempeño; alta rotación por falta de bienestar (Deloitte)	Planes de carrera, bienestar holístico, autonomía y flexibilidad laboral
<b>Generación X</b>	45–60 años	Sobrecarga por liderazgo y familia, estrés tecnológico, normalización del agotamiento	Menos reporte, pero síntomas más crónicos; riesgo de invisibilidad emocional	Coaching individual, reconocimiento de experiencia, opciones de flexibilidad laboral
<b>Baby Boomers</b>	61–79 años	Pérdida de relevancia, aislamiento profesional, problemas físicos y emocionales	Baja búsqueda de ayuda, necesidad de integración y legado	Programas intergeneracionales, capacitaciones digitales, reconocimiento institucional

- **Adaptación a la modalidad de trabajo:**

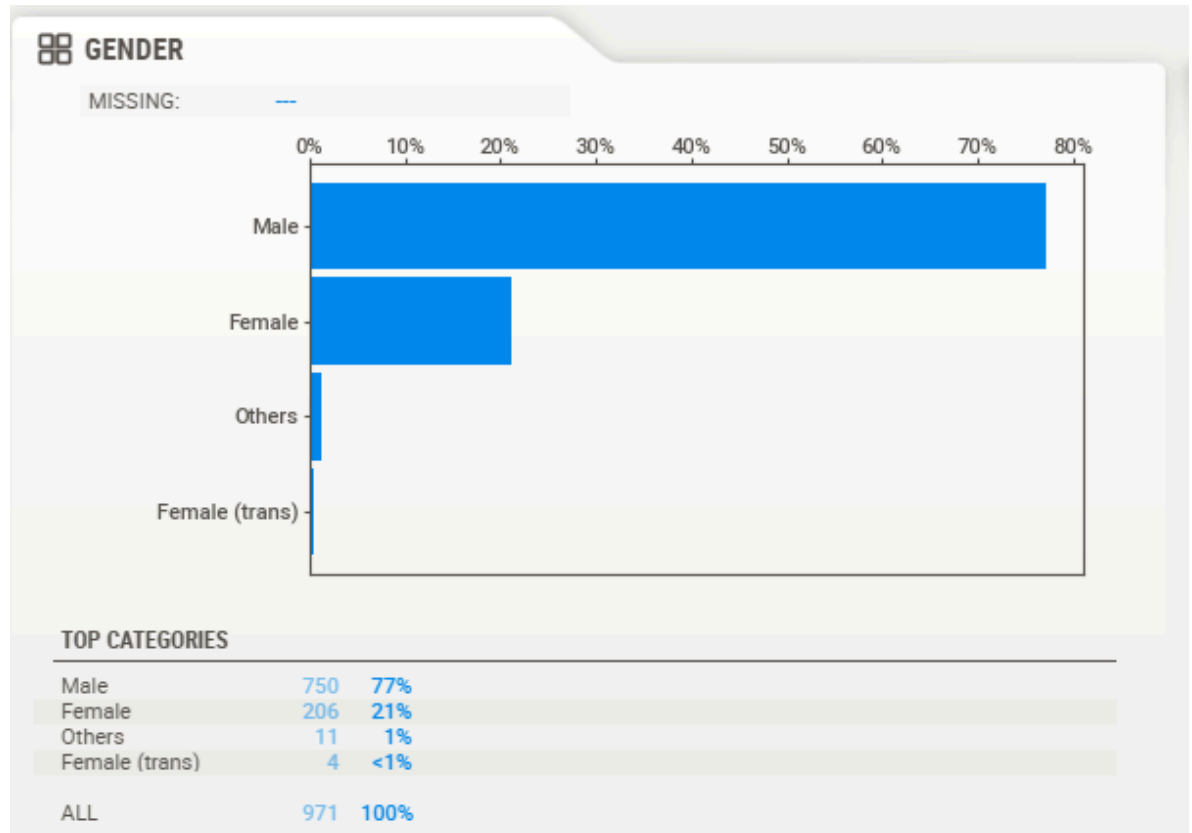
El modelo debe identificar en qué entornos laborales (presencial, remoto o híbrido) la salud mental tiene mayor impacto, facilitando intervenciones más precisas y efectivas.





- **Inclusión y equidad de género:**

El modelo debe evaluar el impacto diferenciado según género, asegurando que las políticas y acciones de bienestar sean inclusivas y respondan a las necesidades específicas de hombres, mujeres y otras identidades.



- **Precisión y confiabilidad:**

El modelo debe alcanzar un alto nivel de precisión (por ejemplo, superior al 80%) en la predicción de afectaciones al rendimiento, garantizando que las decisiones tomadas a partir de sus resultados sean sólidas y confiables.

Estos criterios buscan que el proyecto no solo cumpla con objetivos técnicos, sino que también genere un impacto positivo, humano y sostenible en la organización y en la vida de cada colaborador.

## 2.2. Comprensión de los Datos

### 2.2.1 Descripción de la(s) fuente(s) de datos:

El modelo predictivo se desarrolló utilizando la base de datos “Mental Health in Tech Survey”, recopilada y publicada por Open Sourcing Mental Illness (OSMI) en Kaggle. Este conjunto de datos busca visibilizar cómo es el estado de la salud mental en el entorno laboral, especialmente en la industria tecnológica, y se basa en encuestas voluntarias y anónimas realizadas a profesionales de todo el mundo, aunque la mayoría de los participantes provienen de países de habla inglesa como Estados Unidos y Europa.

La base cuenta con más múltiples registros que incluyen información clave sobre:

- **Datos sociodemográficos:** Edad, género, país de residencia y nivel educativo.
- **Características laborales:** Tipo de empresa, modalidad de trabajo (remoto, presencial), tamaño de la organización y flexibilidad laboral.
- **Condiciones de salud mental:** Diagnósticos previos, percepción de apoyo organizacional, acceso a tratamientos y conocimiento de políticas internas.
- **Percepción y estigma:** Actitudes hacia la salud mental en el trabajo, disposición a buscar ayuda y experiencias previas con condiciones psicológicas.

Aunque la base no contiene registros específicos de Colombia, sus variables universales permiten adaptar el modelo a nuestro contexto, identificando patrones que pueden ser muy útiles para las organizaciones colombianas interesadas en cuidar el bienestar emocional de sus empleados y reducir el impacto de los problemas de salud mental en la productividad. Así, el modelo no solo aporta información valiosa, sino que también abre la puerta a acciones más humanas y preventivas dentro de las empresas.

### 2.2.2 Estructura de los datos (variables, formatos, etc.)

La base de datos utilizada en este proyecto, Mental Health in Tech Survey (OSMI, 2016), está organizada en un formato tabular (.csv) donde cada fila representa a una persona encuestada y cada columna contiene información sobre su perfil, entorno laboral y bienestar emocional. Esto permite analizar cómo distintos factores influyen en la salud mental de los profesionales del sector tecnológico.

#### Principales características de la base de datos:

- **Número de registros y variables:**

Registros: 1.251 personas encuestadas.

Variables: 27 columnas con información diversa sobre cada individuo.

- **Tipos de variables:**

**Categorías nominales:** Género, país, estado, cómo afecta el trabajo a la salud mental, acceso a beneficios, opciones de cuidado, programas de bienestar, consecuencias de hablar sobre salud mental, entre otras.

**Categorías ordinales:** Edad (agrupada en rangos), modalidad de trabajo remoto, disposición a buscar ayuda, impacto de hablar sobre salud mental en entrevistas laborales.

**Booleanas (Sí/No):** Si ha recibido tratamiento, antecedentes familiares, consecuencias observadas, percepción de salud mental vs. física, entre otras.

**Numéricas:** La edad se utiliza como valor numérico antes de ser transformada.

#### **Variables clave del modelo:**

- **Variable objetivo (target):**  
treatment: Indica si la persona ha recibido tratamiento por salud mental (Sí = 1 / No = 0).
- **Variables predictoras:**  
Combinación de características sociodemográficas, laborales, organizacionales y percepciones sobre la salud mental.

#### **Formato y procesamiento de los datos:**

**Formato original:** .csv (valores separados por comas)

### **2.3. Preparación de los Datos**

#### **2.3.1 Limpieza de datos:**

Durante la limpieza de datos, nos enfocamos en estandarizar los nombres de las columnas, eliminar información incompleta o poco relevante y revisar la estructura general de la base para entender sus datos. Prestamos especial atención a variables como edad y género, depurando y uniformando sus valores para evitar errores. Finalmente, tras eliminar registros incompletos y corregir inconsistencias, realizamos una última revisión para asegurarnos de que la base estuviera lista y confiable para su uso. A continuación el detalle de las acciones realizadas:

- **Estandarización de nombres de columnas**

Se transformaron todos los nombres de las columnas a mayúsculas con el fin de mantener consistencia en la información y referenciar fácilmente la data.

- **Revisión de la estructura**

Se validó la cantidad de filas y columnas existentes así como el tipo de datos presentes en cada una con el fin de identificar inconsistencias, errores de tipeo, o formatos no estandarizados.

#	Column	Non-Null	Count	Dtype
0	AGE	1251	non-null	int64
1	GENDER	1251	non-null	object
2	COUNTRY	1251	non-null	object
3	SELF_EMPLOYED	1233	non-null	object
4	FAMILY_HISTORY	1251	non-null	object
5	TREATMENT	1251	non-null	object
6	WORK_INTERFERE	989	non-null	object
7	NO_EMPLOYEES	1251	non-null	object
8	REMOTE_WORK	1251	non-null	object
9	TECH_COMPANY	1251	non-null	object
10	BENEFITS	1251	non-null	object
11	CARE_OPTIONS	1251	non-null	object
12	WELLNESS_PROGRAM	1251	non-null	object
13	SEEK_HELP	1251	non-null	object
14	ANONYMITY	1251	non-null	object
15	LEAVE	1251	non-null	object
16	MENTAL_HEALTH_CONSEQUENCE	1251	non-null	object
17	PHYS_HEALTH_CONSEQUENCE	1251	non-null	object
18	COWORKERS	1251	non-null	object
19	SUPERVISOR	1251	non-null	object
20	MENTAL_HEALTH_INTERVIEW	1251	non-null	object
21	PHYS_HEALTH_INTERVIEW	1251	non-null	object
22	MENTAL_VS_PHYSICAL	1251	non-null	object
23	OBS_CONSEQUENCE	1251	non-null	object

- **Análisis de los valores únicos por columna**

Se revisaron los valores únicos existentes en cada columna para detectar valores atípicos que pudieran afectar la calidad de la data.

Column	Únicos	Faltantes
AGE	45	0
GENDER	46	0
COUNTRY	46	0
SELF_EMPLOYED	2	18
FAMILY_HISTORY	2	0
TREATMENT	2	0
WORK_INTERFERE	4	262
NO_EMPLOYEES	6	0
REMOTE_WORK	2	0
TECH_COMPANY	2	0
BENEFITS	3	0
CARE_OPTIONS	3	0
WELLNESS_PROGRAM	3	0
SEEK_HELP	3	0
ANONYMITY	3	0
LEAVE	5	0
MENTAL_HEALTH_CONSEQUENCE	3	0
PHYS_HEALTH_CONSEQUENCE	3	0
COWORKERS	3	0
SUPERVISOR	3	0
MENTAL_HEALTH_INTERVIEW	3	0
PHYS_HEALTH_INTERVIEW	3	0
MENTAL_VS_PHYSICAL	3	0
OBS_CONSEQUENCE	2	0

- **Depuración de datos poco relevantes para el análisis**

Se eliminan las columnas 'COMMENTS', 'STATE' y 'TIMESTAMP' ya que tenían una cantidad representativa de valores nulos, se realiza esta acción para garantizar que el análisis posterior se realice sobre datos completos y fiables. Ver Anexo Códigos

- **Verificación final de la base de datos**

Por último, se verificó la composición final de la base de datos, asegurándose de que los datos estuvieran limpios y estructurados correctamente antes de su almacenamiento en un archivo CSV.

#	Column	Cantidad	Únicos
0	AGE	971	44
1	GENDER	971	4
2	COUNTRY	971	38
3	SELF_EMPLOYED	971	2
4	FAMILY_HISTORY	971	2
5	TREATMENT	971	2
6	WORK_INTERFERE	971	4
7	NO_EMPLOYEES	971	6
8	REMOTE_WORK	971	2
9	TECH_COMPANY	971	2
10	BENEFITS	971	3
11	CARE_OPTIONS	971	3
12	WELLNESS_PROGRAM	971	3
13	SEEK_HELP	971	3
14	ANONYMITY	971	3
15	LEAVE	971	5
16	MENTAL_HEALTH_CONSEQUENCE	971	3
17	PHYS_HEALTH_CONSEQUENCE	971	3
18	COWORKERS	971	3
19	SUPERVISOR	971	3
20	MENTAL_HEALTH_INTERVIEW	971	3
21	PHYS_HEALTH_INTERVIEW	971	3
22	MENTAL_VS_PHYSICAL	971	3
23	OBS_CONSEQUENCE	971	2

### 2.3.2 Problemas detectados:

- **Género:**

Novedad en la normalización de datos de género, las respuestas eran reiterativas pero escritas de una forma distinta.

Género



- Male
- F
- Male
- Man
- Androgyne
- Male-ish
- Female
- Make
- Cis Male
- Female (trans)
- non-binary
- Female (cis)
- M
- Woman
- Female
- Femake
- femail
- Nah

- **Edad:**

Para la variable edad, se detectaron edades en negativo, valores imposibles o rangos de menores de 18 que no aplicarían para el análisis en el entorno laboral.

Age
-29
11
61
-1726
65
329
8
999999999999
-1
58

### 2.3.3 Transformaciones realizadas (normalización, codificación, etc.)

- **Edad:**

Fue necesario depurar la base ya que presentó las inconsistencias mencionadas en el apartado anterior y se usó IA para hallar un código de depuración.

- **Género:**

Se crea un diccionario de mapeo estandarizado ya que la bbdd contaba con múltiples respuestas en relación al género y venían en formatos distintos, con errores de digitación o usando sinónimos para el mismo género.

GENDER	Cantidad
Male	750
Female	206
Others	11
Female (trans)	4
Total	971

- **Selección de variables**

La selección de variables es una etapa esencial en la creación de un modelo predictivo, ya que de ella dependen la calidad de las predicciones, la eficiencia del análisis y la facilidad para comprender los resultados. En este proyecto, elegimos cuidadosamente aquellas variables que mejor representan aspectos sociodemográficos, laborales y organizacionales, y que además tienen una relación lógica o estadística con el objetivo principal: predecir la probabilidad de que una persona necesite tratamiento de salud mental en su entorno de trabajo. Este enfoque permite que el modelo sea más preciso, útil y fácil de interpretar para la toma de decisiones en las organizaciones.

- **Variable objetivo:**

**TREATMENT:** Variable binaria que indica si el encuestado ha recibido tratamiento por salud mental (Sí = 1, No = 0). Esta es la variable de salida o dependiente que el modelo busca predecir, permitiendo identificar quiénes han requerido atención profesional en salud mental y, en consecuencia, anticipar patrones de riesgo dentro de la organización. Esta variable es clave porque refleja el acceso real a tratamiento, un indicador directo de la necesidad de intervención y cuidado en salud mental en el entorno laboral.



Variable	Tipo	Justificación
AGE	Númerica	Refleja la etapa de vida, ayudando a entender cómo varían las necesidades y percepciones sobre salud mental entre generaciones.
GENDER	Catagórica	Permite identificar brechas de género y diferencias culturales en el reporte y acceso a la salud mental.
FAMILY_HISTORY	Binaria	Tener antecedentes familiares puede aumentar la conciencia o predisposición a buscar ayuda profesional.
WORK_INTERFERE	Catagórica	Indica si los problemas emocionales están afectando el desempeño laboral, mostrando el impacto real en la vida diaria del trabajador.
REMOTE_WORK	Binaria	La modalidad de trabajo (presencial o remoto) influye en el nivel de estrés y el acceso a apoyo organizacional.
TECH_COMPANY	Binaria	Permite diferenciar el entorno laboral, ya que las empresas tecnológicas pueden tener culturas y condiciones únicas.
BENEFITS	Binaria	Muestra si la empresa ofrece cobertura en salud mental, facilitando el acceso a tratamientos y apoyo.
CARE_OPTIONS	Catagórica	Indica si el trabajador conoce o tiene opciones para acceder a servicios de salud mental en su lugar de trabajo.
WELLNESS_PROGRAM	Binaria	Relevante para identificar si existen programas preventivos que protejan el bienestar emocional de los equipos.
SEEK_HELP	Binaria	Refleja la predisposición del colaborador a buscar ayuda, un indicador clave para promover una cultura de apoyo.
ANONYMITY	Binaria	Mide si el trabajador percibe confidencialidad al hablar de su salud mental, lo que puede influir en su disposición a pedir ayuda.
MENTAL_HEALTH_INTERVIEW	Catagórica	Indica la percepción de estigma y comodidad para hablar de salud mental en el proceso de entrevista laboral.

### Criterios de selección:

- **Relevancia teórica y contextual:** Escogimos las variables porque ayudan a entender mejor cómo la salud mental afecta el trabajo. Así, cada variable aporta información útil y relevante para el objetivo del estudio.
- **Calidad de los datos:** Priorizamos aquellas variables con pocos datos faltantes y que fueran fáciles de codificar, para garantizar la fiabilidad y consistencia de nuestros análisis.

Estas variables fueron transformadas, normalizadas o codificadas según su tipo, para asegurar que fueran compatibles con los algoritmos de machine learning y que el modelo pudiera aprender de manera efectiva y confiable.

## 2.4. Modelado

### 2.4.1 Algoritmos seleccionados y justificación

Luego del procesamiento de la base de datos, en el ejercicio de machine learning utilizamos los siguientes 5 modelos con los siguientes resultados

Model	Accuracy	Balanced Accuracy	ROC AUC	F1 Score	Time Taken
AdaBoostClassifier	0.78	0.74	0.74	0.78	0.28
LGBMClassifier	0.78	0.74	0.74	0.77	0.35
XGBClassifier	0.78	0.73	0.73	0.77	1.16
BaggingClassifier	0.73	0.68	0.68	0.72	0.17
RandomForestClassifier	0.73	0.66	0.66	0.71	0.44

Teniendo en cuenta lo anterior, el modelo que mejor equilibrio presenta entre precisión, desbalance y eficiencia referente al tratamiento de la data es el **AdaBoost Classifier** el cual es un algoritmo que se encarga de combinar varios modelos simples para crear uno más fuerte y preciso, este modelo es una opción sólida entre rendimiento y velocidad

Esta capacidad de adaptación y mejora continua lo convierte en una solución sólida para extraer patrones ocultos en los datos y proponer recomendaciones efectivas, optimizando no sólo el rendimiento de los algoritmos, sino también la calidad de las decisiones empresariales orientadas al bienestar y la productividad de los colaboradores.

Si bien tenemos un segundo modelo **LGBMClassifier** que presenta comportamientos similares al anterior descrito, este fue descartado por su tiempo de cómputo que puede afectar la eficiencia, sin embargo, **LGBMClassifier** representa una alternativa sólida y escalable que podríamos considerar en el futuro, especialmente conforme nuestro conjunto de datos crezca y requiera una mayor capacidad de manejo de grandes volúmenes de información, permitiendo así aprovechar sus ventajas en velocidad y precisión a medida que la infraestructura y los recursos lo permita

### 2.4.2 Proceso de entrenamiento y validación

- Damos inicio con la base de datos traducida lo cual al hacer la limpieza de datos esta nos generaba ciertos errores en cuanto al delimitado, al formato de archivo que al momento de correr el código los resultados no eran coherentes, tomamos la decisión de utilizar la base original sin ningún tipo de intervención en donde el proceso de limpieza fue más preciso, como lo hemos descrito anteriormente.
- Para el entrenamiento y resultados del modelo utilizamos la función de “**dropna**” para eliminar las filas con datos faltantes, esto nos permitió hacer una depuración adecuada de los datos que no eran relevantes para el desarrollo del ejercicio, a ello se le suma la definición e identificación de la columna objetivo “**TREATMENT**”; luego de tener esta información la función “**HyperTransformer**” nos ayudó a transformar y preparar los datos para el modelado, para luego

dividirlos en conjunto de entrenamiento (80%) y prueba (20%).

### 2.4.3 Comparación entre modelos:

Modelo	¿Supervisado?	Detalle
AdaBoostClassifier	Sí	Es fácil de implementar y puede mejorar la precisión de clasificadores débiles, lo que lo convierte en un algoritmo flexible para diversas tareas de clasificación. Tiene buenos resultados en precisión (0.78) y es bastante rápido (0.28 seg).
LGBMClassifier	Sí	Muy rápido y eficiente, lo hace adecuado para conjuntos de datos grandes.
XGBClassifier	Sí	ofrece mejora de árboles en paralelo y es conocido por su rendimiento. Buen desempeño pero menos rápido (1.16 seg).
BaggingClassifier	Sí	Una técnica de aprendizaje por conjuntos que implica entrenar múltiples instancias del mismo clasificador en diferentes subconjuntos del conjunto de entrenamiento, y luego combinar sus predicciones mediante votación o promediado. Ayuda a reducir la varianza y a mejorar la estabilidad del modelo. (0.17 seg).
RandomForestClassifier	Sí	Los modelos de bosque aleatorio son conocidos por su alta precisión, robustez y capacidad para manejar grandes conjuntos de datos con muchas características. Precisión más baja (0.73) y más lento que Bagging.

## 2.5. Evaluación

### 2.5.1 Análisis de Resultados

El modelo AdaBoost fue el que mejor desempeño mostró, logrando identificar con confianza a los trabajadores que probablemente han necesitado tratamiento de salud mental. A pesar de trabajar con datos desbalanceados, el modelo pudo reconocer patrones importantes gracias a un cuidadoso proceso de preparación y transformación de la información.

### 2.5.2 Principales hallazgos:

- Variables como el género, la modalidad de trabajo, el tamaño de la empresa y la percepción del apoyo organizacional resultaron ser muy útiles para predecir riesgos.
- Se detectaron grupos con mayor vulnerabilidad, especialmente jóvenes que trabajan en ambientes poco flexibles o donde no se habla abiertamente sobre salud mental.
- Estos resultados demuestran que, con datos básicos, es posible anticipar cuándo alguien podría necesitar apoyo emocional, lo que permite diseñar estrategias de bienestar más efectivas y humanas.

### 2.5.3 Validación con expertos o stakeholders :

En otros países, ya existen modelos similares que ayudan a:

- Predecir el agotamiento laboral (burnout) en sectores como tecnología y salud (EE. UU., Reino Unido).
- Detectar en tiempo real riesgos de depresión y ansiedad usando inteligencia artificial (Singapur, Canadá).
- Apoyar decisiones clínicas en salud mental (Países Bajos, Alemania).
- Estos modelos se aplican en programas corporativos, clínicas y políticas públicas, mostrando que la tecnología puede ser una gran aliada para el bienestar.

#### 2.5.4 ¿Para qué sirve este modelo?

- **Empresas:** Detectar a tiempo problemas emocionales que afectan la productividad, reducir rotación y ausentismo, y ahorrar costos.
- **Clínicas y EPS:** Priorizar atención según el nivel de riesgo, optimizando recursos.
- **Sociedad:** Fomentar ambientes laborales saludables que disminuyan el impacto social y económico de la salud mental no tratada.

#### 2.5.5 Impacto económico y retorno de inversión

- El ausentismo por problemas emocionales puede causar pérdidas de hasta el 20-30% en productividad, según la OMS.
- Invertir en bienestar emocional puede generar un retorno de hasta 4 dólares por cada dólar invertido, según Deloitte.
- Este modelo ayuda a enfocar recursos en donde realmente se necesitan, evitando gastos innecesarios y maximizando el impacto

#### 2.5.6 Potencial para atención clínica

- Priorizar pacientes según urgencia emocional, mejorando la gestión en clínicas.
- Integrar factores sociales y laborales para ofrecer una atención más completa.
- Reducir tiempos de espera clasificando casos por gravedad.
- Apoyar programas comunitarios donde hay pocos especialistas.
- Este modelo puede adaptarse a sistemas de salud ocupacional y atención primaria, especialmente en EPS o IPS que buscan innovar en salud mental.

#### 2.5.7 Lecciones aprendidas

- Los datos laborales sí pueden predecir riesgos emocionales, pero deben manejarse con cuidado y calidad.
- La limpieza y codificación precisa de los datos fue clave para obtener buenos resultados.

- No todo lo que funciona en otros países se puede copiar tal cual; adaptar el modelo al contexto colombiano fue fundamental.
- El éxito requiere colaboración entre psicólogos, tecnólogos, recursos humanos y expertos en ética.
- Traducir los resultados técnicos a términos de negocio (como productividad y retorno de inversión) facilita su aceptación.

### **2.5.8 Limitaciones**

- La base de datos usada es extranjera, lo que puede afectar la precisión en Colombia.
- Hay un desbalance en la muestra, con más personas que han recibido tratamiento, lo que puede sesgar el modelo.
- La encuesta fue voluntaria, por lo que puede haber sesgo hacia quienes ya están interesados en salud mental.
- El modelo aún no ha sido validado clínicamente en Colombia, limitando su uso inmediato en atención médica.
- Es fundamental garantizar marcos éticos y legales estrictos para proteger la privacidad y el uso responsable de los datos.

## **2.6. Implementación**

El despliegue inicial del modelo predictivo se llevará a cabo en cinco pasos fundamentales que buscan facilitar su uso y maximizar su beneficio para las empresas:

- Primero, se prepara todo el entorno técnico necesario, instalando y configurando herramientas como Python, Streamlit y Power BI, que permitirán que el modelo funcione de manera fluida y accesible.
- Luego, se integra el modelo AdaBoost, que ya ha sido entrenado, adaptándolo a una interfaz visual e interactiva; esto significa que las empresas podrán explorar fácilmente los resultados y entender las predicciones sin necesidad de conocimientos técnicos profundos.
- En la tercera etapa, se crean visualizaciones claras y atractivas que muestran los niveles de riesgo detectados, cómo estos pueden afectar la productividad y ofrecen recomendaciones prácticas para mejorar. Después, se realiza una prueba piloto en una o dos empresas colombianas, con el fin de validar que el modelo realmente funcione con datos reales y sea útil en el contexto local.
- Finalmente, se ajusta la herramienta basándose en la retroalimentación recibida, se documenta todo el proceso para que sea transparente y se ofrece soporte básico para asegurar que el modelo esté listo para ser utilizado en el día a día laboral, ayudando a las empresas a tomar decisiones más informadas y efectivas.

Este enfoque busca no solo implementar una tecnología avanzada, sino también acompañar a las empresas en su adopción, garantizando que el modelo sea una herramienta práctica y comprensible que aporte valor real

Semana	Actividad	Herramientas / Salida esperada
1 y 2	Preparación técnica	Entorno Python + Streamlit configurado
3 y 4	Integración del modelo	Modelo AdaBoost funcional en plataforma
5 y 6	Desarrollo de visualizaciones	Dashboards en Power BI o Streamlit
7 y 8	Prueba piloto	Datos reales, retroalimentación empresarial
9 y 10	Ajustes + documentación	Versión final, manual de usuario, soporte

### 2.6.1 Estrategia de mantenimiento continuo:

El mantenimiento y actualización del modelo predictivo se enfocará en garantizar que siga siendo útil y preciso con el paso del tiempo. Aunque inicialmente se entrenó con la base de datos internacional de Kaggle sobre salud mental en tecnología, se implementará un proceso de reentrenamiento periódico que incorporará datos nuevos generados por las empresas colombianas en sus propios entornos. Esto permitirá que el modelo se adapte mejor a las particularidades y realidades locales, mejorando su capacidad para predecir de manera acertada.

Además, se realizará un monitoreo constante del rendimiento del modelo mediante métricas clave como precisión y recall, lo que permitirá detectar cualquier desviación en su comportamiento. En caso de identificar problemas, se actualizarán tanto el algoritmo como las visualizaciones para mantener la calidad y relevancia de la información presentada. También se llevará un control riguroso de las versiones del código y se aplicarán revisiones de seguridad y compatibilidad tecnológica, asegurando que la herramienta siga siendo confiable, funcional y adaptable a los cambios tecnológicos y necesidades del entorno empresarial.

Este enfoque integral de mantenimiento no solo protege la inversión tecnológica, sino que también asegura que las empresas puedan contar con una herramienta que evoluciona junto con sus datos y contextos, facilitando decisiones informadas y oportunas en la gestión de la salud mental laboral.

## 3. Presupuesto estimado

A continuación presentamos el presupuesto necesario para llevar a cabo un proyecto de salud mental apoyado en técnicas de Machine Learning, con una duración prevista de 2.5 meses de acuerdo al despliegue anterior. El proyecto contempla el uso de infraestructura en la nube con capacidad de procesamiento por GPU, ideal para manejar grandes volúmenes de datos y cargas de trabajo intensivas.



Concepto	Detalle / Supuestos clave	Total estimado (COP)
Licencias de software	Power BI Pro (2 usuarios), Streamlit Cloud (Team), GitHub Team, Microsoft 365	\$ 900.000,00
Sueldos Data Scientist (2.5 meses)	1 profesional @ \$8.000.000/mes x 2.5 meses	\$ 20.000.000,00
Infraestructura en la nube (GPU)	VM con GPU Nvidia T4 + Azure SQL Database + Blob Storage + App Service (Streamlit) + Tráfico, logs y respaldos	\$ 6.300.000,00
Visualizaciones / Reportes	Trabajo del visualizador + diseño + interacción	\$ 1.250.000,00
TOTAL GENERAL		\$ 28.450.000

Para asegurar el correcto funcionamiento del proyecto, se utilizarán los siguientes servicios en Microsoft Azure, seleccionados por su eficiencia y capacidad de adaptación a las necesidades del equipo:

- Máquina virtual con GPU (Nvidia T4): \$2,000,000 COP por mes. Permite el procesamiento acelerado de datos y modelos de Machine Learning.
- Azure SQL Database (nivel estándar): \$120,000 COP por mes. Facilita el almacenamiento y gestión segura de la información.
- Almacenamiento Blob (100 GB): \$50,000 COP por mes. Espacio suficiente para guardar archivos, datos y respaldos.
- Azure App Service (despliegue frontend): \$150,000 COP por mes. Permite poner en línea la interfaz de usuario y facilitar el acceso a los resultados.
- Otros (tráfico, respaldo, logs): \$200,000 COP por mes. Incluye gastos adicionales necesarios para el funcionamiento continuo y seguro del sistema.

En resumen, este presupuesto está diseñado para cubrir todos los aspectos clave del proyecto: desde el talento humano especializado, pasando por las herramientas tecnológicas necesarias, hasta la infraestructura que soportará el desarrollo y despliegue de las soluciones basadas en inteligencia artificial. Todo está pensado para garantizar un desarrollo eficiente, seguro y alineado con los objetivos de salud mental propuestos.

## 5. Conclusiones y Recomendaciones

### 5.1 Conclusiones claves del análisis:

- **Alta presencia de problemas de salud mental en profesionales de tecnología:** Cada vez más profesionales del sector tecnológico están enfrentando dificultades relacionadas con su salud mental, al punto que muchos han necesitado tratamiento. Esto confirma que, aunque es un tema común, sigue siendo poco reconocido y subestimado en el ambiente laboral.
- **El trabajo y la organización influyen mucho:** Factores como si se trabaja desde casa o en la oficina, el tamaño de la empresa y la percepción de apoyo que sienten los empleados afectan significativamente la probabilidad de que alguien busque ayuda o necesite tratamiento para su salud mental.
- **El estigma todavía pesa:** Aunque la sociedad está más abierta a hablar sobre salud mental, muchos trabajadores aún sienten miedo o vergüenza de expresar sus dificultades en el trabajo o durante procesos de selección, lo que dificulta que reciban ayuda a tiempo.



- **Los datos nos permiten anticipar riesgos:** Gracias a un modelo predictivo llamado AdaBoost, que utiliza información laboral y personal cuidadosamente procesada, podemos identificar con buena precisión quiénes podrían estar en riesgo, facilitando intervenciones tempranas y efectivas.
- **Adaptando el modelo para Colombia:** Aunque el modelo se creó con datos de otros países, las variables que usa son universales. Esto significa que puede ajustarse para funcionar en Colombia, especialmente en empresas que buscan implementar políticas de bienestar y cuidar a sus colaboradores

## 5.2 Recomendaciones para el Negocio o Futuras Investigaciones

### Para empresas y áreas de talento humano:

- **Implementar evaluaciones predictivas internas:** Utilizar herramientas similares al modelo desarrollado para detectar riesgo temprano en empleados, de forma confidencial y preventiva.
- **Diseñar estrategias personalizadas de bienestar:** A partir de los resultados del modelo, crear planes de acción por tipo de riesgo: talleres de autocuidado, atención psicológica, o cambios en la cultura organizacional.
- **Incluir salud mental en indicadores de productividad:** Integrar métricas de salud emocional como parte de la medición de desempeño, para actuar sobre las causas y no solo sobre los síntomas del bajo rendimiento.
- **Desarrollar una cultura de cuidado:** Fomentar la formación de líderes con sensibilidad mental y generar espacios de conversación segura para disminuir el estigma.

### Para futuras investigaciones:

- **Recolectar datos propios en Colombia:** Crear una línea base local permitirá entrenar un modelo más representativo y aumentar la precisión predictiva.
- **Analizar el impacto económico de la salud mental:** Estimar cuánto afecta la salud emocional en costos como ausentismo, rotación y bajo rendimiento puede fortalecer el caso de negocio.
- **Explorar modelos explicativos:** Combinar modelos predictivos con modelos que expliquen el "por qué" puede ayudar a tomar decisiones más estratégicas en RRHH.

## 5.3 Consideraciones Éticas en el Uso de Modelos Predictivos para Salud Mental

- **Privacidad y anonimato:** Es fundamental que los datos de los empleados se mantengan siempre anónimos y seguros. Los resultados del modelo no deben usarse para juzgar o discriminar a nadie, sino para brindar apoyo y cuidado a quienes lo necesiten.
- **Consentimiento informado:** Cada persona debe entender claramente para qué se usarán sus datos, cómo se protegerán y qué beneficios puede esperar. La transparencia genera confianza y respeto.
- **No etiquetar ni excluir:** El objetivo del modelo es ayudar a prevenir riesgos y mejorar el bienestar, no para etiquetar a alguien o limitar sus oportunidades laborales por su historial emocional.

- **Equidad y respeto** Hay que asegurarse de que las decisiones basadas en el modelo no reproduzcan prejuicios por género, edad, origen o cualquier otra condición. La justicia y la inclusión deben estar siempre presentes.
- **Responsabilidad compartida** La salud mental es un compromiso tanto del individuo como del entorno laboral. Las empresas deben usar la información para crear espacios más saludables y apoyar a sus colaboradores, no solo para señalar problemas individuales.

## Bibliografía

- Engels, F. (1876). El papel del trabajo en la transformación del mono en hombre. En Dialéctica de la naturaleza. Recuperado de <https://www.marxists.org/espanol/m-e/1870s/1876trab.htm>
- Heath, C., & Heath, D. (2007). Made to Stick: Why Some Ideas Survive and Others Die. Random House
- Chiavenato, I. (2009). Gestión del talento humano (3.<sup>a</sup> ed.). McGraw-Hill.
- Shifman, L. (2014). Memes in digital culture. MIT Press.
- El Economista. (s.f.). *¿Cuánto le cuesta a tu empresa el ausentismo?* <https://www.eleconomista.com.mx/el-empresario/Cuanto-le-cuesta-a-tu-empresa-el-ausentismo-Invertir-en-la-salud-laboral-20220922-0089.html>
- PRAX. (s.f.). *¿Cuánto le cuesta a las empresas el trabajo no saludable?* <https://prax.com.co/post/cu%C3%A1nto-le-cuesta-a-las-empresas-el-trabajo-no-saludable>
- Expansión. (2023). *La calculadora que mide el costo de ignorar la salud mental de los trabajadores.* <https://expansion.mx/carrera/2023/10/10/la-calculadora-coste-ignorar-salud-mental-trabajadores>
- La Nota Económica. (s.f.). *Empresas colombianas pueden ahorrar entre un 12% y un 15% si cuidan la salud mental de sus empleados.* <https://lanotaeconomica.com.co/movidas-empresariales/empresas-colombianas-pueden-ahorrar-entre-un-12-y-un-15-si-cuidan-la-salud-mental-de-sus-empleados/>
- Gallup. (2022). *State of the Global Workplace Report.* <https://www.gallup.com/workplace/349484/state-of-the-global-workplace.aspx>
- El Economista. (s.f.). *¿Cuánto le cuesta a tu empresa el ausentismo?* <https://www.eleconomista.com.mx/el-empresario/Cuanto-le-cuesta-a-tu-empresa-el-ausentismo-Invertir-en-la-salud-laboral-20220922-0089.html>
- RRHHDigital. (s.f.). *15 veces más ausentismo y 6 veces más rotación: el costo de no cuidar la salud mental.* <https://www.rrhhdigital.mx/secciones/bienestar/15134/15-veces-mas-ausentismo-y-6-veces-mas-rotacion-el-costo-de-no-cuidar-la-salud-mental>

- Texto del proyecto (información sobre AdaBoost y variables sociodemográficas).
- La Nota Económica. (s. f.). Empresas colombianas pueden ahorrar entre un 12% y un 15% si cuidan la salud mental. <https://lanotaeconomica.com.co/movidas-empresariales/empresas-colombianas-pueden-ahorrar-entre-un-12-y-un-15-si-cuidan-la-salud-mental-de-sus-empleados/>
- El Economista (costo de rotación laboral).
- Seguros SURA. (s. f.). Conoce el impacto económico del ausentismo laboral y cómo gestionarlo. <https://segurossura.com/co/blog/revista/conoce-el-impacto-economico-del-ausentismo-laboral/>
- Open Sourcing Mental Illness. (2016). *Mental health in tech survey* [Dataset]. Kaggle. <https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey>