

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

¿Qué hay detrás del desempeño?

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

¿Qué hay detrás del desempeño?

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

Introducción

El trabajo al ser una necesidad básica que ha formado parte de la historia del ser humano y a su vez, un elemento constitutivo de su qué hacer en el mundo nos lleva a cuestionarnos si ¿están las organizaciones preparadas para entender a sus colaboradores desde una visión integral?

Este estudio revela historias detrás de los datos, a través de la construcción de un modelo predictivo apoyado en técnicas de Machine Learning (ML), busca proponer soluciones estratégicas en relación a la salud mental y constituirse como una herramienta clave para empresas que opten por cuidar a su talento y potenciar sus resultados.



Planteamiento del problema:

¿Puede una empresa realmente alcanzar su máximo potencial si no cuida la salud mental y el bienestar emocional de quienes la hacen posible todos los días?

Perspectiva Humana

- Convertir datos en acciones concretas
- Entender a las personas tras los números
- Identificar oportunidades para ambientes saludables

Modelos de Análisis

- Precisión y eficiencia en el análisis
- Diseño de intervenciones personalizadas
- Promoción del bienestar y optimización organizacional





¿Qué hay detrás del desempeño?

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

01

Objetivo general

- Desarrollar un modelo predictivo utilizando técnicas de aprendizaje automático que determine los factores que afectan el rendimiento laboral debido a problemas de salud mental, en un esfuerzo por el bienestar en el trabajo.

02

Objetivos Específicos

- Examinar los factores asociadas en relación con la accesibilidad de los servicios de salud mental en el lugar de trabajo.
- 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
- Evaluar los resultados del modelo predictivo y formular políticas organizacionales que inviten a la participación activa hacia la optimización de la salud psicológica y mejorar la productividad en las empresas

ALCANCE DEL PROYECTO

A NIVEL INTERNACIONAL



Datos globales



Modelo Predictivo



Impacto Financiero

ENFOQUE EN COLOMBIA



Modalidades de trabajo



Encuesta nacional

Proporciona una estructura clara y sistemática para desarrollar proyectos de análisis de datos y modelado predictivo para desarrollar proyectos de análisis y modelado predictivo

Metodología CRISP-DM (Cross-Industry Standard Process for Data Mining),

Está definido por 6 fases



Comprensión de los datos

Recolectar, describir y explorar los datos para identificar su calidad y relevancia.

Modelado

Aplicar algoritmos y técnicas para crear modelos predictivos o descriptivos

Despliegue

Implementar el modelo en el entorno real para generar valor en el negocio

Comprensión del negocio

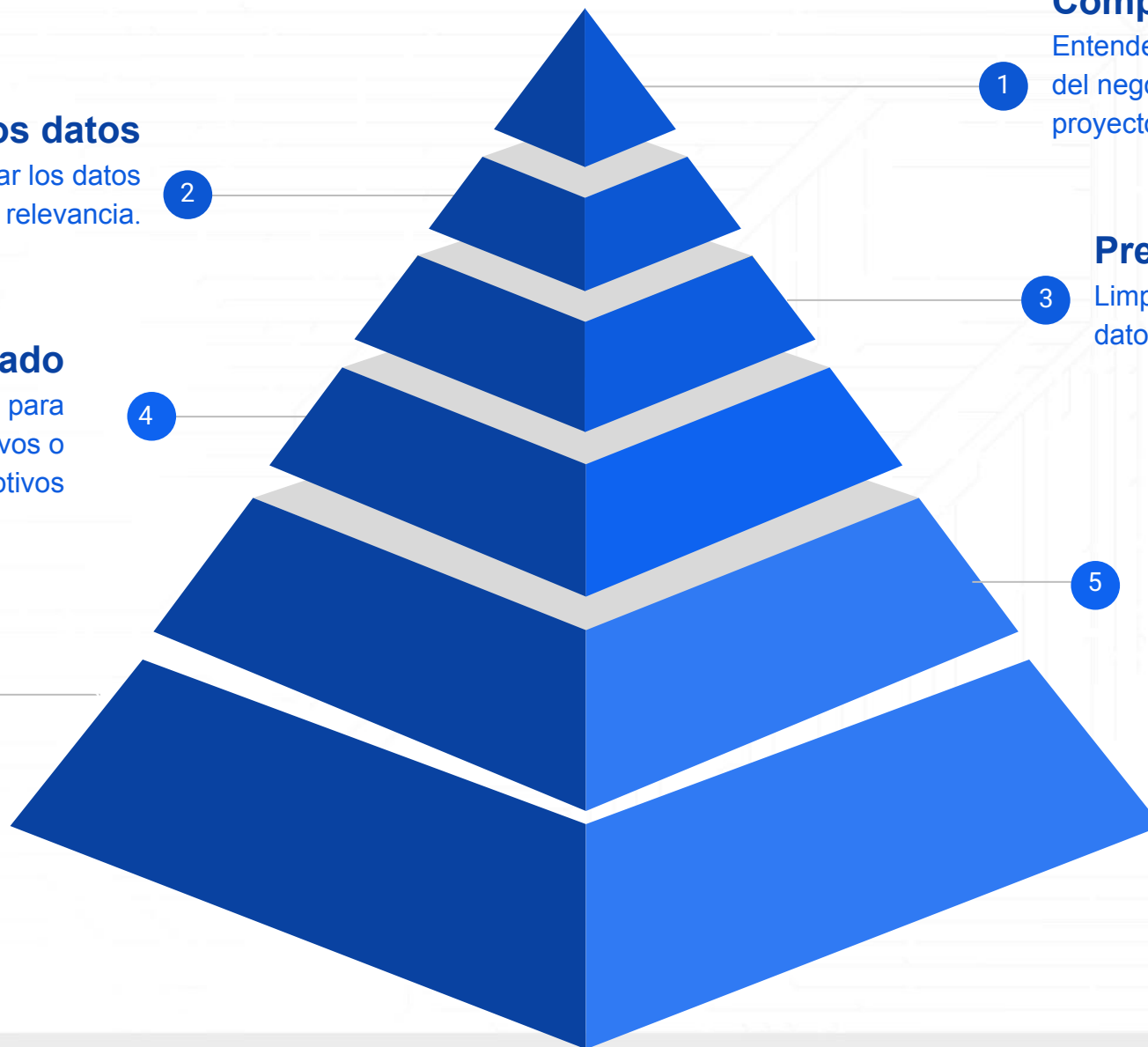
Entender los objetivos y necesidades del negocio para definir metas del proyecto

Preparación de los datos

Limpiar, transformar y seleccionar los datos adecuados para el modelado.

Evaluación

Verificar que los modelos cumplen los objetivos del negocio y seleccionar el mejor.



Comprensión del Negocio

Contexto:

Problemas como el ausentismo, la rotación y la baja productividad generan costos significativos para las organizaciones.
El modelo predictivo permite anticipar riesgos y tomar decisiones estratégicas

De cara al negocio

Objetivo principal

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.

Mejora del bienestar emocional.

Ahorro económico por prevención.

Fortalecimiento de una cultura organizacional más humana y productiva

Detección temprana de riesgos

Acciones preventivas centradas en las personas

Reducción de ausentismo y rotación

Mejora en la eficiencia y bienestar laboral

Beneficios

Resultados Esperados

Definición de Objetivos

Anticipar el impacto del bienestar emocional en el rendimiento laboral.

Recopilación de BBDD

Utilización de la base de datos "Mental Health in Tech Survey" de Kaggle.

1.251 registros, 27 variables clave.

Análisis de Variables

Identificación de variables categóricas, booleanas y numéricas.

Variable objetivo: tratamiento de salud mental.

Desarrollo del Modelo

Creación de un modelo predictivo basado en técnicas de Machine Learning

Factores sociodemográficos, laborales y de percepción como predictores.

Criterios de Éxito

Impacto claro y relevante del modelo.

Medición del efecto en productividad y bienestar.

ANÁLISIS DE VARIABLES

Variable	Tipo	Justificación
AGE	Numérica	Refleja la etapa de vida, ayudando a entender cómo varían las necesidades y percepciones sobre salud mental entre generaciones.
GENDER	Categórica	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	Categórica	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	Categórica	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	Categórica	Indica la percepción de estigma y comodidad para hablar de salud mental en el proceso de entrevista laboral.

PREPARACIÓN DE DATOS: LIMPIEZA Y TRANSFORMACIÓN



BBDD INICIAL

Column	Non-Null	Count	Dtype
AGE	1251	non-null	int64
GENDER	1251	non-null	object
COUNTRY	1251	non-null	object
SELF_EMPLOYED	1233	non-null	object
FAMILY_HISTORY	1251	non-null	object
TREATMENT	1251	non-null	object
WORK_INTERFERE	989	non-null	object
NO_EMPLOYEES	1251	non-null	object
REMOTE_WORK	1251	non-null	object
TECH_COMPANY	1251	non-null	object
BENEFITS	1251	non-null	object
CARE_OPTIONS	1251	non-null	object
WELLNESS_PROGRAM	1251	non-null	object
SEEK_HELP	1251	non-null	object
ANONYMITY	1251	non-null	object
LEAVE	1251	non-null	object
MENTAL_HEALTH_CONSEQUENCE	1251	non-null	object
PHYS_HEALTH_CONSEQUENCE	1251	non-null	object
COWORKERS	1251	non-null	object
SUPERVISOR	1251	non-null	object
MENTAL_HEALTH_INTERVIEW	1251	non-null	object
PHYS_HEALTH_INTERVIEW	1251	non-null	object
MENTAL_VS_PHYSICAL	1251	non-null	object
OBS_CONSEQUENCE	1251	non-null	object

ÚNICOS Y FALTANTES POR COLUMNA

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

TRANSFORMACIÓN

#	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

HALLAZGOS EN VARIABLES:

GÉNERO



■ Male

■ F

■ Male

■ Man

■ Androgyne

■ Male-ish

■ Female

■ Make

■ Cis Male

■ Female (trans)

■ non-binary

■ Female (cis)

■ M

■ Woman

■ Female

■ Femake

■ femail

■ Nah



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

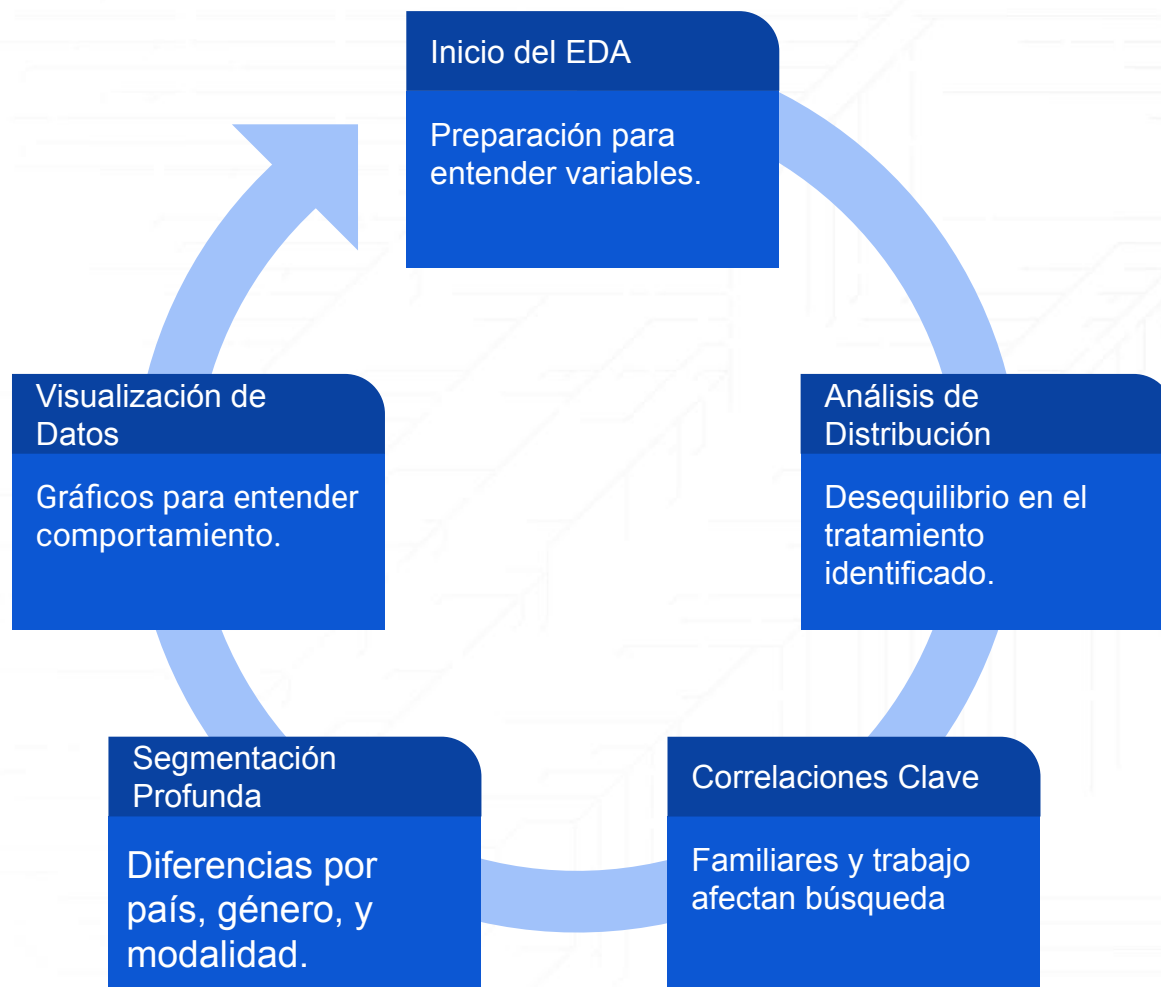
Se detectó una gran diversidad de formas de registrar el género lo que requirió un proceso cuidadoso de normalización mediante un diccionario de mapeo estandarizado.

EDAD

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

Aparecieron valores negativos, incoherentes o menores de 18 años, lo cual no es coherente con un entorno laboral formal por lo tanto, fue transformada a un rango válido y tratada como variable numérica antes de su codificación.

Flujo del Análisis Exploratorio de Datos



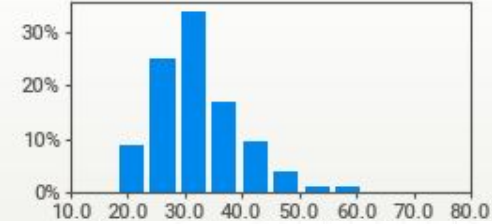
GRAFICAS EDA

AGE

VALUES: 971 (100%)
MISSING: —
DISTINCT: 44 (5%)
ZEREOES: —

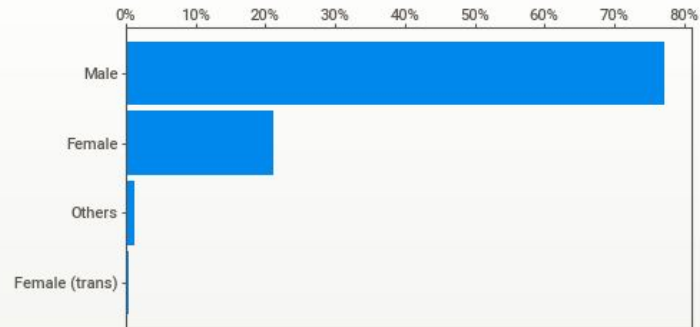
MAX 72.0
95% 45.0
Q3 36.0
AVG 32.3
MEDIAN 31.0
Q1 27.0
5% 23.0
MIN 18.0

RANGE 54.0
IQR 9.00
STD 7.40
VAR 54.7
KURT. 1.74
SKEW 0.989
SUM 31,370



GENDER

MISSING: —

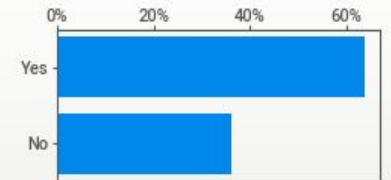


TOP CATEGORIES

Male	750	77%
Female	206	21%
Others	11	1%
Female (trans)	4	<1%
ALL	971	100%

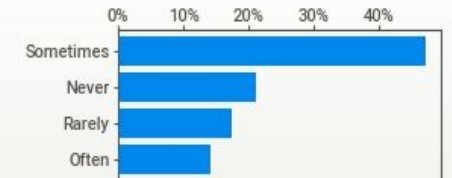
TREATMENT

VALUES: 971 (100%)
MISSING: —
DISTINCT: 2 (<1%)



WORK_INTERFERE

VALUES: 971 (100%)
MISSING: —
DISTINCT: 4 (<1%)



Factores psicosociales críticos:

- **Presión laboral y sobrecarga** → Estrés, ansiedad, burnout.
- **Falta de reconocimiento y discriminación** → Baja motivación.
- **Monotonía laboral** → Riesgos físicos y mentales.

Análisis segmentado:

- 👥 **Generaciones** (Millennials, Gen X, Baby Boomers) → Estrategias adaptadas.
- 📅 **Modalidad de trabajo** (presencial, remoto, híbrido) → Intervenciones precisas.
- 🕒 **Horarios laborales** → Ajuste de recursos.
- ♀️ **Equidad de género** → Acciones inclusivas.



Generación	Edad Aproximada	Factores de Riesgo	Datos Clave	Estrategias de Apoyo
Generación Z	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
Millennials	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
Generación X	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
Baby Boomers	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

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.

Modelado

Algoritmos seleccionados:

Luego del procesamiento de la base de datos, en el ejercicio de Machine Learning utilizamos los siguientes modelos con sus respectivos 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

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.



Implementación

Despliegue en 5 pasos:

1. **Preparación técnica:** Configuración de Python, Streamlite y Power BI.
2. **Integración del modelo:** AdaBoost con interfaz visual e interactiva.
3. **Visualizaciones útiles:** Riesgos, impacto en productividad y recomendaciones.
4. **Prueba piloto:** Validación en empresas colombianas con datos reales.
5. **Ajustes finales:** Correcciones según retroalimentación, documentación y soporte básico.

Estrategia de mantenimiento continuo:

- **Reentrenamiento periódico:** Con nuevos datos locales para mejorar precisión y relevancia.
- **Monitoreo constante:** Uso de métricas clave (precisión, recall) para detectar desviaciones.
- **Actualización del sistema:** Ajustes en algoritmo, visualizaciones y compatibilidad tecnológica.
- **Gestión de versiones y seguridad:** Revisión regular del código y entorno técnico.



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

Presupuesto

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 × 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

Infraestructura (Azure con GPU)

VM con GPU Nvidia T4, Azure SQL Database, Blob Storage, App Service (Streamlit) Tráfico, logs y respaldos

Herramientas clave:

Python, Scikit-learn, Streamlit, Power BI Pro, GitHub Team



Conclusiones

- **Los problemas de salud mental en tecnología son comunes y crecientes**, pero aún subestimados por las empresas.
- **El entorno laboral influye directamente:** tipo de trabajo, tamaño de empresa y apoyo organizacional afectan el riesgo.
- **El estigma sigue siendo un freno:** muchas personas no piden ayuda por miedo o vergüenza.
- **Los datos predicen riesgos:** el modelo AdaBoost permite identificar a tiempo quién podría necesitar apoyo.
- **Modelo adaptable a Colombia:** usa variables universales y es aplicable en empresas que quieran cuidar a su gente.



Recomendaciones

Para las empresas y equipos de talento humano:

- Implementar evaluaciones predictivas internas
- Diseñar estrategias de bienestar personalizadas
- Incluir la salud mental en los indicadores de productividad
- Capacitar a los líderes para que tengan sensibilidad hacia la salud mental
- Crear espacios seguros donde los empleados puedan hablar abiertamente, ayudando a reducir el estigma y fortalecer el equipo.

Para futuras investigaciones:

- Recolectar datos propios en Colombia: Generar una base de datos local permitirá entrenar modelos más precisos y ajustados a nuestro contexto, mejorando la efectividad de las predicciones.
- Analizar el impacto económico de la salud mental: Estimar cuánto afectan los problemas emocionales a costos como el ausentismo, la rotación y la baja productividad, para fortalecer el argumento de invertir en bienestar.
- Explorar modelos explicativos: Combinar modelos que no solo predigan riesgos, sino que también expliquen las causas, ayudará a tomar decisiones más estratégicas y efectivas en recursos humanos.

Consideraciones Éticas del Modelo Predictivo

- **Privacidad ante todo:**
Los datos de las personas siempre deben mantenerse anónimos y usarse únicamente para brindar apoyo, nunca para juzgar o discriminar.
- **Consentimiento claro y transparente:**
Cada persona tiene derecho a saber cómo se usarán sus datos y qué beneficios podrá obtener de ello.
- **Sin etiquetas ni exclusiones:**
El objetivo del modelo es prevenir riesgos y acompañar a quienes lo necesiten, no limitar sus oportunidades laborales ni estigmatizarlos.
- **Equidad y respeto para todos:**
Es fundamental evitar cualquier sesgo por género, edad, origen u otra condición. Las decisiones deben ser justas, inclusivas y respetuosas.
- **Responsabilidad compartida:**
La salud mental es un compromiso de todos: individuos, equipos y organizaciones. El modelo debe servir para mejorar el entorno y no para señalar culpables.



Impacto actual

- Este modelo permite anticipar riesgos de salud mental en el trabajo, ayudando a identificar a tiempo a quienes podrían necesitar apoyo.
- Facilita que las decisiones empresariales sean más humanas y estratégicas, poniendo el bienestar de las personas en el centro.
- Abre la puerta a una cultura organizacional que prioriza el cuidado emocional y el equilibrio, mejorando el ambiente laboral y la productividad.



Líneas Futuras para el Modelo Predictivo

- **Adaptación con datos locales:**
Entrenar el modelo con información real de empresas colombianas para que sea más preciso y muestre mejor nuestra realidad laboral.
- **Integración con plataformas empresariales:**
Incorporar el modelo en las herramientas de Recursos Humanos y de Business Intelligence que ya usan las empresas, facilitando su uso diario y la toma de decisiones.
- **Soporte clínico:**
Utilizar el modelo para mejorar la atención en clínicas de salud mental, ayudando a priorizar casos y ofrecer un seguimiento más efectivo.
- **Investigación continua:**
Ampliar el análisis a otros sectores económicos para entender mejor cómo la salud mental impacta diferentes áreas y adaptar las estrategias de bienestar.



GRACIAS

Por: Ana María Ramírez | Johanna Gonzalez | Santiago Quintero
Mayo 2025