Predicción de Riesgo de Burnout Laboral Mediante Machine Learning



Yulieth Andrea Salazar · Kevin Felipe Marín Roa · Corinne Hernández Beltrán
Tutor: Ricardo Andrés Villalba Rivera
Universidad EAN · Especialización en Machine Learning · Bogotá, 2026

El Problema

35%

Incremento del estrés laboral en la última década (OIT, 2023)

28%

Trabajadores colombianos con síntomas de burnout (MinSalud, 2024)

6–12

Meses de desfase en métodos tradicionales de detección

¿Cuál es el desafío?

los métodos tradicionales actúan de forma reactiva y tardia . Para cuando el problema se detecta, el empleado ya sufrió consecuencias graves en su salud.

Solución propuesta

Modelo predictivo de ML que identifica el riesgo ANTES de que se manifieste, usando 5 variables psicosociales clave del contexto colombiano.

Pregunta de Investigación

¿Qué modelo de machine learning permite predecir de forma más precisa el riesgo de burnout laboral y cuales son las variables que tienen mayor influencia?

Dataset sintético · 1.500 registros · Financiero | Medios | Publicidad

Variables predictoras del modelo

DEMANDA	DEMANDA	DEMANDA	DEMANDA	RECURSO
Carga Laboral	Estrés Percibido	Horas de Trabajo	Conflicto de Rol	Apoyo Liderazgo

Objetivo General

Desarrollar un modelo predictivo basado en machine learning que permita identificar el riesgo de burnout laboral a partir de variables psicosociales y laborales; además comparando diferentes algoritmos para identificar cual ofrece el mejor desempeño.

Objetivos Específicos

OE1 Identificar las variables del burnout con base en la literatura especializada (modelo JD-R). Karasek, Bakker y Demerouti.

OE2 Implementar y comparar 4 algoritmos supervisados: Reg. Logística, Random Forest, XGBoost y MLP.

OE3 Evaluar el desempeño del MLP frente a modelos clásicos e identificar variables con mayor poder predictivo.

OE4 Generar recomendaciones para implementar el modelo como herramienta de alerta temprana en RRHH.

Variables del Modelo Predictivo

Variable	Tipo de Dato	JD-R	Descripción Operativa
Carga_Laboral	Numérico 1–10	**DEMANDA**	Volumen de trabajo, complejidad y plazos ajustados
Estrés	Numérico 1–10	**DEMANDA**	Tensión y presión percibida en actividades laborales
Horas_Trabajo	Numérico (6–11h)	**DEMANDA**	Exposición objetiva a demandas: horas diarias laboradas
Conflicto_Rol	Numérico 1–10	**DEMANDA**	Ambigüedad o incompatibilidad en expectativas del rol
Apoyo_Liderazgo	Numérico 1–10	**RECURSO**	Respaldo percibido de supervisores. Factor PROTECTOR ↑apoyo = ↓riesgo

Variable objetivo: Riesgo_Burnout — binaria (Alto / Bajo)

Diseño Metodológico

Enfoque	Datos	División	Validación	Algoritmos
Cuantitativo Predictivo	Sintético 1.500 registros Basado en Modelos estadísticos	Train 80% Test 20%	K-Fold CV (k=5) Early Stopping MLP	4 modelos comparados

Pipeline metodológico

01 **Revisión Literatura**
Karasek, Bakker y Demerouti.

02 **Definición Variables**

03 **Dataset Sintético**

04 **Preprocesa-miento**

05 **4 Modelos ML**

06 **Evaluación Métricas**

Datos sintéticos: control experimental sin comprometer privacidad — cumple Ley 1581/2012. La comparación evalúa parsimonia: ¿mayor complejidad = mejor desempeño?

Cuatro Algoritmos Comparados

RL

Modelo Lineal · Baseline

Regresión Logística

Función sigmoide. Alta interpretabilidad mediante coeficientes β. Sensible a escala ⟶ StandardScaler.



Alta interpretabilidad

RF

Ensemble · Bagging

Random Forest

200 árboles independientes combinados por votación. Robusto ante ruido. Invariante a la escala.



Alta robustez

XG

Ensemble · Boosting

XGBoost

Árboles secuenciales correctivos por gradiente. scale_pos_weight para desbalance. Alto rendimiento.



Alto rendimiento

MLP

Deep Learning

Red Neuronal (MLP)

Dense(32)+ReLU+BN+Dropout ⟶ Dense(16)+L2 ⟶ Sigmoid. Early Stopping. Captura patrones no lineales.



No lineal

¿Qué encontramos?

🏆 **Mejor desempeño Regresión Logística**

84%
Accuracy
De cada 100 acertó en 84

9 de 10
AUC-ROC 0.929
Separa riesgo alto de bajo

4 de 5
Recall 81.1%
Personas en riesgo real detectadas
Minimiza Falsos Negativos

¿Y qué lo provoca?

Barra más larga = más peso en la predicción

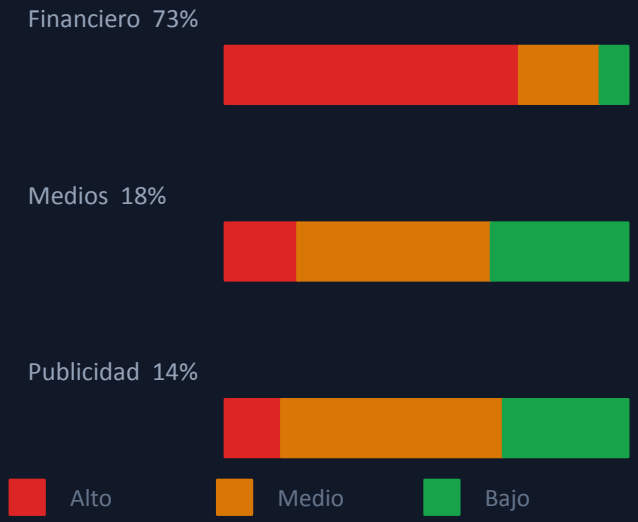


- El estrés es el factor que más mueve la predicción
- Apoyo del lider: va al revés — más apoyo, menos riesgo
- Valida los modelos teóricos de Karasek, Bakker y Demerouti.

Burnout Predictive Analytics

247	23%	41%	36%
Total empleados	Alto Riesgo	Riesgo Medio	Bajo Riesgo
3 sectores	↑ 3% vs mes anterior	Monitoreo activo	Sin intervención

Distribución de Riesgo por Sector



Peso de Variables



⚠ Alertas activas

A. Torres	Financiero	CRÍTICO
M. Vargas	Medios	ALTO
L. Ospina	Publicidad	ALTO
P. Cruz	Financiero	ALTO

¿Qué aprendimos?

01

Más complejo no significa mejor

La Regresión Logística — el modelo más sencillo — le ganó a la red neuronal y al XGBoost. Para este problema, la solución directa es la correcta.

02

Los resultados tienen sentido

Estrés y Carga Laboral como principales predictores coincide con la ciencia del burnout. El modelo aprendió algo real, no patrones al azar.

03

Esto tiene impacto real, hoy

AUC > 0.90 en los 4 modelos: identifica correctamente 4 de cada 5 personas en riesgo — antes de que el problema explote.

"El burnout no avisa —
pero los datos sí pueden hacerlo."

Muchas gracias

Yulieth Andrea Salazar · Kevin Felipe Marín Roa · Corinne Hernández Beltrán