

# ANÁLISIS DE ESTRÉS EN ESTUDIANTES

## Comparación: Datos Mundiales vs UTEQ

Utilizando Redes Neuronales Artificiales

*Fecha del análisis: 07 de November de 2025*

### 1. INTRODUCCIÓN

Este documento presenta un análisis comparativo del estrés en estudiantes universitarios utilizando técnicas de aprendizaje automático, específicamente Redes Neuronales Artificiales (RNA). El estudio compara datos reales de estudiantes a nivel mundial con datos simulados de la Universidad Tecnológica de Querétaro (UTEQ).

### 2. METODOLOGÍA

#### 2.1 Fuentes de Datos

**Datos Mundiales:** Se utilizó un conjunto de datos con 2,028 registros de estudiantes de diversas universidades internacionales. Los datos incluyen información demográfica y respuestas a 10 preguntas sobre estrés académico medidas en una escala de 0-4.

**Datos UTEQ:** Se generó un conjunto de 1,000 registros simulados para estudiantes de la UTEQ, manteniendo características similares a los datos mundiales pero con variaciones propias del contexto educativo mexicano. La simulación incluyó carreras específicas de la UTEQ como Ingeniería en Desarrollo de Software, Energías Renovables, Nanotecnología, entre otras.

#### 2.2 Variables del Estudio

Variable	Descripción
Edad	Rango de edad del estudiante (18-22, 23-26, 27-30)
Género	Masculino o Femenino
Universidad	Institución educativa
Departamento	Carrera o área de estudio
Año Académico	Nivel en la carrera (1º, 2º, 3º año u otro)

CGPA	Promedio académico actual (escala 2.0-4.0)
Beca	Si recibe beca o apoyo financiero
Preguntas 1-10	Mediciones de estrés (escala 0-4)

### 2.3 Arquitectura de Red Neuronal

Se implementó una Red Neuronal Multicapa (MLP - Multi-Layer Perceptron) con la siguiente configuración:

Componente	Especificación
Capa de Entrada	10 neuronas (una por cada pregunta de estrés)
Capas Ocultas	3 capas: 100, 50 y 25 neuronas respectivamente
Capa de Salida	3 neuronas (para clasificación triclase)
Función de Activación	ReLU (Rectified Linear Unit)
Optimizador	Adam (Adaptive Moment Estimation)
Máximo de Iteraciones	500 épocas
Early Stopping	Activado (parada temprana si no hay mejora)

### 2.4 Proceso de Entrenamiento

Los datos se dividieron en conjuntos de entrenamiento (80%) y prueba (20%) utilizando estratificación para mantener la proporción de clases. Se aplicó normalización StandardScaler para estandarizar las características. El modelo se entrenó con validación cruzada de 5-fold para evaluar su capacidad de generalización.

## 3. RESULTADOS

### 3.1 Distribución de Niveles de Estrés

Nivel de Estrés	Datos Mundiales	Datos UTEQ
Bajo (Low Stress)	115 (5.7%)	0 (0%)
Moderado (Moderate)	1,348 (66.5%)	781 (78.1%)
Alto (High Perceived)	565 (27.8%)	219 (21.9%)
Total	2,028	1,000

**Observación:** Los datos simulados de UTEQ muestran una tendencia hacia estrés moderado-alto, sin casos de estrés bajo. Esto puede reflejar las demandas particulares del sistema educativo tecnológico mexicano.

### 3.2 Rendimiento de los Modelos

Métrica	Modelo Mundial	Modelo UTEQ
Accuracy Entrenamiento	99.88%	96.25%
Accuracy Prueba	98.52%	95.00%
Iteraciones Necesarias	54	36
Precision (Promedio)	97%	95%
Recall (Promedio)	95%	90%
F1-Score (Promedio)	96%	92%

**Interpretación:** Ambos modelos alcanzaron un rendimiento excelente, superando el 95% de precisión en datos de prueba. El modelo mundial mostró un accuracy ligeramente superior (98.52% vs 95.00%), posiblemente debido al mayor volumen de datos de entrenamiento (2,028 vs 1,000 muestras). No hay evidencia de sobreajuste significativo ya que la diferencia entre entrenamiento y prueba es razonable en ambos casos.

## 4. ANÁLISIS DE CURVAS DE APRENDIZAJE

Las curvas de aprendizaje muestran cómo evoluciona el rendimiento del modelo a medida que aumenta el tamaño del conjunto de entrenamiento. Son fundamentales para diagnosticar problemas como underfitting (subajuste) o overfitting (sobreajuste).

### 4.1 Interpretación de las Curvas

**Datos Mundiales:** La curva muestra una convergencia rápida y estable. El modelo alcanza alto rendimiento con relativamente pocas muestras (~400) y se mantiene consistente con más datos. La pequeña brecha entre las curvas de entrenamiento y validación indica buen balance entre sesgo y varianza.

**Datos UTEQ:** Similar comportamiento al modelo mundial, aunque con ligeramente mayor variabilidad inicial. La convergencia se logra alrededor de las 200 muestras. El modelo demuestra buena capacidad de generalización.

### 4.2 Gráficos Generados

Se generaron tres visualizaciones principales:

- **Curva de Aprendizaje - Datos Mundiales:** Muestra la evolución del accuracy para el conjunto de datos internacional, comparando entrenamiento vs validación.
- **Curva de Aprendizaje - Datos UTEQ:** Presenta el mismo análisis para los datos simulados de la Universidad Tecnológica de Querétaro.
- **Comparación Mundial vs UTEQ:** Gráfico comparativo directo que permite visualizar las diferencias de aprendizaje entre ambos conjuntos de datos.
- **Matrices de Confusión:** Muestran la distribución de predicciones correctas e incorrectas para cada clase de estrés en ambos modelos.

## 5. CONCLUSIONES Y HALLAZGOS

### 5.1 Principales Hallazgos

**1. Efectividad del Modelo:** Las Redes Neuronales Artificiales demostraron ser altamente efectivas para clasificar niveles de estrés en estudiantes, logrando precisiones superiores al 95% en ambos conjuntos. **2. Patrones de Estrés:** Los datos UTEQ simulados muestran predominancia de estrés moderado a alto, sin casos de estrés bajo, lo que sugiere un entorno académico exigente característico de las instituciones tecnológicas. **3. Generalización:** Ambos modelos presentan buena capacidad de generalización, evidenciada por la pequeña diferencia entre accuracy de entrenamiento y prueba. **4. Eficiencia del Entrenamiento:** El modelo UTEQ convergió más rápido (36 iteraciones vs 54), posiblemente debido a la menor complejidad de los datos (solo 2 clases vs 3 en el modelo mundial). **5. Tamaño de Muestra:** Las curvas de aprendizaje indican que con ~400-500 muestras se alcanza un rendimiento cercano al óptimo, sugiriendo que no es necesario un conjunto masivo de datos para este tipo de clasificación.

### 5.2 Recomendaciones

- **Para la UTEQ:** Implementar encuestas reales basadas en las 10 preguntas de estrés para validar y mejorar el modelo con datos reales de estudiantes.
- **Intervención Temprana:** Utilizar el modelo como herramienta de detección temprana para identificar estudiantes en riesgo de estrés alto y ofrecer apoyo psicológico oportuno.
- **Seguimiento Longitudinal:** Realizar mediciones periódicas a lo largo del semestre para monitorear cambios en los niveles de estrés.
- **Análisis por Carrera:** Realizar análisis específicos por departamento/carrera para identificar programas académicos con mayores niveles de estrés.
- **Mejoras al Modelo:** Incorporar variables adicionales como horas de sueño, actividad física, y apoyo social para mejorar las predicciones.

## 6. ARCHIVOS Y RECURSOS GENERADOS

Los siguientes archivos fueron generados durante este análisis:

<b>Datos:</b>	• stress_uteq.csv - Conjunto de datos simulados para UTEQ (1,000 registros)
<b>Gráficos:</b>	• curva_aprendizaje_mundial.png - Curva de aprendizaje para datos mundiales
<b>Código:</b>	• curva_aprendizaje_uteq.png - Curva de aprendizaje para datos UTEQ
<b>Código:</b>	• curva_comparacion_mundial_vs_uteq.png - Comparación directa de ambas curvas
<b>Código:</b>	• matrices_confusion.png - Matrices de confusión de ambos modelos
<b>Código:</b>	• stress_analysis.py - Script completo de Python con todo el análisis

*Este reporte fue generado automáticamente utilizando Python, scikit-learn y ReportLab.*