

# Predicción Automática del Estrés con Machine Learning

**Nombre del Proyecto:** R2-D-Data

**Autores:**

Álvaro Ruiz

Fernando Beneytez Maesa

Gabriela Schneider Vannoni

Marcela Motta

Mario Pérez Ramos

Pablo Aguinaga

**Bootcamp:** Big Data, IA & Machine Learning, XIV edición.

Este modelo explora la predicción automática del nivel de estrés durante tareas laborales, empleando técnicas avanzadas de Machine Learning y Deep Learning sobre datos multimodales para ofrecer evaluaciones objetivas del comportamiento bajo presión.





# El Problema del Estrés Laboral

## Impacto Significativo

El estrés laboral afecta la productividad, el bienestar y la toma de decisiones en las organizaciones modernas, especialmente en trabajos del conocimiento con altas exigencias cognitivas.

## Evaluación Subjetiva

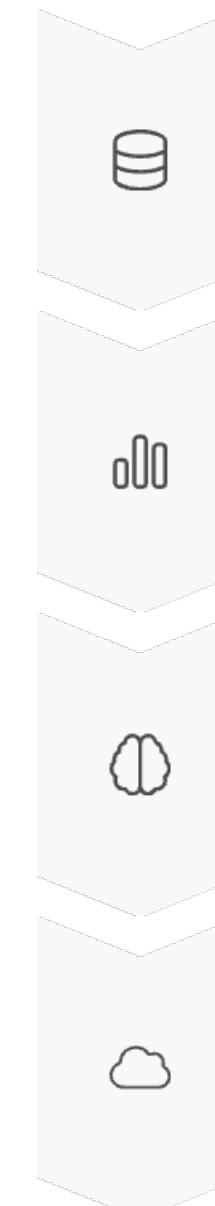
Los métodos tradicionales de evaluación del estrés dependen de instrumentos subjetivos como encuestas o autoevaluaciones, que no siempre reflejan con precisión el estado real del individuo.

## Necesidad de Automatización

Se requieren sistemas objetivos y automatizados que permitan medir el estrés de forma fiable y en tiempo real, aplicables en selección de personal, ergonomía digital o diseño de entornos laborales.



# Objetivos del Proyecto



## Integración de Datos Multimodales

Recopilar, integrar y transformar datos de expresiones faciales, postura, actividad informática y señales fisiológicas del dataset SWELL-KW.

## Análisis Completo

Aplicar limpieza, preprocesamiento, análisis exploratorio y selección de características relevantes para construir modelos efectivos.

## Desarrollo de Modelos Predictivos

Implementar y comparar algoritmos de clasificación tradicionales y redes neuronales para predecir niveles de estrés con precisión.

## Arquitectura Escalable

Diseñar un sistema en Google Cloud que permita alojar el modelo, automatizar tareas y facilitar su despliegue en entornos reales.

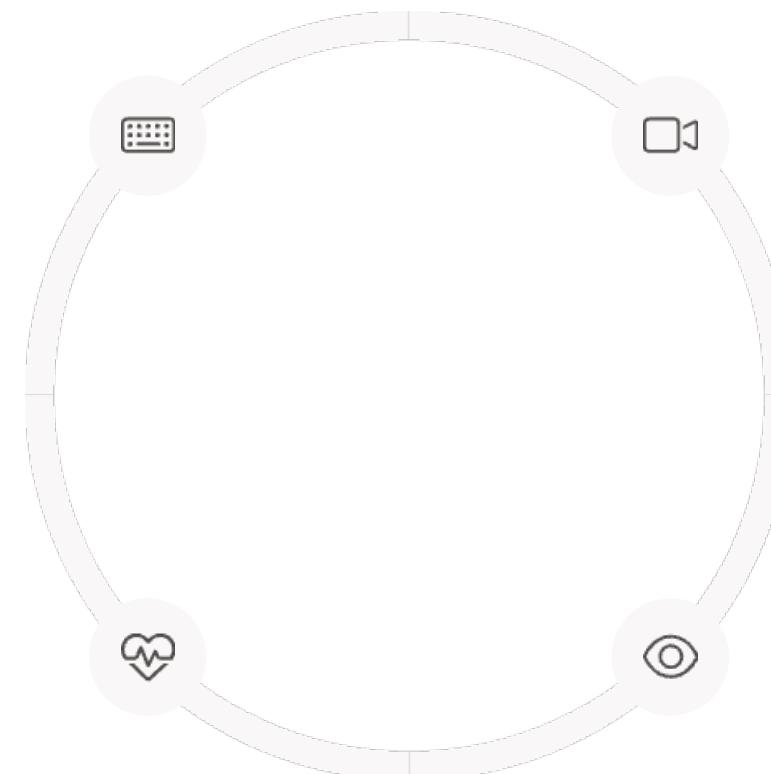
# Dataset SWELL-KW

## A. Interacción Informática

Eventos de teclado, ratón y ventanas activas registrados en formato uLog durante tareas de oficina.

## C. Señales Fisiológicas

Registradas con sensores Mobi, incluyendo frecuencia cardíaca, variabilidad del ritmo cardíaco y conductancia de la piel.



## B. Expresiones Faciales

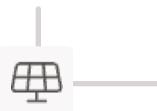
Grabaciones de vídeo procesadas con FaceReader para extraer emociones, atención y valencia.

## D. Posturas Corporales

Captadas con sensor Kinect 3D, analizando la posición de diversas partes del cuerpo durante el trabajo.

El experimento incluyó a 25 participantes realizando tareas habituales de oficina bajo condiciones controladas, con factores de estrés como interrupciones frecuentes y presión de tiempo.

# Metodología de Análisis



## ETL y Limpieza

Integración de datos desde múltiples formatos (.csv, .xlsx, .ods).

Corrección de fechas inválidas, estandarización de metadatos clave y limpieza de valores nulos o anómalos.

Normalización y tipificación de variables numéricas para asegurar coherencia entre fuentes.



## Análisis Exploratorio

Cálculo de estadísticas y visualización de distribuciones.

Ánalysis de correlación por condición experimental.

Detección de valores atípicos y exploración de diferencias según niveles de estrés.



## Preprocesamiento

Codificación de variables categóricas (momento del día, condición).

Transformaciones estadísticas (log, Yeo-Johnson, Box-Cox) para corregir sesgos.

Balanceo de clases y división en train/test para evaluación objetiva.



## Selección de Características

Transformación y reducción de variables con alto sesgo o redundancia.

Evaluación de importancia con modelos como Random Forest, XGBoost y CatBoost.

Selección final manual para mejorar rendimiento e interpretabilidad.



# Modelos Implementados y Resultados

## Regresión Logística

Modelo base entrenado con 100 variables faciales, posturales y de interacción.

Buen rendimiento general con leve sobreajuste.

Nos sirvió como punto de referencia frente a modelos más complejos.

## XGBoost

Modelo de boosting avanzado, eficiente y con alta capacidad de generalización.

Entrenado con 100 variables + momento del día como input categórico.

Rendimiento sobresaliente, error muy bajo y sin sobreajuste. Útil para identificar las variables más influyentes en la predicción del estrés.

## Catboost

Modelo de boosting avanzado, eficaz con variables categóricas y numéricas.

Entrenado con 100 variables de comportamiento + optimización por RMSE.

Rendimiento sólido y sin sobreajuste, ideal para contextos de predicción de estrés.

## Random Forest

Modelo no lineal entrenado con las mismas 100 variables que el modelo base. Mejor generalización y robustez frente al ruido gracias a su estructura de árboles.

Menor sobreajuste y útil para identificar las variables más relevantes en la predicción del estrés.

## Red Neuronal LSTM

Modelo diseñado para capturar la evolución temporal del estrés. Entrenado con 100 variables + momento del día y condición experimental.

Procesó secuencias minuto a minuto para preservar la dinámica del comportamiento.

Rendimiento competitivo y útil para analizar el estrés en contexto temporal.

## Decision Tree

Modelo explicativo y fácil de interpretar para clasificar estrés. Entrenado con 100 variables faciales, posturales y de interacción.

Buena base de partida y útil para entender qué variables influyen más en la predicción.

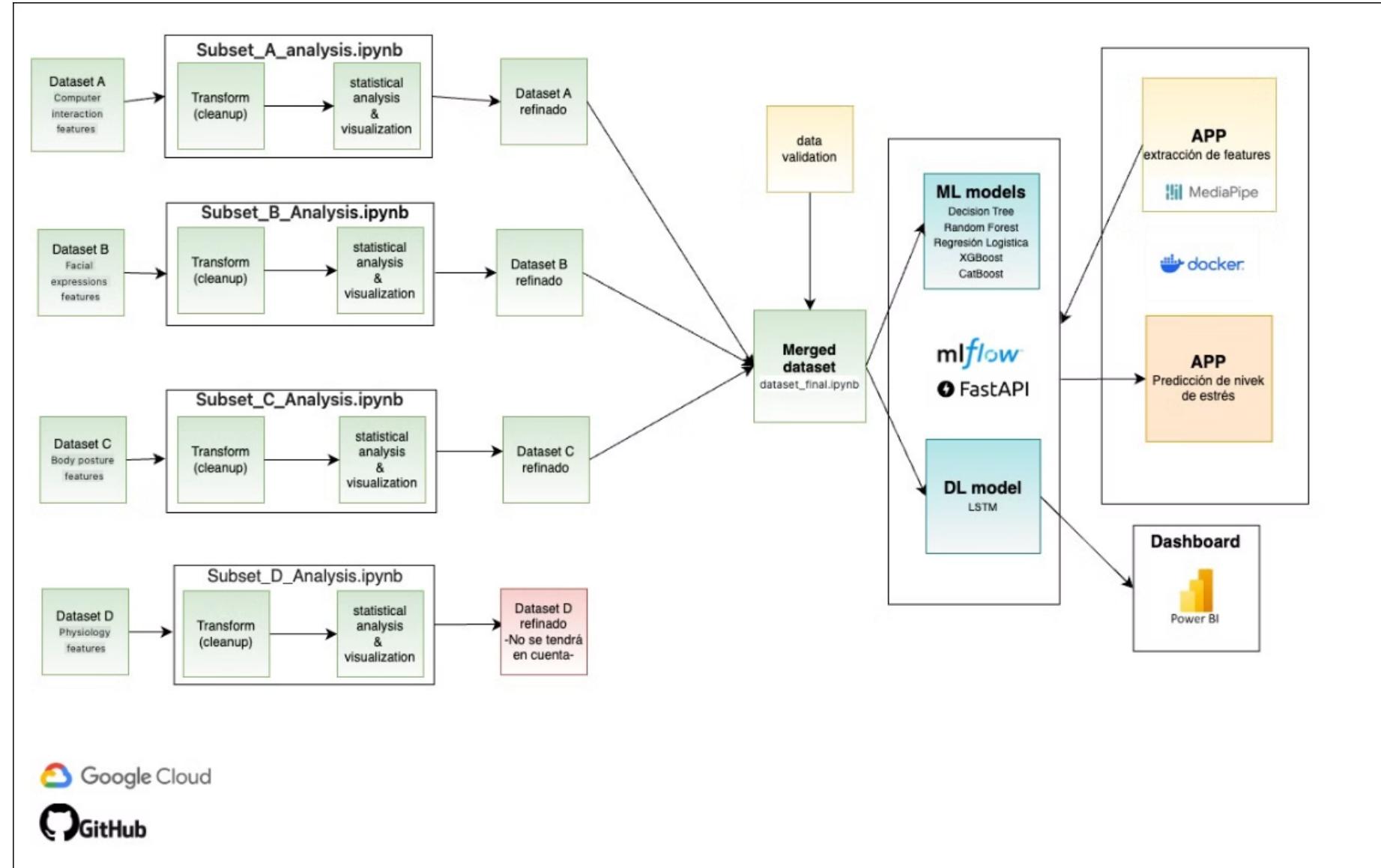
Los modelos fueron evaluados mediante métricas como accuracy, precisión, recall, F1-score y análisis de sobreajuste. Entre ellos, XGBoost y LSTM obtuvieron los mejores resultados globales en la clasificación binaria del estrés (alto vs. bajo), destacando por su capacidad de generalización y precisión. Otros modelos como Random Forest, CatBoost y Decision Tree también ofrecieron un rendimiento sólido, aportando interpretabilidad y robustez al análisis.

# Despliegue del Sistema:

## Arquitectura Cloud

Diseño en Google Cloud Platform:

La arquitectura en Google Cloud incluye almacenamiento en Cloud Storage, funciones en Cloud Run y bases de datos en BigQuery



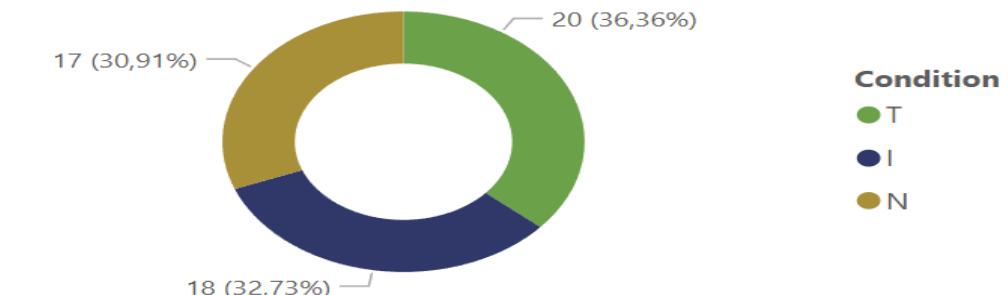
# Despliegue del Sistema (II)

## Visualización

Dashboards interactivos en Power BI para Visualizar

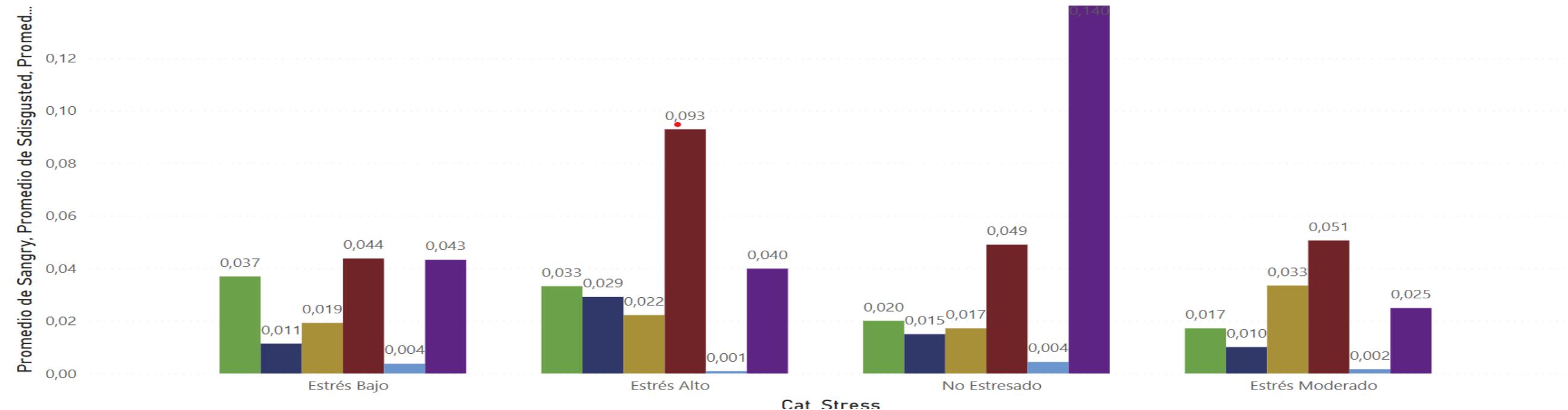
- Niveles de estrés y estados de ánimo asociados a través de diferentes indicadores.
- Comparar modelos predictivos.

Relación entre condición y esfuerzo físico



Análisis entre estados de ánimo y estrés

● Promedio de Sangry ● Promedio de Sdisgusted ● Promedio de Shappy ● Promedio de Ssad ● Promedio de Sscared ● Promedio de Ssurprised



- indicadores
- CONDICION
  - PERSONA
  - MOMENTO DEL DIA
  - NIVEL DE ESTRES

Personas con estrés bajo

0,32

Personas con estrés moderado

0,26

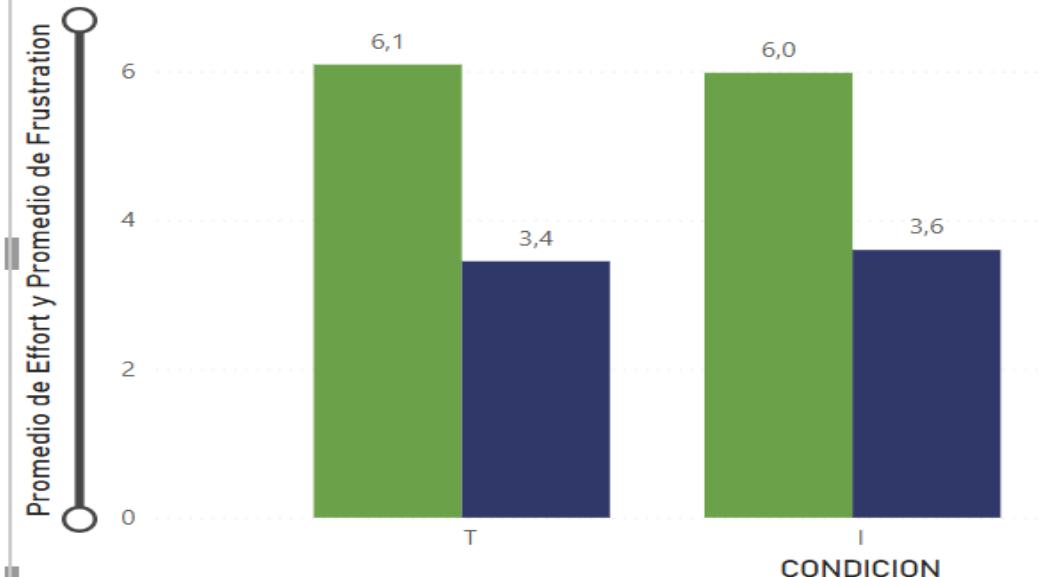
Personas sin estrés

0,36



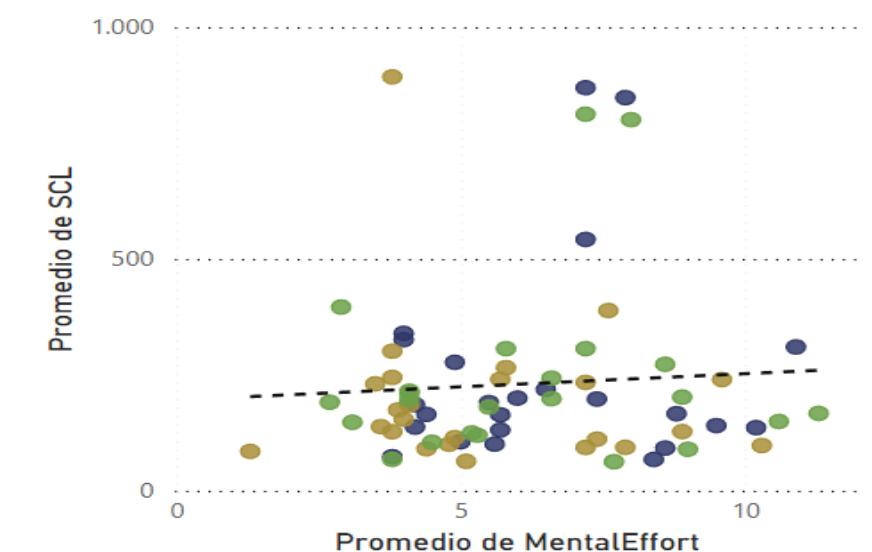
#### Promedio de Effort y Promedio de Frustration por CONDICION

● Promedio de Effort ● Promedio de Frustration



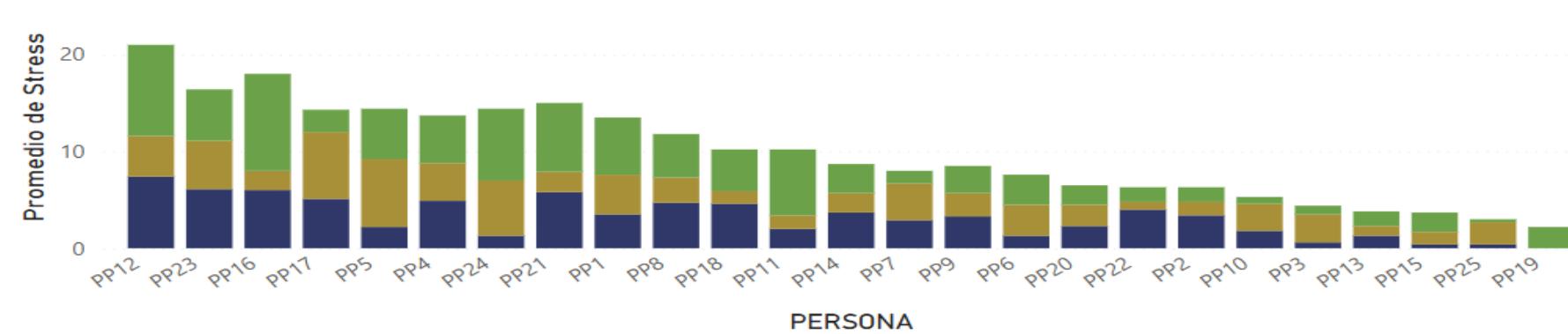
#### Promedio de MentalEffort y Promedio de SCL por PERSONA y CONDICION

CONDICION ● I ● N ● T

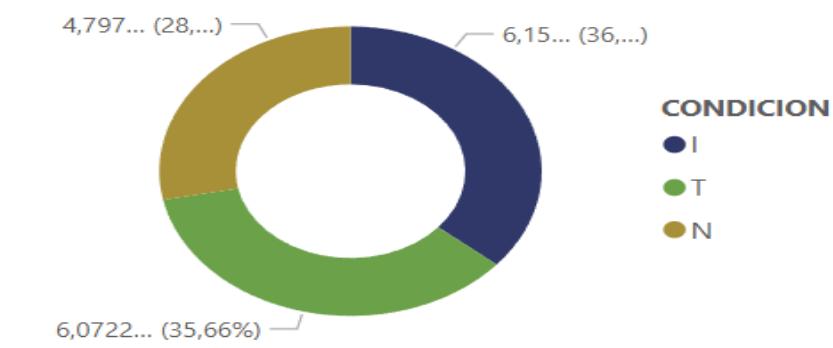


#### Promedio de Stress por PERSONA y CONDICION

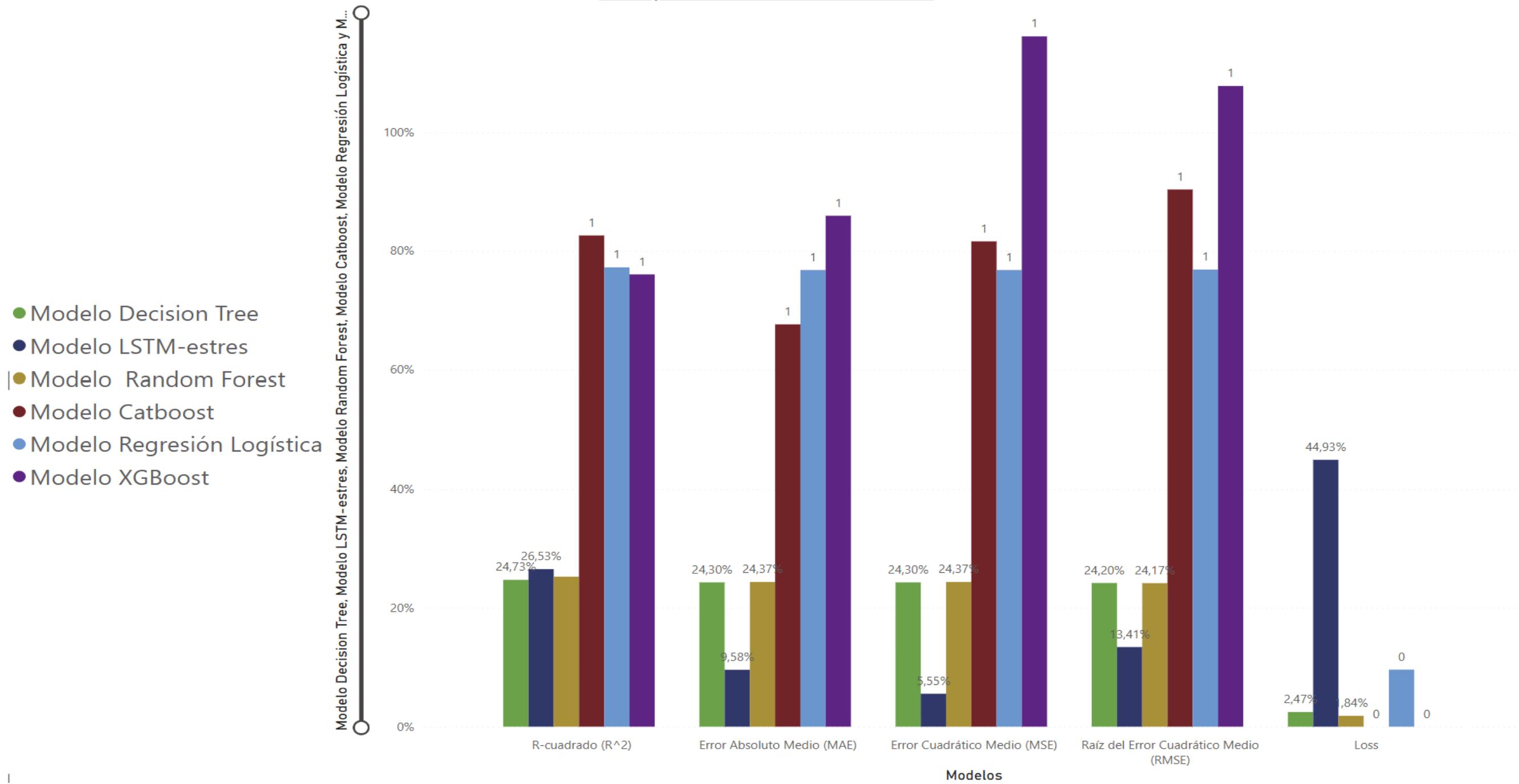
CONDICION ● I ● N ● T



#### Promedio de Performance\_rc por CONDICION



## Comparación entre modelos



# Despliegue del Sistema (III)

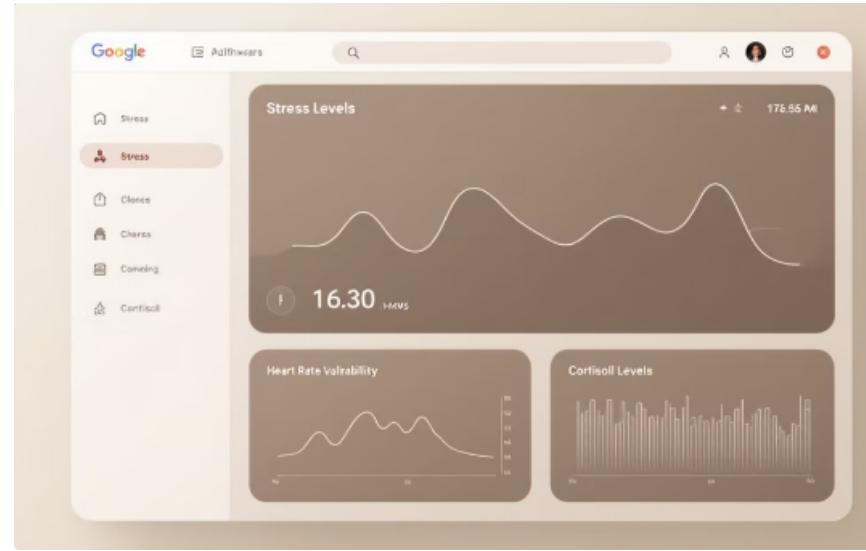
## Desarrollo API

Implementación con FastAPI

El sistema se desplegó mediante una API REST construida con FastAPI, que expone el modelo entrenado y permite realizar predicciones a partir de datos estructurados.

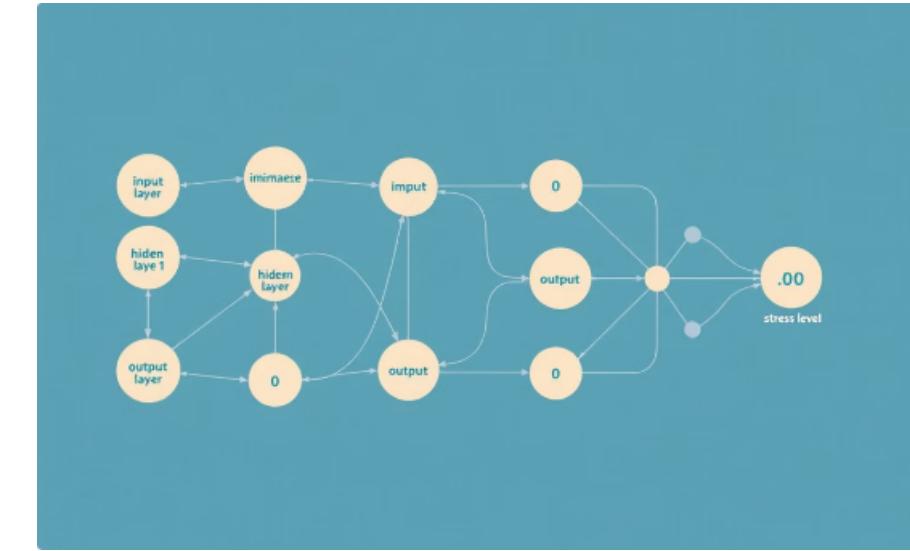


# Conclusiones y Trabajo Futuro



## Monitorización en Tiempo Real

Implementar captura de datos directamente desde cámaras o sensores conectados, permitiendo análisis instantáneo del estrés en entornos laborales reales.

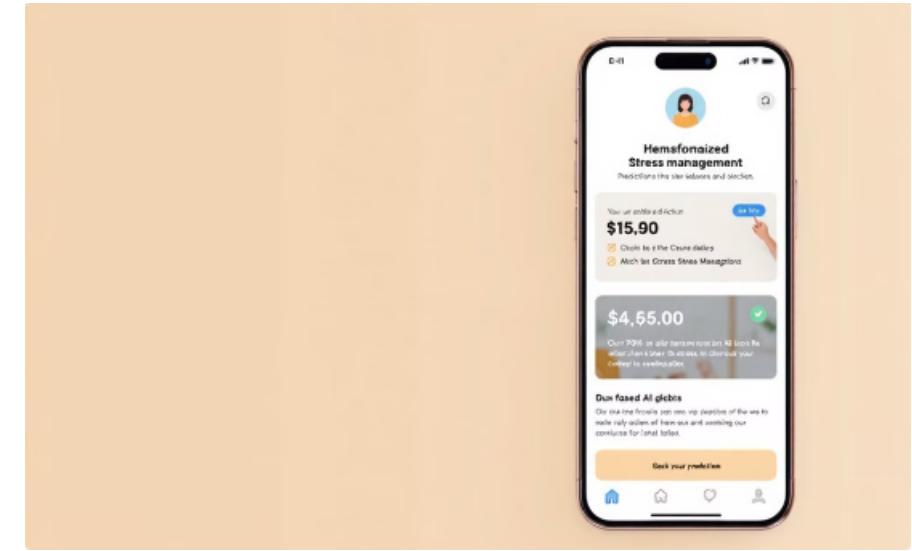


## Refinamiento de Modelos

Aplicar técnicas más avanzadas como redes neuronales convolucionales a vídeo o señales fisiológicas crudas para mejorar la precisión predictiva.

Este proyecto ha demostrado que es posible predecir el nivel de estrés de una persona usando datos de distintas fuentes (como gestos faciales, postura, actividad con el ordenador y contexto).

Los resultados obtenidos con modelos de inteligencia artificial han sido muy positivos, lo que abre la puerta a aplicar este sistema en situaciones reales, como en el entorno laboral, para mejorar el bienestar y la productividad, o como base para nuevas investigaciones en el futuro.



## Aplicaciones Prácticas

Desarrollar interfaces para recursos humanos, ergonomía digital o herramientas de bienestar que aprovechen las predicciones del sistema en contextos reales.