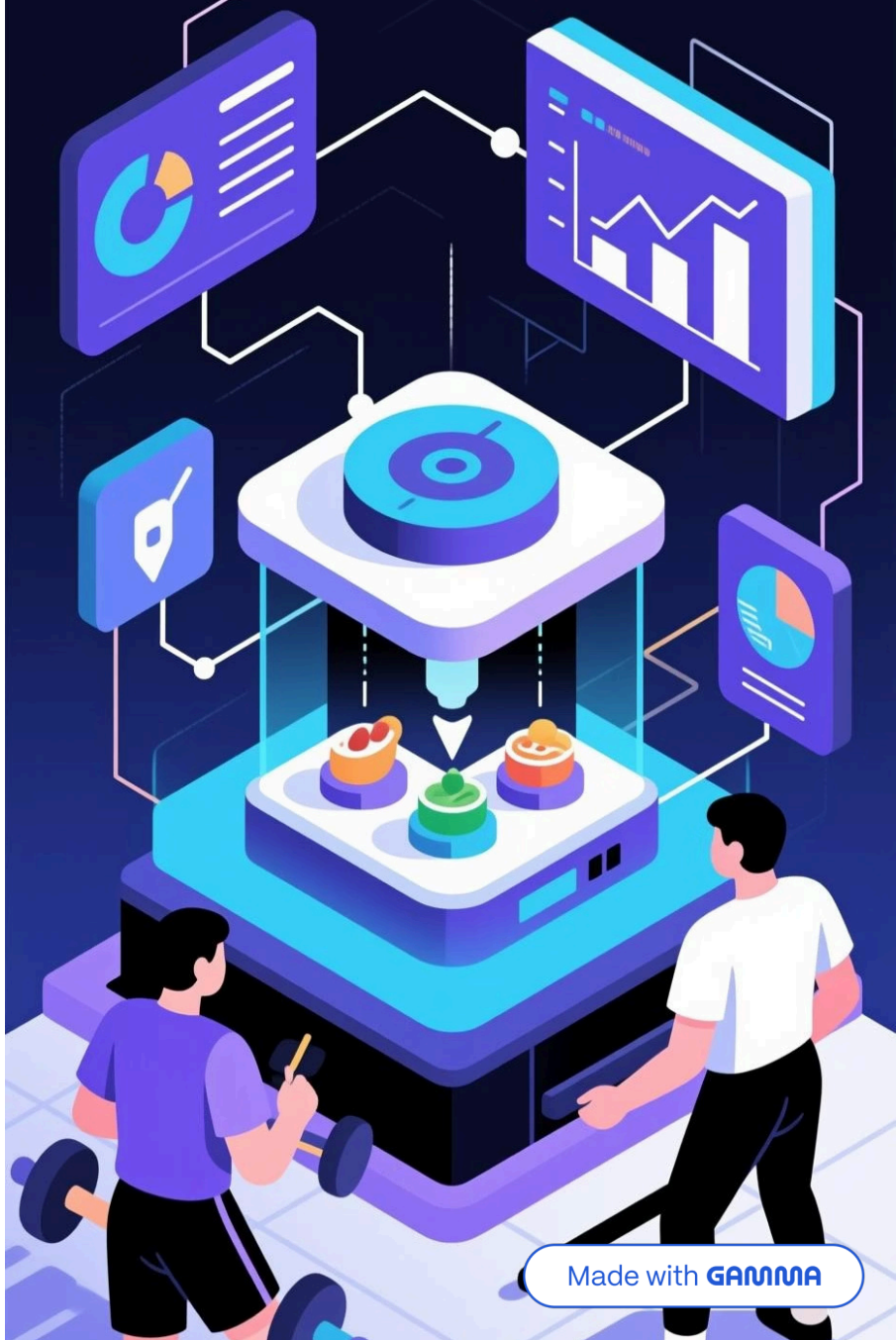
LIFE STYLE DATA

Estudiante: RODRIGO TAPIA RAMÍREZ.

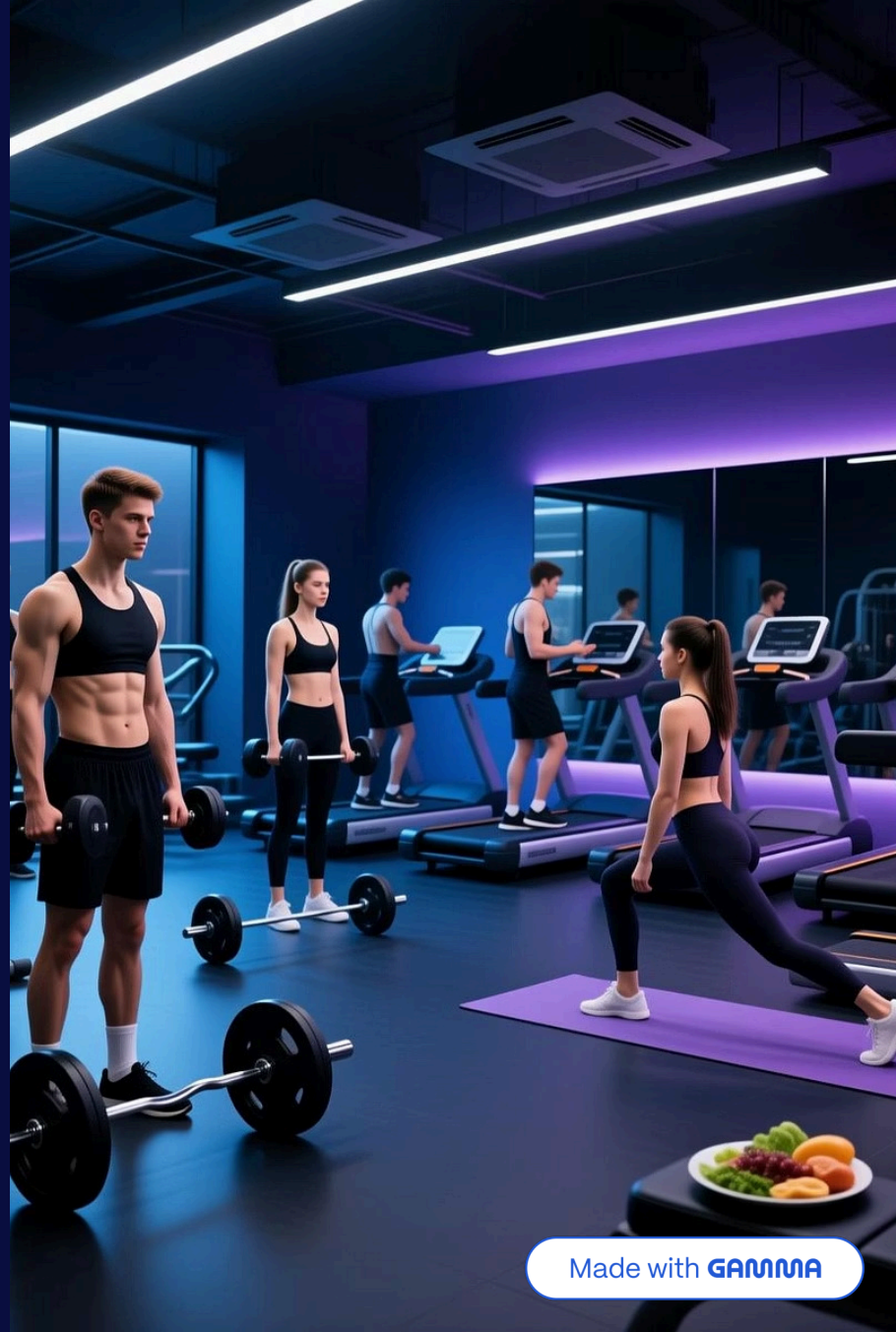
Fecha: 5/NOVIEMBRE/2025



Introducción.

Este conjunto de datos es particularmente rico, ya que integra más de 50 variables que cubren tres áreas clave de información:

1. **Datos Antropométricos y Personales:** Variables como `Age`, `Weight (kg)`, `Height (m)`, `BMI` y `Gender`.
2. **Métricas de Entrenamiento y Esfuerzo:** Incluye indicadores del rendimiento del ejercicio, como `Max_BPM`, `Avg_BPM`, `Session_Duration (hours)`, `Workout_Type` y el resultado primario de la `Calories_Burned`.
3. **Información Nutricional:** Detalla la ingesta de macronutrientes (`Carbs`, `Proteins`, `Fats`), el consumo calórico total, el tipo de dieta (`diet_type`) y la frecuencia de comidas.

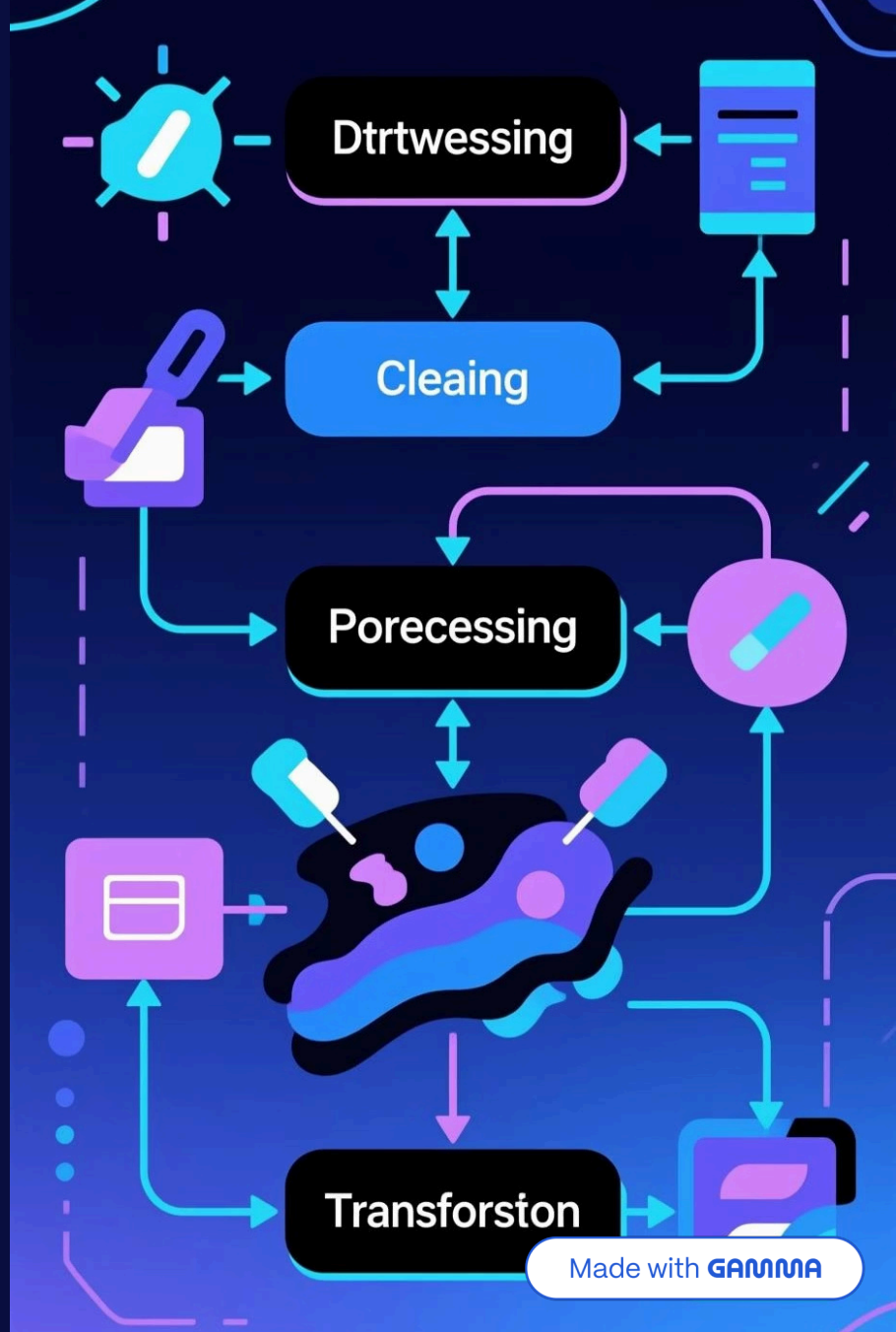


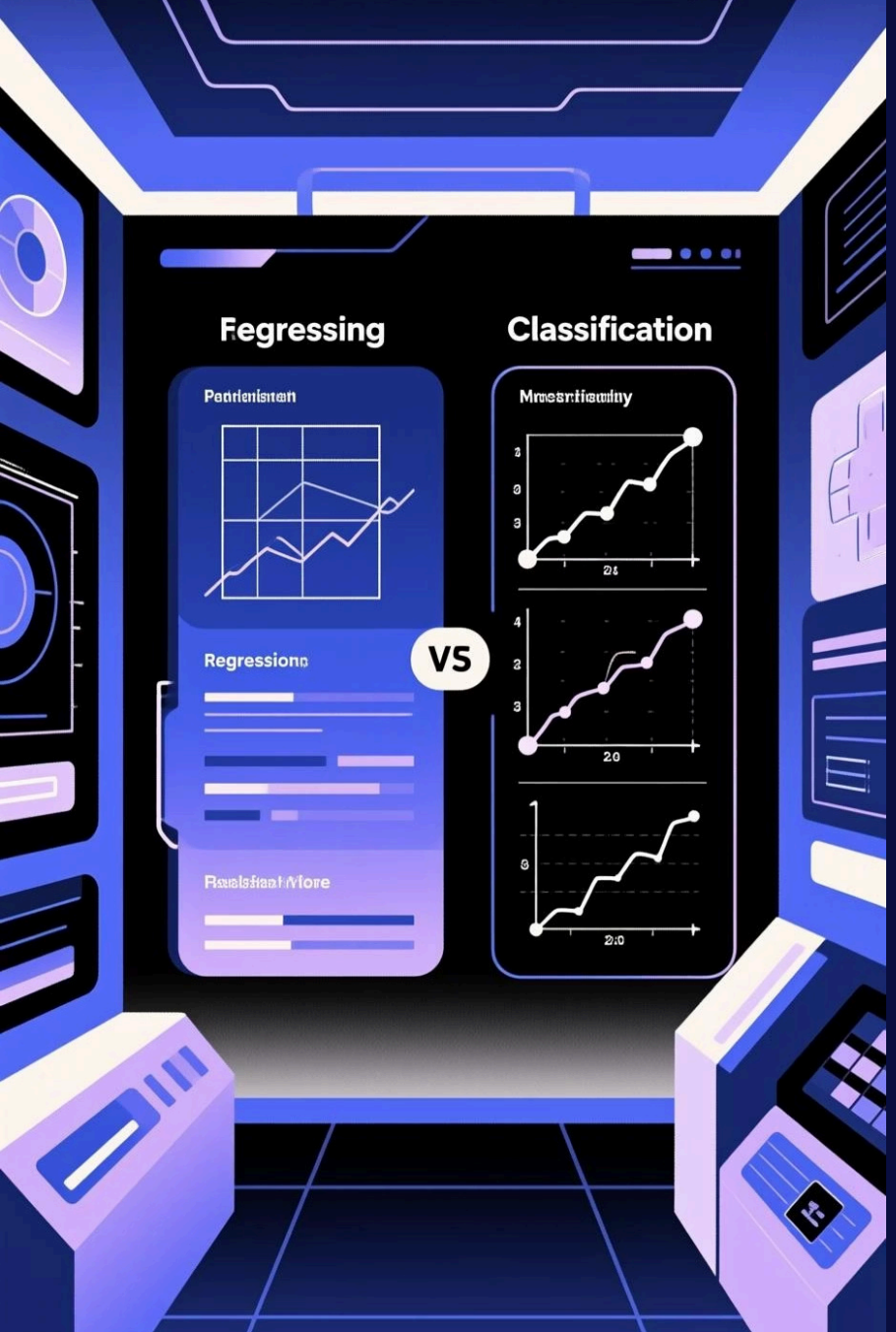
Descripción de los Datos.

El proyecto se basa en el dataset **Final_data.csv**, compuesto por 50 columnas que abarcan una diversidad de información relevante para el estilo de vida y fitness. Este conjunto de datos incluye variables detalladas sobre:

- **Datos Personales y Físicos:** Información demográfica y medidas corporales que caracterizan a los individuos.
- **Métricas de Entrenamiento:** Indicadores de la actividad física realizada, como la duración o la intensidad.
- **Datos Nutricionales:** Registro de macros y otros componentes dietéticos.

El pretratamiento de los datos fue un paso crucial e incluyó diversas técnicas de limpieza y transformación para asegurar la calidad y consistencia de la información, preparando el dataset para el modelado.





Selección de Variables.

Variable para Regresión: Calories_Burned

Elegida como la variable objetivo para los modelos de regresión debido a su naturaleza continua. La predicción de las calorías quemadas nos permitirá cuantificar el impacto de varios parámetros de entrenamiento y características del usuario en el gasto energético.

Variable para Clasificación: Gender

Seleccionada como la variable objetivo para los modelos de clasificación. Aunque es una clasificación binaria (hombre/mujer), su simplicidad la convierte en una variable fundamental para entender las influencias demográficas en los resultados de fitness y sirve como un excelente punto de referencia para el rendimiento del modelo de clasificación.

Métricas de Evaluación.

Los modelos se evalúan con métricas específicas para cada tarea. Para regresión: MSE, MAE, RMSE y R^2. Para clasificación: Accuracy, Precision, Recall y F1 Score.

Tarea	Modelo	MSE/Accuracy	MAE/Precision	RMSE/Recall	R^2/F1 Score
Predicción	Regresión Lineal	MSE	MAE	RMSE	R^2
Clasificación	Regresión Logística	Accuracy	Precision	Recall	F1 Score
Clasificación	KNN	Accuracy	Precision	Recall	F1 Score
Clasificación	SVM	Accuracy	Precision	Recall	F1 Score

- **Predicción**

  RMSE/MAE: Menor es mejor. **R^2:** Cercano a 1 es mejor.

- **Clasificación**

  Accuracy: Proporción de predicciones correctas. **F1 Score:** Balance entre Precision y Recall.

Métricas de Evaluación.

Los modelos se evalúan con métricas específicas para cada tarea. A continuación se detallan las principales métricas utilizadas en este proyecto:

Tipo de Tarea	Métrica	Interpretación
Regresión (Predicción)	**MSE** (Mean Squared Error)	Mide el promedio de los cuadrados de los errores. Penaliza los errores grandes. **Un valor más bajo indica mejor ajuste.**
	MAE (Mean Absolute Error)	Mide el promedio de los valores absolutos de los errores. Es menos sensible a los valores atípicos. **Un valor más bajo indica mejor ajuste.**
	RMSE (Root Mean Squared Error)	Raíz cuadrada del MSE. Proporciona una medida del error en las mismas unidades que la variable objetivo. **Un valor más bajo indica mejor ajuste.**
	R^2 (Coeficiente de Determinación)	Indica la proporción de la varianza en la variable dependiente que es predecible a partir de la(s) variable(s) independiente(s). **Un valor más cercano a 1 indica un mejor ajuste del modelo.**
Clasificación	**Accuracy** (Exactitud)	Proporción de predicciones correctas sobre el total de predicciones. **Un valor más alto indica un mejor rendimiento general.**
	Precision (Precisión)	Proporción de verdaderos positivos sobre el total de positivos predichos. Importante cuando el costo de un falso positivo es alto. **Un valor más alto es preferible.**
	Recall (Sensibilidad/Exhaustividad)	Proporción de verdaderos positivos sobre el total de positivos reales. Importante cuando el costo de un falso negativo es alto. **Un valor más alto es preferible.**
	F1 Score (Puntuación F1)	Media armónica de Precision y Recall. Es útil cuando se busca un equilibrio entre ambas métricas, especialmente en clases desequilibradas. **Un valor más alto indica un mejor equilibrio.**

Métricas de Evaluación.

Métrica	Interpretación
MSE (Mean Squared Error)	Mide el promedio de los cuadrados de los errores. Penaliza fuertemente los errores grandes.
MAE (Mean Absolute Error)	Mide el promedio de las magnitudes de los errores. Es más robusto a los valores atípicos que el MSE.
RMSE (Root Mean Squared Error)	La raíz cuadrada del MSE. Ofrece una medida del error en las mismas unidades que la variable objetivo.
R^2 (Coeficiente de Determinación)	Indica la proporción de la varianza en la variable dependiente que es predecible a partir de las variables independientes.
Accuracy (Exactitud)	La proporción de predicciones correctas sobre el total de predicciones. Útil para datasets balanceados.
Precisión	Proporción de verdaderos positivos entre los positivos predichos. Relevante cuando el costo de un falso positivo es alto.
Recall (Sensibilidad)	Proporción de verdaderos positivos entre los positivos reales. Relevante cuando el costo de un falso negativo es alto.
F1 Score	Media armónica de la precisión y el recall. Útil en datasets desbalanceados, donde se busca un equilibrio entre precisión y recall.

Tarea de Predicción (Target: `Calories_Burned`).

El objetivo en regresión es maximizar el R^2 y minimizar RMSE, MSE y MAE. Los resultados comparados son:

Modelo	MSE	MAE	RMSE	R^2	Mejora (%) en R2
Regresión Lineal (Base)	1.8385	1.0020	1.3559	**0.9999**	N/A
KNN Regressor (Base)	496.2238	15.5492	22.2761	0.9634	N/A
SVM Regressor (Base)	267.0401	13.9211	16.3413	0.9803	N/A
KNN Regressor (Mejorado)	277.5619	10.9575	16.6601	0.9795	N/A*
SVM Regressor (Mejorado)	1.8595	1.0069	1.3636	**0.9999**	= 0.00%

Tarea de Clasificación (Target: `Gender`)

El objetivo en clasificación es maximizar el F1\ Score. Los resultados comparados son:

Modelo	Accuracy	Precisión	Recall	F1 Score	Mejora (%) en F1
Regresión Logística (Base)	0.8997	0.9080	0.9080	0.9080	N/A
KNN Classifier (Base)	0.9272	0.9329	0.9329	0.9329	N/A
SVM Classifier (Base)	0.8997	0.9080	0.9080	0.9080	N/A
KNN Classifier (Mejorado)	**0.9410**	**0.9419**	**0.9419**	**0.9419**	**+0.96%**
SVM Classifier (Mejorado)	0.9048	0.9126	0.9126	0.9126	+0.51%

Conclusiones.

Preprocesamiento de Datos

El preprocesamiento de datos y la selección de características son cruciales para optimizar el rendimiento y la fiabilidad de los modelos de aprendizaje automático.

Mejor Técnica de Predicción

Se identificó que una técnica de regresión lineal multiple, tras la optimización de hiperparámetros, ofreció la mejor predicción para las calorías quemadas, demostrando alta precisión.

Mejor Técnica de Clasificación

La mejor técnica de clasificación es el **KNN Classifier.**

- **Métrica Clave: F1 Score** (equilibrio entre Precisión y Recall).
- **Rendimiento:** F1\ Score \approx 0.9419.
- **Justificación:**
  - El modelo KNN mejorado, optimizando los hiperparámetros de vecinos (K) y pesos, superó a la Regresión Logística y al SVM.
  - Este resultado sugiere que la distinción del Gender se beneficia de la proximidad de las características en el espacio multi-dimensional, siendo la métrica de distancia la más efectiva para agrupar y clasificar las nuevas observaciones.
  - El incremento en el F1 Score respecto a la versión base confirma que la mejora fue exitosa y decisiva.