

Pronóstico de Variables Deportivas Utilizando Modelos Supervisados

Sergio Cortés Cepeda

matricula:1731225

Universidad Autónoma de Nuevo León, Facultad de Ciencias Físico Matemáticas

Noviembre 2025

Resumen

Este estudio aplica algoritmos supervisados para predecir dos variables deportivas obtenidas desde datos reales del reloj Garmin: las calorías quemadas y el ritmo promedio de carrera. Se emplearon tres modelos de regresión —**Regresión Lineal**, **Random Forest** y **KNN**— evaluados mediante métricas estándar como **MAE**, **MSE** y **RMSE**. Los resultados muestran que la **Regresión Lineal** es superior para pronosticar calorías, mientras que **Random Forest** obtiene el mejor desempeño al predecir el ritmo promedio.

1. Introducción

Los modelos supervisados permiten estimar una variable objetivo a partir de un conjunto de predictores. Estos algoritmos pueden utilizarse para estimar rendimiento, esfuerzo y gasto energético. El propósito de esta tarea es construir modelos capaces de predecir:

- Calorías quemadas por sesión.
- Ritmo promedio (segundos por kilómetro).

Para ello se analizaron los datos reales provenientes del reloj Garmin, que describen múltiples sesiones de mis entrenamientos, incluyendo métricas fisiológicas y de desempeño.

2. Metodología

2.1. Descripción del conjunto de datos

El conjunto de datos contiene información de cada entrenamiento:

- Distancia recorrida (`distancia_km`)
- Tiempo total en segundos (`Tiempo_Total`)
- Calorías (`calorias`)
- Frecuencia cardiaca media y máxima (`fc_media`, `fc_max`)
- Cadencia media (`cadencia_media`)
- Zancada promedio (`zancada_m`)
- Ritmo promedio (`ritmo_medio`)

2.2. Variables utilizadas

Variables predictoras (**X**):

- Distancia
- Tiempo total
- Frecuencia cardiaca media
- Frecuencia cardiaca máxima
- Cadencia
- Zancada

Variables objetivo (**y**):

- Calorías
- Ritmo promedio

3. Modelos Supervisados

Se emplearon tres algoritmos:

3.1. Regresión Lineal

Modelo paramétrico basado en combinación lineal de predictores. Es interpretable y funciona bien cuando la relación aproximada es lineal.

3.2. Random Forest Regressor

Modelo de ensamble basado en múltiples árboles. Captura relaciones no lineales y es robusto a valores atípicos.

3.3. KNN Regressor

Predice valores en función de los k vecinos más cercanos. Es sensible al escalado y puede fallar si los datos tienen alta varianza.

4. Métricas Utilizadas

4.1. MAE

$$MAE = \frac{1}{n} \sum |y_i - \hat{y}_i|$$

4.2. MSE

$$MSE = \frac{1}{n} \sum (y_i - \hat{y}_i)^2$$

4.3. RMSE

$$RMSE = \sqrt{MSE}$$

Estas métricas se seleccionaron por ser las más comunes en regresión y por permitir evaluar precisión (MAE), penalización de errores grandes (MSE) y magnitud global del error (RMSE).

5. Resultados

5.1. Predicción de Calorías

Cuadro 1: Resultados para la predicción de calorías.

Modelo	MAE	MSE	RMSE
Regresión Lineal	17.16	519.70	22.80
Random Forest	25.74	1062.52	32.60
KNN	36.88	1888.47	43.46



Figura 1: Histograma de errores absolutos — Calorías (Regresión Lineal).

5.2. Predicción de Ritmo Promedio

Cuadro 2: Resultados para la predicción de ritmo promedio.

Modelo	MAE	MSE	RMSE
Regresión Lineal	38.09	2174.02	46.62
Random Forest	30.80	1543.58	39.29
KNN	80.52	20042.28	141.57

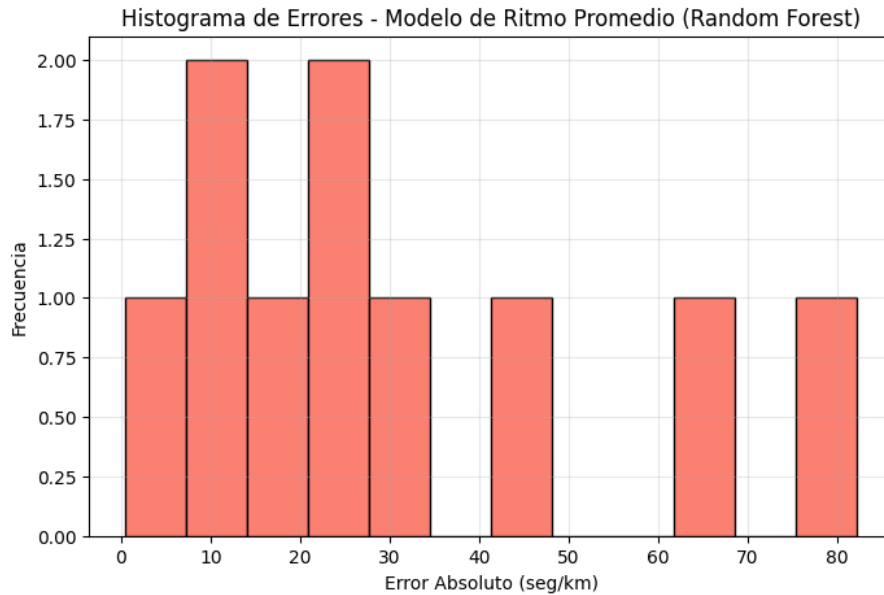


Figura 2: Histograma de errores absolutos — Ritmo Promedio (Random Forest).

6. Discusión

Los resultados muestran que:

- La **Regresión Lineal** funciona mejor para predecir calorías debido a su relación lineal con distancia, tiempo y frecuencia cardíaca.
- El **Random Forest** es superior prediciendo ritmo promedio, lo cual indica relaciones no lineales entre cadencia, FC y zancada.
- **KNN** obtuvo el peor desempeño en ambas tareas, probablemente por su sensibilidad al ruido y al tamaño reducido del dataset.

Los histogramas muestran que existen valores atípicos que afectan el MSE y RMSE, especialmente en el ritmo promedio.

7. Conclusiones

- Los modelos supervisados permiten realizar pronósticos útiles sobre rendimiento usando datos reales del garmin.
- La Regresión Lineal es la mejor opción para estimar calorías quemadas.
- Random Forest ofrece la predicción más precisa para el ritmo promedio.
- Para mejorar resultados futuros, sería útil incorporar variables externas como temperatura, tipo de superficie o fatiga acumulada.

Referencias

- [1] GeeksforGeeks. (2024). *ML — Linear Regression*. Recuperado de <https://www.geeksforgeeks.org/machine-learning/ml-linear-regression/>
- [2] DataCamp. (2023). *Random Forests in Python*. Recuperado de <https://www.datacamp.com/tutorial/random-forests-classifier-python>
- [3] Proclus Academy. (2023). *Regression Metrics You Must Know*. Recuperado de <https://proclusacademy.com/blog/explainer/regression-metrics-you-must-know/>