**Justificación: Regresión Lineal**

**1. Naturaleza del Problema**

La regresión lineal es un modelo supervisado de aprendizaje automático ideal para **predicción de valores continuos**, como en este caso, donde el objetivo es predecir **"Calories\_Burned"** (calorías quemadas). Esta variable dependiente es continua y cuantitativa, lo cual hace que la regresión lineal sea un enfoque natural para resolver el problema planteado.

El dataset proporciona múltiples variables independientes (predictoras) como **"Age"**, **"Weight"**, **"Max\_BPM"**, **"Session\_Duration"**, entre otras. Estas características tienen una relación plausible con la cantidad de calorías quemadas, ya que factores fisiológicos (como el peso, altura y frecuencia cardiaca) y patrones de ejercicio (duración y tipo de entrenamiento) afectan directamente el gasto energético de una persona.

**2. Ventajas de la Regresión Lineal**

La regresión lineal presenta varias ventajas que la convierten en una herramienta adecuada para el problema propuesto:

* **Interpretabilidad:** La regresión lineal es uno de los algoritmos más interpretables. Los coeficientes obtenidos en el modelo pueden mostrar la influencia de cada variable en la predicción de calorías.
* **Simplicidad:** Es un método sencillo, eficiente y rápido, especialmente con un dataset de tamaño moderado como el proporcionado (973 muestras).
* **Relaciones lineales:** Aunque el ejercicio y el gasto calórico pueden tener cierta no linealidad, una regresión lineal es una excelente aproximación inicial. Las características como **"Session\_Duration"** y **"Max\_BPM"** probablemente tengan una relación lineal significativa con **"Calories\_Burned"**.
* **Requisitos de preprocesamiento:** La regresión lineal es robusta siempre que las variables estén correctamente escaladas y no exista multicolinealidad severa.

**3. Evaluación de la Relación entre Variables**

Para justificar más formalmente el uso de regresión lineal, se puede realizar un **análisis exploratorio de datos (EDA)** que incluya:

* Análisis de correlación entre las variables predictoras y la variable objetivo (**Calories\_Burned**).
* Visualización de gráficos de dispersión para variables clave como **"Session\_Duration"** y **"Max\_BPM"** contra las calorías.
* Validación de supuestos básicos de regresión lineal como la homocedasticidad y normalidad de residuos.

De forma intuitiva:

* **"Session\_Duration"** (duración de la sesión) y **"Max\_BPM"** (frecuencia cardiaca máxima) son características clave, ya que una mayor duración y esfuerzo físico típicamente resultan en un mayor consumo calórico.
* **"Weight"** y **"Fat\_Percentage"** también influyen, pues personas con mayor masa corporal tienden a gastar más energía.

**4. Impacto del Preprocesamiento**

El preprocesamiento de datos, aunque no indispensable para el funcionamiento básico de la regresión lineal, puede mejorar significativamente el modelo:

* **Normalización/Estandarización:** Alinear la escala de variables como peso, altura y BPM facilita la interpretación y el ajuste del modelo.
* **Imputación de datos faltantes:** Asegura que no haya pérdida de información en el entrenamiento.
* **Transformación de variables categóricas:** El uso de **"Workout\_Type"** y **"Gender"** requiere codificación (e.g., One-Hot Encoding) para ser utilizadas en un modelo lineal.

**5. Fundamentación Teórica y Referencias**

La regresión lineal es un modelo clásico y fundamental en problemas de predicción de valores continuos en el campo de **Machine Learning** y estadística. Su uso en problemas relacionados con la salud y ejercicio está ampliamente documentado:

* **Raschka, S., & Mirjalili, V. (2019).** *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2.* Packt Publishing. ISBN: 978-1789955750.
* **Seber, G. A. F., & Lee, A. J. (2012).** *Linear Regression Analysis.* Wiley. DOI: 10.1002/9781118274446.
* **Ruppert, D. (2004).** *Statistics and Data Analysis for Financial Engineering.* Springer. DOI: 10.1007/978-1-4757-5215-7.

**Conclusión**

La regresión lineal es un modelo idóneo para predecir la variable **"Calories\_Burned"** debido a la naturaleza continua de los datos y su relación intuitiva con las características independientes. Además, su simplicidad, interpretabilidad y bajo costo computacional la convierten en una opción robusta y eficaz para este problema.