Propuesta Integral para la Monitorización Deportiva

Desarrollo de Modelos de ML para:   
1. Clustering de Perfiles de Rendimiento Técnico  
2. Carga Semanal Recomendada  
3. Predicción del Riesgo de Lesión

# I. Modelo 1: Clustering de Perfiles de Rendimiento Técnico

## 1. Objetivo y Justificación

Objetivo:  
 - Agrupar deportistas según su comportamiento competitivo y estado fisiológico.   
 - Identificar perfiles: A (Óptimos), B (Sobrecargados), C (Inestables), D (Bajo estímulo).  
  
Justificación:  
 - Uso combinado de variables de bienestar (Sueño, Fatiga, Estrés, Dolor, Molestia) e indicadores de esfuerzo (PSE, Tiempo), con información contextual (Modalidad, Deporte) para personalizar entrenamientos y prevenir sobrecargas.

## 2. Exploración y Preprocesamiento

Fuente: Archivo 'df\_final.csv' con >31,000 registros.  
Variables clave: Bienestar, Esfuerzo, Contexto.  
Acciones: Imputación de valores nulos (media/mediana o forward-fill por deportista), codificación de 'Molestia', y normalización con StandardScaler.

## 3. Feature Engineering

Construcción del vector de características combinando variables normalizadas y, si es necesario, ponderándolas según su relevancia.

## 4. Desarrollo del Modelo y Validación

- Algoritmo: KMeans (inicialmente 4 clústeres).  
 - Validación: Método del codo y Silhouette Score.  
 - Visualización: Proyección 2D (PCA o t-SNE) para interpretar los clusters.

## 5. Resultados y Aplicación

Interpretación de perfiles para personalizar entrenamientos, emitir alertas y generar informes para el cuerpo técnico y médico.

# II. Modelo 2: Carga Semanal Recomendada

## 1. Objetivo y Justificación

Objetivo:  
 - Estimar la carga semanal ideal para cada atleta, basándose en indicadores de esfuerzo, bienestar y cumplimiento.  
  
Justificación:  
 - Variables de esfuerzo (PSE y Tiempo) combinadas con indicadores de bienestar (Sueño, Dolor, Estrés, Fatiga, Molestia) y cumplimiento (Tipo\_enc) permiten modelar matemáticamente la carga ideal para prevenir sobrecargas.

## 2. Exploración, Preprocesamiento y Feature Engineering

Fuente: 'df\_final.csv'  
Acciones: Imputación de valores nulos, codificación, normalización y creación de la columna 'Semana'.  
Cálculos:  
 - Carga = PSE x Tiempo.  
 - Carga\_4s = Promedio móvil de las 4 semanas anteriores.

## 3. Desarrollo del Modelo Predictivo y Validación

Algoritmos:  
 - Linear Regression, Lasso Regression, XGBoostRegressor.  
Validación: Uso de TimeSeriesSplit con métricas MSE, MAE y R².

## 4. Resultados y Aplicaciones

Predicción de la carga ideal semanal por atleta, permitiendo comparar con la carga real, ajustar la planificación y generar dashboards interactivos.

# III. Modelo 3: Predicción del Riesgo de Lesión

## 1. Objetivo y Justificación

Objetivo:  
 - Clasificar el riesgo de lesión en 3 niveles: 0 (Bajo), 1 (Moderado) y 2 (Alto), basado en variables de carga y estado físico.  
  
Justificación:  
 - El desequilibrio entre la carga y la recuperación incrementa el riesgo de lesión. Variables clave: ACWR, Dolor, Molestia, Sueño y Días consecutivos.

## 2. Exploración, Preprocesamiento y Feature Engineering

Fuente: 'df\_final.csv'  
Acciones: Imputación, codificación y normalización.  
Cálculos:  
 - Carga = PSE x Tiempo.  
 - ACWR = (Carga de la semana actual) / (Carga\_4s).  
 - Días\_consecutivos: Número de días seguidos de entrenamiento a partir de la fecha.

## 3. Desarrollo del Modelo y Manejo del Desbalance

Algoritmos:  
 - RandomForestClassifier y XGBoostClassifier.  
Balanceo: Aplicación de SMOTE para corregir el desbalance en las clases.  
Validación: f1\_score, recall y matriz de confusión, con especial atención en la detección del alto riesgo.

## 4. Resultados y Aplicaciones

El modelo etiquetará cada atleta con un riesgo (0, 1 o 2) permitiendo generar alertas automáticas y reportes diarios, integrables en dashboards interactivos para el seguimiento en tiempo real.

# IV. Integración y Despliegue

Cada modelo incluye su propio proceso de extracción, preprocesamiento, feature engineering, modelado y validación, adaptado a sus objetivos específicos:  
  
 - Modelo 1: Clustering de perfiles para personalización del entrenamiento.  
 - Modelo 2: Predicción de la Carga Semanal Recomendada para planear la carga de cada atleta.  
 - Modelo 3: Clasificación del Riesgo de Lesión para intervenciones preventivas.  
  
Se recomienda desarrollar pipelines modulares y mantener la actualización y monitorización continua de cada modelo.

# V. Cronograma y Recomendaciones Finales

Fase 1 (2-3 semanas): Exploración y preprocesamiento de datos para cada modelo.  
Fase 2 (1-2 semanas): Feature engineering específico (Cálculo de Carga, ACWR, Días consecutivos, columna 'Semana').  
Fase 3 (2-3 semanas): Desarrollo y ajuste de cada modelo (KMeans, regresores y clasificadores), con validación particular.  
Fase 4 (1-2 semanas): Integración en dashboards interactivos y generación de reportes/alertas.  
Fase 5 (Continuo): Despliegue, monitoreo y actualización periódica.  
  
Recomendaciones:  
 - Validar continuamente con el equipo técnico/médico.  
 - Mantener flujos independientes para facilitar ajustes de cada modelo.  
 - Desarrollar dashboards que integren la información en tiempo real.

# VI. Conclusiones

La propuesta integral detalla tres caminos de desarrollo independientes para abordar diferentes necesidades en la monitorización deportiva:   
1. Clustering de Perfiles de Rendimiento Técnico para personalizar entrenamientos.  
2. Carga Semanal Recomendada para definir la planificación óptima de entrenamiento.  
3. Predicción del Riesgo de Lesión para implementar intervenciones preventivas.  
  
Se aconseja trabajar con pipelines modulares, garantizar la actualización continua y validar cada modelo en conjunto con profesionales del área para lograr la máxima eficiencia en la toma de decisiones.