Universidad Icesi - Algoritmos y Programación III
Mariana Agudelo Salazar y Natalia Vargas

Proyecto final - Algoritmos y Programación III
Caso de Estudio: Sistema de Anotación de video
Entrega 1

Contexto

El objetivo central del proyecto es desarrollar una herramienta de software capaz de analizar actividades específicas de una persona (caminar hacia la cámara, caminar hacia atrás, voltear, sentarse y levantarse), realizar un seguimiento de los movimientos articulares y posturales y clasificarlos apropiadamente en tiempo real, esto con el objetivo de ayudar a personas mayores de edad con ejercicios de refuerzo de movilidad.

Aspectos Éticos

Para la recolección de los datos utilizados para el entrenamiento de los modelos es esencial garantizar la privacidad y seguridad de los datos personales de los involucrados, además de tener consentimiento explícito y comunicar los objetivos y propósitos del proyecto para el cual prestarán su imagen. El modelo a desarrollar debe tener una amplia muestra para evitar el sesgo y exclusión de personas. El principio de transparencia también debe ser eje guía para el desarrollo del proyecto, mostrar los resultados reales, los datos y explicar el funcionamiento del sistema con claridad son elementos que los desarrolladores deben tener en cuenta en todo momento.

Preguntas de interés

- ¿Cuál es la mejor forma de representar el movimiento (ángulos o trayectorias) para lograr una clasificación precisa de las actividades y posturas?
- ¿Cómo influye la variación de perspectiva o velocidad del movimiento en los resultados del modelo de clasificación?

Tipo de Problema

El proyecto trata un problema de Clasificación Multiclase Supervisada, busca identificar el movimiento de una persona en base a secuencias de movimiento extraídas de videos.

Metodología

Se utilizará como metodología central CRISP-DM (Cross Industry Standard Process for Data Mining) la cual consiste en 6 etapas:

Fase CRISP-DM	Descripción del trabajo realizado
I. Comprensión empresarial	Definición de los objetivos del proyecto, establecimiento de las etapas de trabajo y delimitación de las cinco clases de movimiento utilizadas (*caminar-adelante, caminar-atrás, girar, sentarse, pararse*) para la obtención y análisis de datos.
II. Comprensión de datos	Establecimiento de parámetros estándar para la captura de videos, recopilación de secuencias con las acciones requeridas y análisis preliminar de los datos recolectados para verificar su

	consistencia y calidad.
III. Preparación de datos	Normalización y limpieza del conjunto de datos, extracción de landmarks corporales mediante *MediaPipe Pose* y selección de herramientas de anotación (Label Studio / CVAT) para segmentar las acciones.
IV. Modelado	Selección de los modelos supervisados a implementar (Random Forest y XGBoost), entrenamiento inicial de los algoritmos y ajuste de hiperparámetros para optimizar el desempeño.
V. Evaluación	Comparación del rendimiento de los modelos utilizando métricas estándar (Accuracy, Precision, Recall, F1-Score), análisis de relevancia de características y verificación del cumplimiento de los objetivos planteados.
VI. Despliegue	Implementación del sistema de anotación en tiempo real, desarrollo de una interfaz visual que muestre los resultados de manera clara y documentación del proceso de uso.

Métricas que usaremos para medir el progreso

- **Accuracy (Exactitud):** Porcentaje total de predicciones correctas sobre el total de las muestras evaluadas
- **Precision (Precisión):** Proporción de las predicciones clasificadas como positivas fueron realmente correctas.
- **Recall (Sensibilidad):** Proporción de los casos positivos reales identificados correctamente por los modelos.
- **F1-Score:** Precision + Recall, promedio para equilibrar ambos valores.

Datos Recolectados

Para la toma de datos se utilizó un mismo dispositivo móvil y se establecieron estándares a seguir para minimizar el ruido entre los datos obtenidos.
La grabación de los videos iniciales se realizó a una misma luz artificial con un trípode para evitar el movimiento innecesario de la cámara, a una distancia que permitiera el enfoque de cuerpo completo.
Se grabaron 18 individuos realizando los 5 movimientos a utilizar, luego se cortaron los videos en edición para clasificarlos en carpetas nombradas con los movimientos (caminar-adelante, caminar-atras, girar, pararse, sentarse) para simplificar el proceso de análisis.

Análisis Exploratorio de los Datos

El análisis exploratorio se realizó sobre los dos conjuntos de datos generados: datos_analisis.csv, que contiene metadatos de los videos, y datosmediapipe.csv, con los 33 puntos de referencia corporales (landmarks) por fotograma. El propósito de este análisis fue **evaluar la calidad, consistencia y balance del conjunto de datos**, así como **identificar patrones estructurales** que permitan anticipar el desempeño de los modelos de clasificación multiclase que se desarrollarán en etapas posteriores.

1. **Análisis de Metadatos de Video (datos_analisis.csv)**
   a. **Balance de Clases:**

   El conjunto de datos presenta un balance **perfecto** en cuanto al número de videos por clase (18 por cada una de las 5 acciones: *caminar-adelante, caminar-atras, girar, sentarse, pararse*).
   Este balance garantiza una base sólida para el entrenamiento de modelos supervisados, evitando sesgos de representación entre categorías.

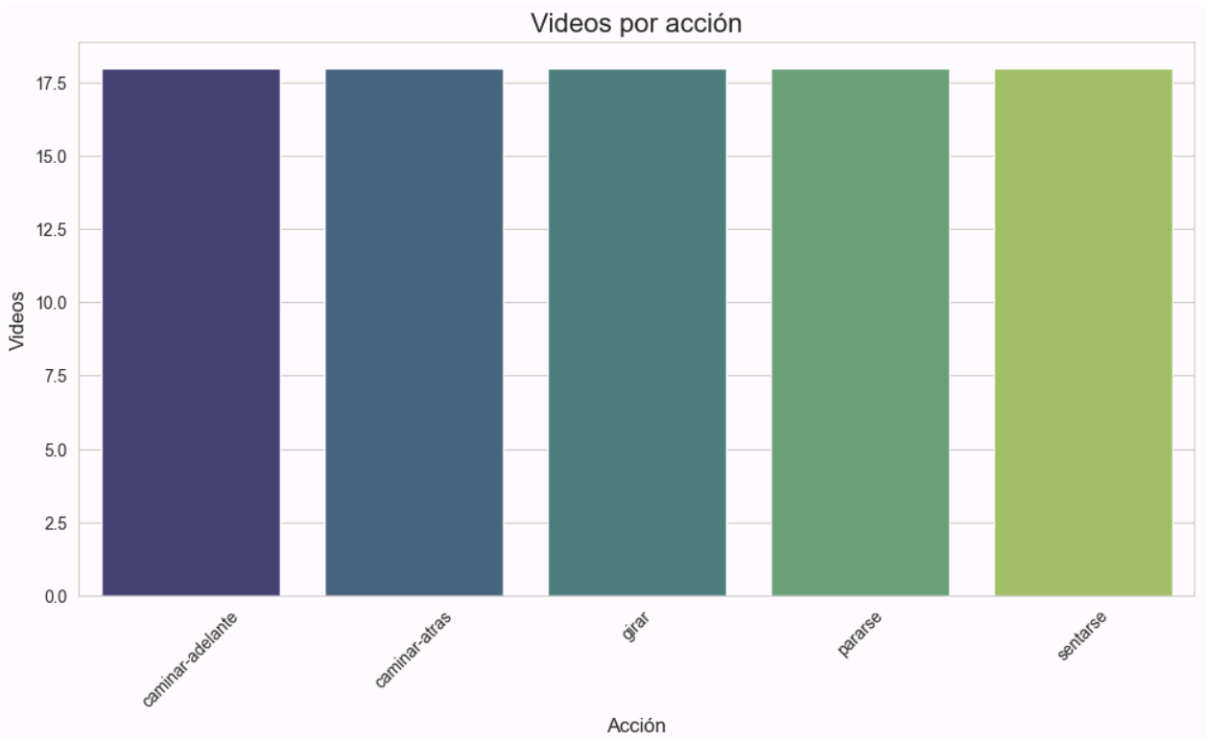


Figura 1

   b. Distribución de Fotogramas:

   A pesar del balance de videos, el número de fotogramas por acción es desigual, lo cual es esperado debido a que cada acción tiene una duración diferente. Por ejemplo, la acción "caminar-adelante" es la más larga, acumulando 1,827 fotogramas, mientras que "girar" y "pararse" son las más cortas, con aproximadamente 950 fotogramas cada una. Esta disparidad en la cantidad de datos de entrenamiento por clase deberá ser considerada en la etapa de modelado.

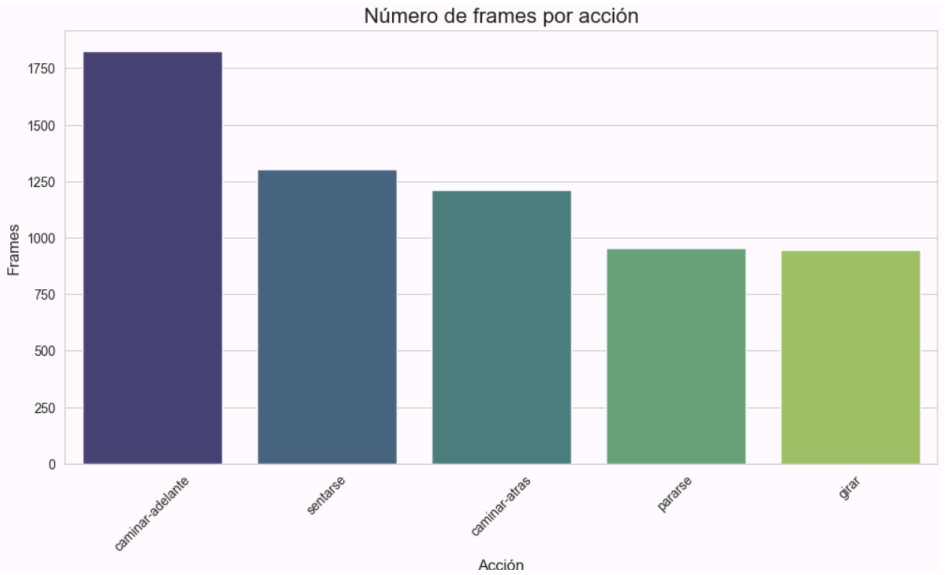


Figura 2

El análisis de la duración de los videos confirma lo observado en los fotogramas: las acciones de caminar son más largas (2.5-3.0 segundos en promedio) que las de girar o pararse (1.5-2.0 segundos).

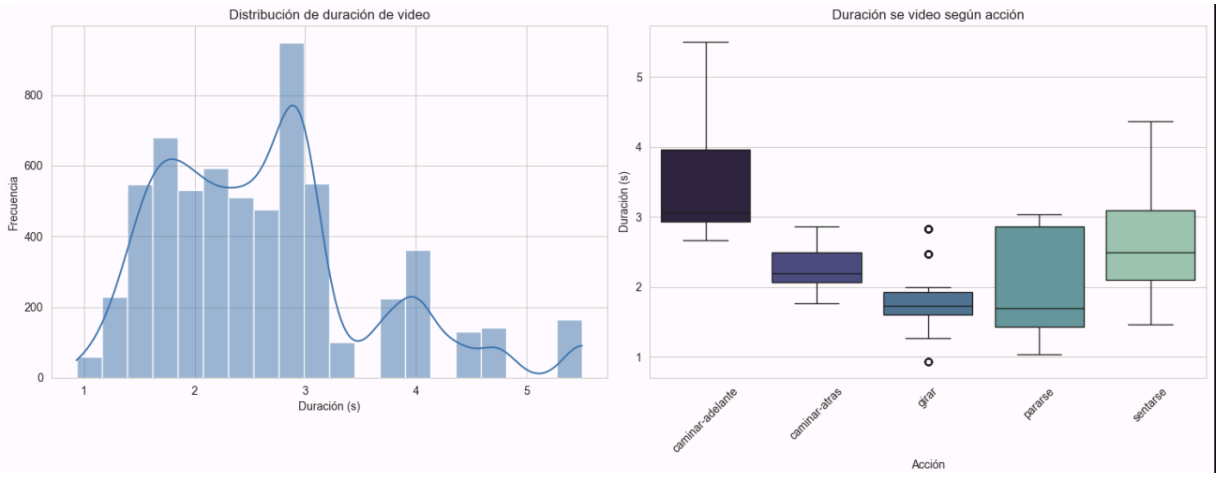


Figura 3

c. **Condiciones de Grabación:**

Por su parte, las métricas de luminancia media (luminance_mean) y desviación estándar (luminance_std) evidencian una iluminación homogénea entre videos, validando la consistencia experimental en las condiciones de captura (misma cámara, luz artificial controlada, distancia constante)

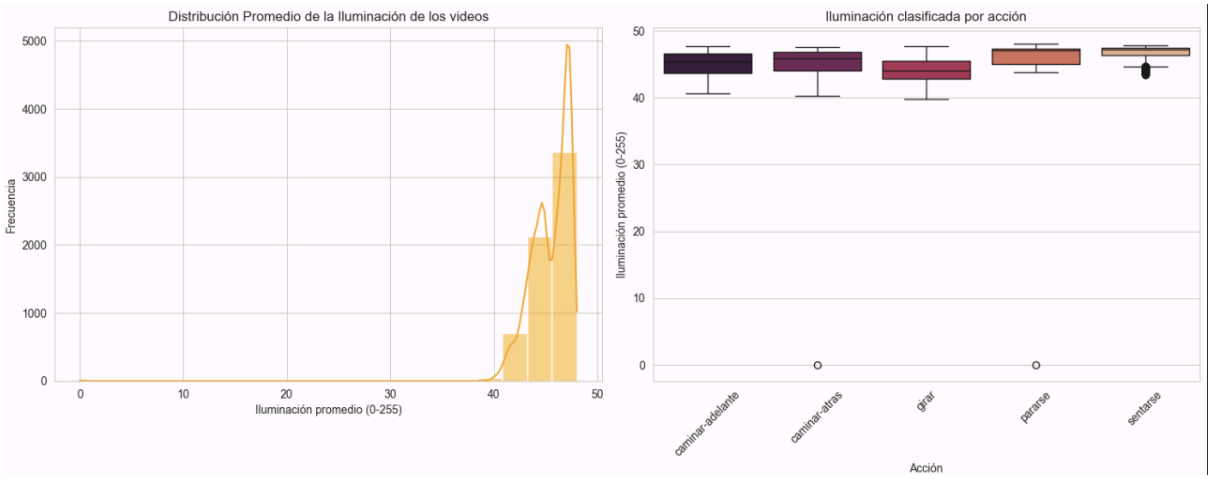


Figura 4

2. Análisis de Puntos de Referencia Corporales (datosmediapipe.csv)

a. Calidad de la Detección (Visibilidad):

La herramienta MediaPipe demostró una alta fiabilidad en la detección de los puntos corporales. La visibilidad promedio de los landmarks fue excelente, especialmente para el torso y el rostro (ojos, boca, hombros), con valores cercanos a 1.0. Las extremidades (tobillos, muñecas, talones) presentaron una visibilidad ligeramente menor, lo cual es normal debido al desenfoque por movimiento o a que ocasionalmente salen del encuadre. En general, la calidad de los datos de pose es muy alta.

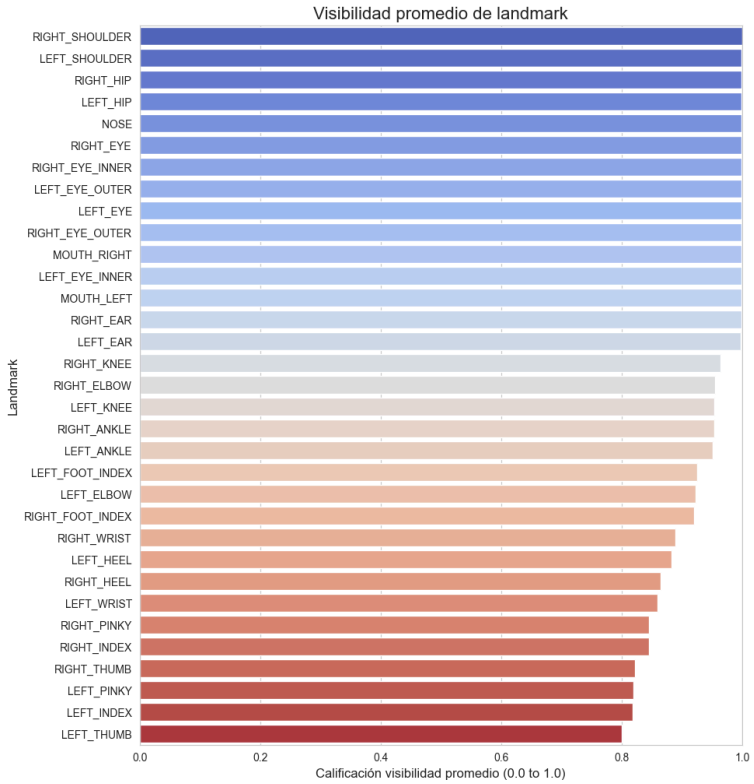
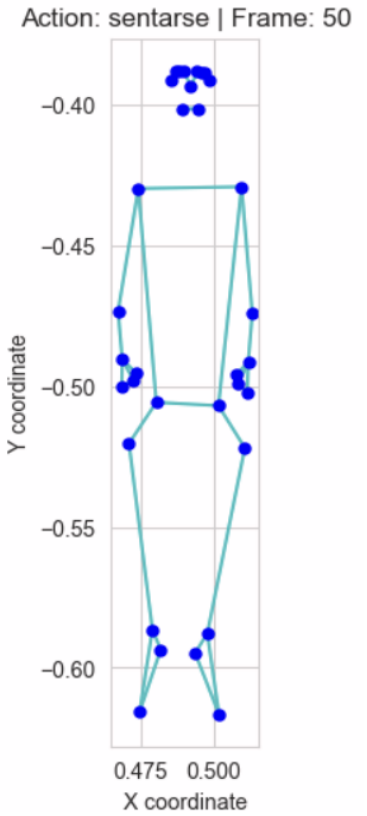
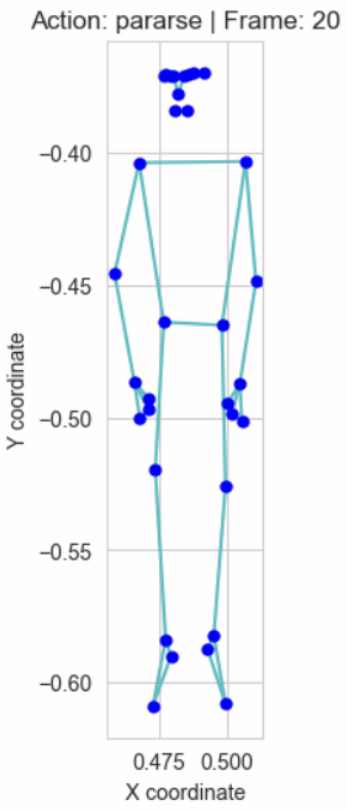


Figura 5

b. Visualización estructural de la pose

Para verificar la correcta extracción de coordenadas, se graficaron esqueletos 2D de ejemplo por acción, empleando las conexiones estándar de MediaPipe.





Las visualizaciones evidencian que la estructura corporal se conserva entre frames, y que las posiciones relativas de articulaciones son coherentes con los movimientos etiquetados.

c. **Separabilidad de Clases:**

El análisis de la distribución de landmarks clave reveló patrones distintivos entre las acciones, lo cual es muy prometedor para el modelo de clasificación. El hallazgo más significativo se observó en la posición vertical promedio de la cadera (hip_y_avg):

i. La acción **"sentarse"** muestra una distribución de valores claramente separada del resto, ya que la cadera se encuentra en una posición mucho más baja en el fotograma (valor de y más alto).
ii. Acciones como **"pararse"** y **"girar"** también presentan distribuciones características, aunque con mayor superposición con las acciones de "caminar".

d. Correlación de Landmarks clave:

Con el fin de analizar la relación entre los movimientos de diferentes partes del cuerpo, se construyó una matriz de correlación de coordenadas verticales (y) de seis landmarks representativos: hombros (11–12), caderas (23–24) y rodillas (25–26).

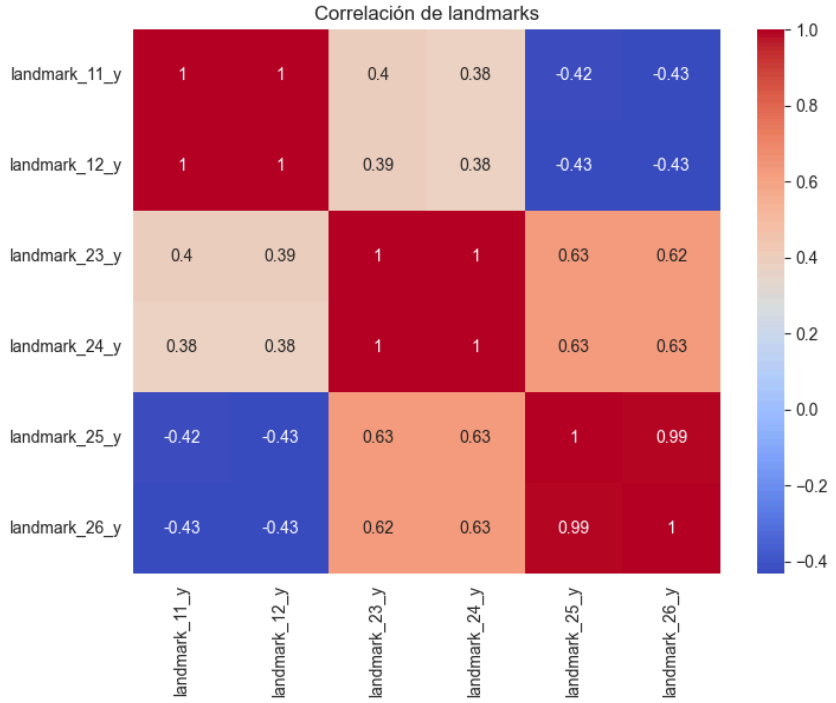


Figura 8

El análisis muestra correlaciones positivas muy altas entre los puntos simétricos del cuerpo; por ejemplo, entre las rodillas izquierda y derecha ($r \approx 0.99$) y entre las caderas ($r \approx 1.00$), llo que confirma la consistencia postural y simetría anatómica en la captura de datos.

Asimismo, se observa una correlación negativa moderada (-0.42) entre los hombros y las rodillas, lo cual indica que los movimientos verticales de las extremidades inferiores ocurren de manera inversa a los del tronco: cuando la persona se agacha o se sienta (rodillas descienden), los hombros tienden a bajar en menor proporción o incluso mantenerse estables.

Este patrón de correlaciones demuestra que los landmarks seleccionados son coherentes y biomecánicamente dependientes, por lo que resultan útiles para la futura ingeniería de características, permitiendo estimar ángulos, inclinaciones y sincronías articulares entre regiones corporales.

3. Conclusión del EDA:

El análisis confirma que el conjunto de datos es de alta calidad, balanceado y representativo, y que las características espaciales obtenidas con MediaPipe permiten distinguir claramente entre las distintas acciones humanas. Las correlaciones obtenidas entre articulaciones (Figura 8) confirman la coherencia biomecánica del conjunto de datos y respaldan su idoneidad para la extracción de características derivadas como ángulos y distancias.

En particular, la variable hip_y_avg demostró un poder discriminante notable entre posturas estáticas y dinámicas, lo que respalda la elección de un enfoque de clasificación supervisada multiclase en las etapas siguientes.

Asimismo, la consistencia en iluminación y detección de landmarks sugiere que el modelo será robusto ante variaciones leves de entorno.

Con base en estos hallazgos, los próximos pasos se centrarán en la ingeniería de características (ángulos, velocidades, distancias) y en la evaluación comparativa de clasificadores (Random Forest, XGBoost, SVM) bajo métricas de Accuracy, Precision, Recall y F1-Score.

Siguientes Pasos

Con base en los resultados del Análisis Exploratorio, se determinó que el conjunto de datos posee la calidad, balance y coherencia estructural necesarias para avanzar hacia las etapas de ingeniería de características y modelado supervisado.

Los siguientes pasos se enfocarán en construir un pipeline analítico que permita transformar los landmarks en variables más informativas para el clasificador y evaluar distintos algoritmos de aprendizaje automático.

1. Ingeniería de características

A partir de las coordenadas (x, y, z) y la visibilidad de los 33 landmarks se extraerán nuevas variables que representen **relaciones geométricas y cinemáticas** del cuerpo humano. Entre ellas:

- **Ángulos articulares:** Cálculo de ángulos entre segmentos corporales, como rodilla–cadera–hombro y tobillo–rodilla–cadera, para describir flexión y extensión.
- **Distancias relativas:** Medición de distancias euclidianas entre puntos clave (por ejemplo, mano–hombro, cadera–rodilla, hombros–caderas) para capturar proporciones posturales.
- **Velocidades y aceleraciones:** Derivadas temporales de landmarks consecutivos para representar la dinámica del movimiento y distinguir acciones continuas (caminar) de acciones discretas (sentarse, pararse).
- **Normalización espacial:** Reescalamiento de las coordenadas con respecto a la distancia entre hombros o caderas, con el fin de eliminar la dependencia de la altura o posición del individuo en el encuadre.

Estas transformaciones permitirán construir un conjunto de características robustas, invariante a escala y posición, apto para modelado supervisado.

2. Entrenamiento y evaluación de modelos

Se implementará un pipeline de entrenamiento con **división 80/20** para *train/test*, aplicando validación cruzada de 5 pliegues.
Los modelos iniciales a comparar serán:

- **Random Forest:** modelo base de referencia por su interpretabilidad y capacidad para manejar datos tabulares.
- **Support Vector Machine (SVM):** adecuado para problemas de clasificación multiclase con fronteras no lineales.
- **XGBoost:** modelo ensamble basado en boosting, óptimo para maximizar la precisión y reducir sesgo.

Las métricas a utilizar serán **Accuracy, Precision, Recall y F1-Score**, complementadas con una **matriz de confusión** para analizar la distribución de errores entre clases.

3. Validación experimental y comparación

Cada modelo será evaluado tanto en desempeño promedio como en su capacidad de generalización entre diferentes sujetos y condiciones.
Además, se explorará la relevancia de características mediante:

- Importancia de variables en Random Forest / XGBoost.
- Análisis PCA para reducción de dimensionalidad y visualización en 2D.

Esto permitirá identificar qué articulaciones o combinaciones de movimientos son más influyentes en la clasificación final.

4. Despliegue y visualización

Una vez seleccionado el modelo óptimo, se desarrollará una **interfaz gráfica sencilla** que permita visualizar:

- El video en tiempo real.
- Los landmarks detectados (MediaPipe).
- La clase de acción estimada por el modelo.
- Métricas básicas de postura (por ejemplo, inclinación del tronco, ángulo de rodillas).

Esta interfaz se implementará en Python y servirá como **entorno de validación visual** para pruebas con nuevos usuarios.

5. Extensión del dataset y control ético

Para mejorar la robustez del modelo, se planea:

- Incorporar nuevos videos con distintas condiciones de luz y ángulo de cámara.
- Aumentar la diversidad de participantes. (Mayor rango de edad)
- Mantener registro de consentimiento informado y anonimización de datos visuales.

Etapa	Objetivo	Resultado esperado
Ingeniería de características	Generar variables geométricas	Dataset enriquecido y

	y temporales a partir de landmarks	normalizado
Entrenamiento y evaluación	Comparar desempeño de modelos supervisados	Selección del mejor clasificador
Validación experimental	Analizar generalización y relevancia de características	Identificación de variables clave
Despliegue	Visualizar en tiempo real la clasificación y métricas posturales	Prototipo funcional
Ampliación del dataset	Aumentar diversidad y calidad de datos	Mayor robustez y equidad del sistema