

## **Integrantes**

Sharik Camila Rueda  
Juan Camilo Molina  
Nicolas Cardona  
Daniel Gonzalez

## **Preguntas de interés**

1. ¿Es posible entrenar un modelo supervisado capaz de reconocer correctamente en tiempo real las actividades de una persona (caminar hacia la cámara, caminar de regreso, girar, sentarse y ponerse de pie) a partir de los puntos articulares detectados por MediaPipe?
2. ¿Qué tan preciso puede ser el sistema al calcular la inclinación lateral del tronco y los ángulos de flexión de cadera y rodilla, tomando como base los puntos clave proporcionados por MediaPipe?
3. ¿El modelo mantiene un rendimiento estable cuando se aplican variaciones en las condiciones de captura, como cambios de iluminación, fondo, vestimenta o diferentes personas?
4. ¿Qué características (velocidad articular, ángulos relativos, inclinación del tronco, simetría de movimientos) aportan mayor valor a la clasificación y análisis postural?
5. ¿Qué modelo de aprendizaje supervisado ofrece mejor balance entre precisión y tiempo de inferencia?

## **Tipo de problema**

Con base en el análisis previo que el equipo realizó, decimos que el proyecto aborda dos tipos de problemas. Por un lado, el sistema debe identificar qué actividad realiza una persona (por ejemplo, caminar hacia la cámara, caminar de regreso, girar, sentarse o ponerse de pie); por otro, debe calcular y evaluar las posturas corporales involucradas en dichas acciones (por ejemplo, la inclinación del tronco o los ángulos de flexión de cadera y rodilla). Esto implica combinar modelos de aprendizaje supervisado con métodos de regresión geométrica. A continuación, se detalla cada tipo de problema identificado:

1. **Clasificación supervisada multiclasa:** Consiste en que el sistema aprenda a asignar una etiqueta de clase (actividad) a cada segmento temporal de video, basándose en las coordenadas de los puntos articulares obtenidos mediante herramientas como MediaPipe u OpenPose. En este escenario, cada muestra de datos representa una ventana temporal de keypoints normalizados, y el objetivo del modelo es reconocer patrones de movimiento característicos de cada actividad.
2. **Estimación geométrica y regresión de variables posturales:** El segundo tipo de problema que aborda el proyecto es cuantitativa y continua, y se centra en el análisis postural. Consiste en medir la posición y orientación del cuerpo en términos de ángulos e inclinaciones derivados de los keypoints corporales. Por tanto, el sistema debe ser capaz de calcular ángulos articulares, inclinación lateral del tronco y la simetría corporal, esto lo podemos enfocar con el modelado predictivo, el cual consiste en entrenar modelos que, dados los keypoints, predigan valores continuos de ángulos o desviaciones posturales, especialmente cuando hay ruido o pérdida de precisión en la detección.

## **Metodología**

El proyecto se desarrollará siguiendo la metodología CRISP-DM (Cross Industry Standard Process for Data Mining), un estándar ampliamente utilizado para la construcción de proyectos de análisis de datos e inteligencia artificial. Esta metodología permite avanzar de forma estructurada desde la comprensión del problema hasta la implementación de una solución funcional, garantizando que el proceso se lleve a cabo de forma correcta.

A continuación, se describe cómo se aplicará cada una de sus fases al desarrollo del sistema de anotación y clasificación de actividades humanas en video.

### **Comprensión del problema**

En esta primera etapa, el equipo busca comprender a fondo el propósito del sistema y las necesidades que pretende resolver. El objetivo principal es desarrollar una herramienta capaz de detectar en tiempo real distintas actividades humanas como caminar hacia la cámara, caminar de regreso, girar, sentarse o ponerse de pie y de analizar la postura corporal mediante el cálculo de ángulos articulares e inclinaciones del tronco.

Durante esta fase se definirán las preguntas de interés que orientarán el desarrollo del proyecto, así como el tipo de problema al que nos estamos enfrentando y las métricas de evaluación que permitirán medir el desempeño del modelo, tales como la precisión, el F1-score o el error medio en la estimación de ángulos.

Finalmente, se propondrá una estrategia para ampliar el conjunto de datos, ya sea mediante la incorporación de nuevas sesiones de grabación, la inclusión de participantes adicionales o el uso de técnicas de aumento de datos (data augmentation) que incrementen la variabilidad de las muestras y fortalezcan la capacidad de generalización del modelo.

### **Comprensión de los datos**

Una vez definido el objetivo, se procederá a comprender y caracterizar los datos con los que se trabajará. Para ello, el equipo capturará videos propios en los que distintas personas realizarán las actividades mencionadas. Estos videos constituirán la base del conjunto de datos y deberán representar variabilidad en sujetos (altura, complejión, género, vestimenta) y condiciones (iluminación, fondo, distancia a cámara).

Cada video será anotado manualmente para identificar los segmentos donde ocurre cada actividad. Esta anotación se realizará con herramientas de código abierto como Label Studio o CVAT, que permiten etiquetar de forma precisa los intervalos de tiempo donde se ejecuta cada acción.

Posteriormente, se aplicará MediaPipe (o en algunos casos OpenPose) para extraer los puntos clave (keypoints) del cuerpo humano: cabeza, hombros, caderas, rodillas, tobillos y muñecas. Estos puntos proporcionan información numérica (coordenadas x, y y nivel de confianza) que servirá como base para entrenar los modelos. Durante esta fase se realizará

también un análisis exploratorio de los datos, verificando la calidad de la detección de puntos, la proporción de datos faltantes, la estabilidad de las trayectorias y la consistencia entre distintos sujetos.

## Preparación de los datos

En esta etapa se transformarán los datos crudos en información lista para el modelado. Primero se realizará una limpieza de los datos eliminando keypoints con baja confianza o frames incompletos. Luego se aplicará un proceso de normalización, centrando las coordenadas en el punto medio de la cadera y escalándolas según la altura del sujeto, para evitar que la distancia a la cámara afecte las predicciones.

También se aplicará un filtro de suavizado a las trayectorias de las articulaciones, con el fin de eliminar pequeños saltos o ruido generados por la detección automática. Posteriormente, los datos serán segmentados en ventanas temporales (por ejemplo, de 1,5 segundos) que contendrán la secuencia de movimiento correspondiente a cada actividad.

A partir de estas ventanas, se generarán características (features) que representen el movimiento:

- Geométricas, como los ángulos entre cadera, rodilla y tobillo o la inclinación del tronco.
- Dinámicas, como las velocidades y aceleraciones de las articulaciones.
- Posturales, como la simetría entre el lado izquierdo y derecho del cuerpo.

Finalmente, los datos serán divididos en conjuntos de entrenamiento, validación y prueba, procurando que los sujetos sean diferentes en cada grupo para evaluar correctamente la capacidad de generalización del modelo.

## Modelado

Una vez que los datos estén preparados, se entrenarán diferentes modelos de aprendizaje supervisado para la tarea de clasificación de actividades.

En primer lugar, se probarán modelos tradicionales como SVM (Máquina de Vectores de Soporte), Random Forest y XGBoost, que son adecuados para trabajar con conjuntos de características tabulares.

Estos modelos permitirán identificar cuál ofrece un mejor equilibrio entre precisión, interpretabilidad y velocidad de ejecución.

En una segunda etapa, se evaluará la posibilidad de utilizar modelos secuenciales que puedan aprovechar la información temporal de los movimientos para mejorar la clasificación de transiciones entre actividades.

De forma complementaria, se implementarán cálculos geométricos para obtener los ángulos articulares y las inclinaciones corporales. Estos se basarán en relaciones trigonométricas entre los puntos detectados.

Si el nivel de ruido es alto, se considerará entrenar un modelo de regresión supervisado que corrija los errores de los ángulos estimados.

Durante esta fase se realizará la búsqueda de hiperparámetros, la comparación entre modelos y la selección del mejor con base en las métricas definidas (precisión, F1-score, error medio absoluto y tiempo de respuesta).

## Evaluación

Con los modelos entrenados, se procederá a una evaluación exhaustiva del desempeño del sistema.

Se medirá su capacidad para reconocer correctamente las actividades, la estabilidad de las predicciones y la precisión de los ángulos posturales.

Las métricas principales incluirán F1-score macro, precisión y recall por clase, matriz de confusión para detectar confusiones comunes, y error medio absoluto (MAE) para los valores angulares.

Además, se realizarán pruebas de robustez y generalización, utilizando datos de personas y condiciones que no hayan sido vistas durante el entrenamiento. También se analizará la latencia por frame y la tasa de cuadros por segundo (FPS) para comprobar que el sistema cumple con los requisitos de tiempo real.

Por último, se elaborará un análisis de errores para comprender en qué situaciones el modelo falla más (por ejemplo, confundir “sentarse” con “ponerse de pie”) y proponer mejoras.

## Implementación

En la fase final, todos los componentes desarrollados se integrarán en un prototipo funcional.

Este sistema recibirá la señal de video en tiempo real, detectará las articulaciones mediante MediaPipe, procesará los datos con el modelo seleccionado y mostrará en pantalla la actividad reconocida, los ángulos corporales y la inclinación del tronco.

La interfaz será sencilla e intuitiva, mostrando en el mismo cuadro el esqueleto de la persona, la etiqueta de la acción y las medidas principales. Se emplearán herramientas como OpenCV o Streamlit para la visualización interactiva.

En esta fase también se verificará el cumplimiento de las consideraciones éticas, asegurando que no se almacenen rostros ni información personal, y que todos los participantes hayan otorgado su consentimiento informado.

Finalmente, se documentará el proceso completo, incluyendo las configuraciones, los scripts de entrenamiento y las métricas obtenidas, para garantizar la reproducibilidad del proyecto y su posible ampliación en futuras investigaciones.

## Métricas de desempeño

Para evaluar el desempeño del sistema se implementaron un conjunto de métricas biomecánicas y cinemáticas derivadas del modelo MediaPipe Pose, las cuales permiten analizar cuantitativamente el movimiento y la postura de las personas en los videos procesados.

En primer lugar, se calculó la velocidad promedio de la cadera (`hip_speed_avg_px_per_frame`), que mide el desplazamiento del punto medio entre ambas caderas en píxeles por fotograma. Esta métrica es un indicador clave de la intensidad y continuidad del movimiento, pues permite diferenciar acciones dinámicas (como caminar o girar) de aquellas más estáticas o transicionales (como sentarse o pararse). Una mayor velocidad implica un desplazamiento corporal sostenido, mientras que valores cercanos a cero indican estabilidad o ausencia de movimiento.

En segundo lugar, se analizó el movimiento promedio entre fotogramas (`movement_avg_px`), obtenido a partir de la variación de la posición de las articulaciones entre cuadros consecutivos. Esta métrica complementa la velocidad de la cadera, ya que cuantifica el grado general de movimiento del cuerpo, representando la fluidez y amplitud de la acción. Valores elevados reflejan una actividad con desplazamiento amplio, mientras que los bajos se asocian a movimientos localizados o pausas.

Posteriormente, se consideró el ángulo promedio de la rodilla (`knee_angle_avg_deg`), formado entre las articulaciones de la cadera, rodilla y tobillo. Esta métrica biomecánica es esencial para describir el grado de flexión o extensión de las piernas, y por tanto, permite identificar con precisión las fases posturales de una acción.

Asimismo, se incorporó la inclinación promedio de los hombros (`shoulder_tilt_avg_px`), calculada como la diferencia vertical entre el hombro derecho y el izquierdo. Este valor refleja el balance y la orientación direccional del cuerpo, siendo positivo o negativo dependiendo de la inclinación hacia uno u otro lado. Esta métrica es especialmente útil para distinguir desplazamientos laterales o giros, ya que las variaciones de inclinación acompañan naturalmente el sentido del movimiento.

De forma adicional, se registró el brillo promedio del video (`brightness_avg`), una métrica técnica que no influye directamente en la postura o el movimiento, pero sirve para verificar la homogeneidad de las condiciones de captura y garantizar que las diferencias observadas entre videos provienen del comportamiento corporal y no de factores de iluminación. También se conservaron los parámetros de duración, número de fotogramas y frecuencia de cuadros por segundo (FPS) como indicadores del rendimiento temporal y la consistencia de las grabaciones.

### **Estrategia para ampliar el conjunto de datos**

- 1. Colaboración con otros grupos o cursos que desarrollen el mismo proyecto**

Una estrategia práctica y de bajo costo consiste en compartir datasets con otros grupos de estudiantes que estén realizando el mismo proyecto. Esto permitiría reunir un conjunto más amplio de ejemplos de las actividades a clasificar (caminar, girar, sentarse, etc.), capturados con distintos participantes, entornos y cámaras.

Para garantizar la coherencia del conjunto combinado, sería necesario acordar un protocolo común de captura y etiquetado (nombres de clases, formato de los datos, resolución de video, FPS y estructura de los keypoints). De esta forma, se podrían unificar fácilmente los datasets sin introducir inconsistencias o errores de formato.

## 2. Búsqueda y uso de datasets públicos disponibles en Internet

Otra posibilidad es buscar bases de datos ya existentes que contengan secuencias de personas realizando actividades similares a las que se analizarán en este proyecto. Si se encuentran datasets que incluyan las acciones objetivo (caminar, girar, sentarse, ponerse de pie), podrían emplearse para entrenar parcialmente el modelo o reforzar la fase de validación. No obstante, esta estrategia requiere verificar las licencias de uso, garantizar la compatibilidad de formatos y asegurarse de que no existan datos personales sensibles (rostros visibles, contexto privado).

## 3. Captura adicional de videos propios con nuevos participantes

Una alternativa igualmente efectiva es ampliar la recolección de datos con voluntarios adicionales. El grupo puede diseñar un protocolo de grabación sencillo, accesible para cualquier persona con una cámara o teléfono móvil, indicando condiciones básicas como la distancia, el encuadre y la iluminación.

De esta manera, se podrán obtener grabaciones variadas en contextos reales que aporten diversidad de entornos y personas. Posteriormente, se anonimizarán los videos ocultando rostros o guardando solo los keypoints y se integrarán al conjunto principal.

### Estrategia seleccionada

La estrategia más adecuada para ampliar nuestro conjunto de datos es la **colaboración con otros grupos o cursos que estén desarrollando el mismo proyecto**, ya que representa una alternativa realista, eficiente y completamente alineada con el contexto académico. Esta opción permite obtener una mayor cantidad de ejemplos sin necesidad de realizar grabaciones adicionales ni incurrir en costos logísticos, aprovechando los datos ya capturados por otros equipos bajo condiciones similares de grabación, formato y actividades. Además, al compartir datasets entre grupos que siguen los mismos lineamientos metodológicos y éticos, se garantiza la coherencia en las etiquetas, la calidad de los registros y la compatibilidad de los datos, lo que facilita su integración sin comprometer la precisión del modelo.

### Aspectos éticos

1. Privacidad y consentimiento informado: todas las personas filmadas deben conocer el objetivo del proyecto y autorizar el uso de sus videos exclusivamente con fines académicos.
2. Anonimización: se recomienda ocultar el rostro o procesar solo los puntos articulares (sin almacenar imágenes).

3. Minimización de datos: conservar únicamente la información necesaria (coordenadas y etiquetas).
4. Equidad: asegurar la representación de diferentes tipos de cuerpo, edades y géneros para evitar sesgos en el modelo.
5. Transparencia: documentar claramente los límites del sistema (no es herramienta médica ni de vigilancia).
6. Seguridad: mantener los datos almacenados en entornos seguros y con acceso restringido.
7. Responsabilidad social: promover el uso ético del sistema, únicamente en contextos de investigación o mejora ergonómica, no para monitoreo sin consentimiento.