# Reconocimiento de Actividades Humanas en Tiempo Real mediante Poses Corporales y Modelos de Aprendizaje Automático Supervisado

## Facultad de Ingeniería, Diseño y Ciencias Aplicadas Departamento de Computación y Sistemas Inteligentes Ingeniería de Sistemas

Docentes: Uram Sosa Aguirre, Milton Sarria Paja

Asignatura: Inteligencia Artificial I

Institución: Universidad Icesi

Integrantes:

David Santiago Malte, Samuel Ibarra, Alejandro Muñoz

Semestre: 2025 - 1

Fecha de Entrega:

13 de junio 2025

### Introducción

Este informe presenta el diseño e implementación de un sistema inteligente para el reconocimiento de actividades humanas en tiempo real a partir de poses corporales. Utilizando la biblioteca MediaPipe para la detección de poses y algoritmos de aprendizaje automático como Random Forest, XGBoost y redes neuronales, el sistema identifica cinco movimientos clave: caminar hacia adelante, caminar hacia atrás, sentarse, levantarse y girar.

El sistema se desarrolló bajo la metodología CRISP-DM, abarcando desde la exploración y preparación de los datos, hasta el modelado, validación y despliegue de una interfaz en tiempo real. El objetivo principal fue comparar el rendimiento de diferentes modelos y seleccionar el más robusto para ser integrado en la aplicación final.

### Contexto

En el marco del curso de Inteligencia Artificial 1 (2025-1), se propuso el diseño de un sistema inteligente capaz de reconocer, en tiempo real, actividades humanas a partir de poses esqueléticas. El objetivo central fue comparar el desempeño de diferentes algoritmos de clasificación sobre un mismo conjunto de datos, obteniendo métricas objetivas de desempeño y desarrollando una aplicación funcional que permita evidenciar su utilidad práctica.

### Objetivos

* Objetivo general:

Desarrollar un sistema de reconocimiento de actividades humanas en tiempo real, comparando el desempeño de modelos de clasificación supervisada con base en features biomecánicos extraídos de las poses corporales detectadas con MediaPipe.

* Objetivos específicos:
* Implementar un pipeline de preprocesamiento y extracción de características.
* Aplicar técnicas de EDA para explorar y validar la calidad de los datos.
* Entrenar, validar y comparar al menos tres modelos: Random Forest, XGBoost y red neuronal.
* Evaluar el rendimiento mediante métricas estándar (accuracy, F1-score).
* *I*ntegrar el modelo más robusto en una interfaz gráfica interactiva.

### Metodología (CRISP-DM)

1. Entendimiento del negocio

El reconocimiento automático de actividades humanas es clave en aplicaciones como la rehabilitación, monitoreo de pacientes, asistencia a adultos mayores, análisis deportivo y vigilancia. El reto propuesto se centra en detectar movimientos humanos simples pero relevantes (caminar adelante, atrás, sentarse, levantarse, girar) en tiempo real, evaluando la precisión y robustez de los modelos.

1. Entendimiento de los datos

Se capturaron más de 18,000 frames de poses esqueléticas con MediaPipe, cada muestra consiste en 33 puntos del cuerpo con coordenadas (x, y) y visibilidad (v), lo que genera 135 columnas por frame. El dataset se distribuye en cinco clases balanceadas.

1. Preparación de los datos

**Análisis exploratorio (EDA):**

Se cargaron más de **18,000 registros (frames)** con 135 columnas, cada una representando coordenadas (x,y,v) de los landmarks detectados.

* + - * Distribución de clases:

Se buscó que las clases estén balanceadas, procurando tener la misma cantidad de datos para cada tipo de movimiento: caminar\_adelante, caminar\_atras, vuelta, sentarse, levantarse.

* + - * **Análisis de nulos**:

Gracias a la eficiencia de Mediapipe, no fue necesaria la limpieza de datos nulos.

* + - * **Filtrado por visibilidad**:

Se eliminaron frames con visibilidad baja.

**Feature Engineering:**

A partir de los 33 puntos del cuerpo se extrajeron features biomecánicas relevantes como:

* + - Ángulos de rodilla y cadera (izquierda y derecha).
    - Inclinación del tronco.
    - Distancia entre hombros.
    - Distancia entre caderas.

Este enfoque redujo la dimensionalidad de 132 a solo 7 características clave por muestra.

### Modelado

Se entrenaron 3 modelos principales con validación cruzada y búsqueda de hiperparámetros:

* Random Forest

Accuracy: 0.92

Precision media: 0.80

Problemas de sobreajuste con la clase caminar hacia atrás.

* XGBoost

Accuracy: 0.85

Mejor manejo de outliers y relaciones no lineales, sin embargo presentó problemas de sobreajuste

* Red Neuronal

Accuracy: 0.70%

Arquitectura: 4 capas densas y una capa de salida

Fue el modelo final utilizado en la app real.

### Evaluación de resultados

Se evaluaron tres enfoques de clasificación para identificar actividades humanas. A continuación se resumen sus resultados:

|  |  |  |  |
| --- | --- | --- | --- |
| Modelo | Accuracy | F1-score | Observaciones clave |
| Random Forest (base) | 0.80 | 0.80 | Buen desempeño general, pero pobre en clases como “Caminar hacia atrás”. |
| Random Forest (Grid Search) | 0.93 | 0.92 | Mejoró notablemente el rendimiento de la versión base. Altamente competitivo. |
| XGBoost | 0.85 | 0.84 | Mejor equilibrio en todas las clases. Sobresalió en "Levantarse" y "Sentarse". |
| Red Neuronal | 0.75 | **0.74** | Tuvo problemas en clases de "caminar", aunque buen desempeño en "sentarse", "vuelta" y "levantarse". |

Red neuronal – detalles

|  |  |  |  |
| --- | --- | --- | --- |
| Clase | Precisión | Recall | F1-score |
| Caminar hacia adelante | 0.62 | 0.60 | 0.61 |
| Caminar hacia atrás | 0.67 | 0.60 | 0.63 |
| Levantarse | 0.83 | 0.77 | 0.80 |
| Sentarse | 0.79 | 0.86 | 0.83 |
| Vuelta | 0.79 | 0.88 | 0.83 |

Esto demuestra que el modelo neuronal tenía una alta capacidad para reconocer movimientos estáticos (como sentarse o girar), pero no fue suficientemente preciso para movimientos dinámicos como caminar.

### Conclusiones

* La extracción de features biomecánicos fue crucial para la generalización del modelo.
* Random Forest optimizado fue el mejor clasificador offline.
* La red neuronal, aunque con menor accuracy, fue ideal para integrar por eficiencia y simplicidad.
* El uso de MediaPipe facilitó un pipeline robusto de captura de poses sin requerir sensores externos.
* La metodología CRISP-DM permitió un desarrollo iterativo, controlado y estructurado.