**Proyecto Final APO3: Sistema de Anotación de Video**

**Informe de Avance 1 (Semana 12)**

**Integrantes:**

* Juan Pablo Parra A00398004
* Joshua Rivera A00399847
* Santiago Espinosa A00399531
* Thomas Brueck A00399947

**Fecha:** 13 Octubre

**1. Comprensión del Negocio (Business Understanding)**

**1.1. Contexto y Problema**

El análisis de actividades humanas (Human Activity Recognition - HAR) es un campo de la visión por computadora con aplicaciones en áreas como la salud (rehabilitación, monitoreo de pacientes), la seguridad (vigilancia inteligente) y la interacción humano-robot. Tradicionalmente, este análisis requiere supervisión humana, lo cual es costoso en tiempo y propenso a errores.

Este proyecto aborda la necesidad de un sistema automatizado que pueda capturar, analizar y clasificar actividades humanas básicas en tiempo real. El sistema propuesto busca resolver el desafío de interpretar secuencias de vídeo para extraer información semántica sobre las acciones que una persona está realizando, así como sus datos posturales.

**1.2. Preguntas de Interés para el Proyecto**

1. Sobre los Datos y Características (Feature Engineering)

El objetivo aquí es entender qué información extraída del esqueleto es realmente útil para identificar las actividades.

* **P1: ¿Cuáles son los *landmarks* (articulaciones) más informativos para distinguir entre cada par de actividades?**
  + ¿Son los tobillos y rodillas suficientes para diferenciar "caminar" de "sentarse", o se necesita información de la cadera y los hombros para mejorar la precisión? Esto permitirá decidir si se usan los 33 landmarks de MediaPipe o un subconjunto.
* **P2: ¿Qué tipo de características de ingeniería (*feature engineering*) tiene mayor poder predictivo para clasificar las acciones?**
  + Comparar el rendimiento de un modelo usando solo coordenadas normalizadas vs. uno que además incluya ángulos articulares (ej. ángulo de la rodilla) vs. otro que también incorpore velocidades y aceleraciones de los landmarks.
* **P3: ¿Cómo afecta la variabilidad entre diferentes personas (tamaño, velocidad de ejecución) al conjunto de datos y cómo podemos mitigarlo?**

**2. Sobre el Modelado y la Evaluación (Modeling & Evaluation)**

Estas preguntas se centran en encontrar el mejor algoritmo y configurarlo de manera óptima.

* **P4: De los modelos seleccionados (SVM, Random Forest, XGBoost), ¿cuál ofrece el mejor balance entre precisión (F1-Score) y velocidad de inferencia para este problema?**
  + *Metodología:* Entrenar cada modelo con el mismo conjunto de datos y evaluar tanto su F1-Score con validación cruzada como el tiempo que tarda en predecir la actividad para un nuevo conjunto de fotogramas.
* **P5: ¿Qué hiperparámetros tienen el mayor impacto en el rendimiento de cada modelo y cuáles son sus valores óptimos?**
* **P6: ¿Qué actividades son las más difíciles de clasificar para el modelo (las que más confunde) y por qué?**

**1.3. Tipo de Problema**

El problema a resolver se enmarca dentro del campo del aprendizaje automático supervisado. Específicamente, se trata de un problema de clasificación multiclase.

* Es supervisado porque se entrenará un modelo utilizando un conjunto de datos previamente etiquetado, donde cada secuencia de video o fotograma está asociada a una de las actividades conocidas.
* Es de clasificación porque el objetivo es asignar una etiqueta categórica y discreta (ej. "caminando", "sentado") a cada nueva observación (los datos del esqueleto de un fotograma).
* Es multiclase debido a que existen más de dos posibles clases de salida (en este caso, cinco actividades distintas).

**1.4. Objetivos del Proyecto**

El objetivo principal es desarrollar una herramienta de software que clasifique cinco actividades específicas y realice un seguimiento de los movimientos articulares clave de una persona a partir de una señal de vídeo en tiempo real.

Los objetivos secundarios son:

1. Construir un conjunto de datos propio (dataset) de videos para entrenar y evaluar los modelos.
2. Extraer y preprocesar datos de esqueletos 2D (landmarks corporales) utilizando la librería MediaPipe.
3. Diseñar y entrenar al menos tres modelos de clasificación supervisada (SVM, Random Forest, XGBoost) para identificar las actividades definidas.
4. Evaluar y comparar el rendimiento de los modelos utilizando métricas apropiadas (precisión, F1-Score) para seleccionar el más adecuado.
5. Implementar una interfaz gráfica simple que muestre el video en tiempo real, la clasificación de la actividad y la visualización del esqueleto.

**2. Metodología de Trabajo y Métricas:**

**2.1. Un Enfoque Iterativo Basado en CRISP-DM**

Para abordar este proyecto de manera sistemática y rigurosa, se ha adoptado el Marco de Referencia del Proceso Estándar Inter-industrias para la Minería de Datos (CRISP-DM). Esta metodología, reconocida por su enfoque estructurado e iterativo, se adaptará para guiar el ciclo de vida del desarrollo, desde la concepción del problema hasta la implementación de la solución. El proceso se concibe como un ciclo continuo, donde los hallazgos de una fase retroalimentan y refinan las fases anteriores.

A continuación, se detalla la aplicación de cada fase al contexto de este proyecto:

**Fase 1: Comprensión del Problema (Business Understanding)**

* **Objetivo:** Traducir el requerimiento general ("crear un sistema de anotación") en un problema técnico de machine learning bien definido.
* **Tareas Específicas:**
  + Definición del Problema: Establecer el problema como una tarea de clasificación supervisada multiclase, donde la entrada son datos de series temporales (secuencias de coordenadas de landmarks) y la salida es una etiqueta de actividad.
  + Definición de Criterios de Éxito: Cuantificar el éxito a través de métricas técnicas (F1-Score > 90%) y operacionales (Inferencia > 15 FPS), asegurando que la solución no solo sea precisa, sino también funcional en tiempo real.
  + Análisis de Restricciones: Identificar limitaciones como la calidad de la cámara, las condiciones de iluminación y el hardware disponible.

**Fase 2: Comprensión de los Datos (Data Understanding)**

* Objetivo: Recolectar y familiarizarse con los datos para identificar su calidad, estructura y características iniciales.
* Tareas Específicas:
  + Recolección de Datos Iniciales: Ejecución del protocolo de grabación para generar el dataset de videos.
  + Extracción de Estructuras de Datos: Procesamiento de los videos con MediaPipe para transformar los datos crudos (píxeles) en datos estructurados (coordenadas x, y, z, visibility por fotograma en archivos CSV).
  + Análisis Exploratorio de Datos (EDA): Investigación profunda de los datos de landmarks para descubrir patrones, anomalías y relaciones. Esto incluye la visualización de trayectorias de articulaciones, el análisis de la distribución de visibilidad de los landmarks y la verificación del balance de clases.

**Fase 3: Preparación de los Datos (Data Preparation)**

* **Objetivo:** Transformar los datos crudos de landmarks en un formato limpio, consistente y óptimo para el entrenamiento de los modelos de clasificación. Esta es la fase más crítica y laboriosa.
* **Tareas Específicas:**
  + Limpieza de Datos: Implementar estrategias para manejar datos ruidosos o faltantes, como la interpolación lineal para landmarks ocluidos (baja visibilidad) o el uso de filtros (ej. Savitzky-Golay) para suavizar las trayectorias.
  + Ingeniería de Características (Feature Engineering): Crear variables predictivas de alto nivel a partir de las coordenadas básicas. Esto incluye:
    - Normalización Postural: Hacer el modelo invariante a la posición y escala del sujeto, recalculando las coordenadas relativas a un punto central del cuerpo (ej. el centro de la cadera).
    - Características Cinemáticas: Calcular ángulos entre articulaciones (ej. ángulo de la rodilla, inclinación del torso) y velocidades/aceleraciones de landmarks clave.
  + Formateo de Datos: Estructurar los datos en secuencias o ventanas temporales, ya que una actividad depende no solo de una postura instantánea, sino de la evolución de las posturas en el tiempo. Finalmente, escalar las características (ej. StandardScaler) para optimizar el rendimiento de algoritmos sensibles a la escala como SVM.

**Fase 4: Modelado (Modeling)**

* **Objetivo:** Seleccionar, entrenar y optimizar diferentes algoritmos de machine learning para encontrar el que mejor se ajuste al problema.
* **Tareas Específicas:**
  + Selección de Algoritmos: Entrenar modelos de diferentes familias (SVM, Random Forest, XGBoost) para comparar su capacidad de aprender de los datos posturales.
  + Protocolo de Experimentación: Diseñar un protocolo de evaluación robusto utilizando validación cruzada estratificada (Stratified K-Fold). Esto garantiza que cada modelo se evalúe sobre diferentes subconjuntos de datos, proporcionando una estimación fiable de su rendimiento y evitando sesgos.
  + Ajuste de Hiperparámetros: Realizar una búsqueda sistemática de los mejores hiperparámetros para cada modelo (ej. con GridSearchCV) para maximizar su poder predictivo.

**Fase 5: Evaluación (Evaluation)**

* **Objetivo:** Medir el rendimiento de los modelos finales de manera objetiva y determinar si cumplen con los criterios de éxito definidos.
* **Tareas Específicas:**
  + Cálculo de Métricas: Evaluar los modelos en un conjunto de prueba nunca visto durante el entrenamiento, utilizando las métricas definidas (F1-Score, matriz de confusión).
  + Análisis de Errores: Analizar la matriz de confusión para identificar qué actividades son sistemáticamente confundidas y formular hipótesis sobre las causas (ej. características insuficientes, transiciones ambiguas).
  + Revisión del Proceso: Contrastar los resultados con los objetivos. Si el rendimiento no es satisfactorio, los hallazgos de esta fase retroalimentarán el proceso, llevando a decisiones como recolectar más datos, diseñar nuevas características (volver a Fase 3) o probar arquitecturas de modelos más complejas.

**Fase 6: Despliegue (Deployment)**

* **Objetivo:** Integrar el modelo entrenado en una aplicación funcional.
* **Tareas Específicas:**
  + Creación de Pipeline de Inferencia: Construir una función que tome un fotograma de video, aplique en tiempo real las mismas transformaciones de la Fase 3, y use el modelo para predecir la actividad.
  + Visualización: Desarrollar una interfaz gráfica simple (usando OpenCV o Tkinter) que muestre el video, la superposición del esqueleto y la etiqueta de la actividad predicha en tiempo real.

**2.2. Métricas para medir progreso**

Para medir el avance y el rendimiento del proyecto, se utilizarán dos tipos de métricas:

* **Métricas de Evaluación del Modelo:**
  + Accuracy: Proporción de predicciones correctas.
  + Precision, Recall y F1-Score (por clase): Para evaluar el rendimiento en cada actividad específica, especialmente si hay desbalance. El F1-Score macro será la métrica principal para comparar modelos.
  + Matriz de Confusión: Para analizar visualmente qué actividades son confundidas entre sí por el modelo.
* **Métricas de Desempeño del Sistema:**
  + Fotogramas por Segundo (FPS): Medirá la velocidad del pipeline completo (captura, preprocesamiento, inferencia) para asegurar que la aplicación funciona en tiempo real (objetivo: >15 FPS).

**3. Recolección de Datos y Análisis Exploratorio**

**3.1. Plan de Recolección de Datos**

Se creará un conjunto de datos propio siguiendo este protocolo:

* **Sujetos:** Los 4 integrantes del equipo.
* **Actividades:** Caminar hacia la cámara, caminar de regreso, girar, sentarse, ponerse de pie.
* **Volumen:** Cada integrante grabará 10 videos por cada una de las 5 actividades, para un total inicial de 200 videos.
* **Entorno Controlado:** Los videos se grabarán en un entorno con fondo uniforme y buena iluminación para garantizar la calidad de la detección de MediaPipe.
* **Especificaciones:** Formato .mp4, resolución 720p, 30 FPS.

**3.2. Análisis Exploratorio de Datos (Plan)**

Una vez extraídos los landmarks con MediaPipe y almacenados en formato CSV, se realizará el siguiente análisis:

1. **Visualización de Esqueletos:** Se generarán gráficos de los esqueletos sobrepuestos en los fotogramas para verificar visualmente la calidad de la detección.
2. **Estadísticas de Visibilidad:** Se calculará la tasa de detección (visibilidad > 0.7) para cada *landmark* en el total de los fotogramas. Esto5 permitirá identificar articulaciones problemáticas (ej. pies ocluidos).
3. **Distribución de Clases:** Se verificará que el número de fotogramas por cada actividad esté balanceado para evitar sesgos en el entrenamiento del modelo.

**3.3. Estrategias para Incrementar el Conjunto de Datos**

Si el análisis exploratorio o los resultados iniciales sugieren que el modelo necesita más datos para generalizar, se aplicarán las siguientes estrategias:

* **Data Augmentation:** Se aplicarán transformaciones a los videos existentes, como el **volteo horizontal (mirroring)**. Esto duplicará efectivamente el tamaño del dataset, especialmente útil para actividades como "sentarse" o "girar".
* **Inclusión de más Sujetos:** Se solicitará a 2-3 personas adicionales (fuera del equipo del proyecto) que graben videos siguiendo el mismo protocolo para aumentar la variabilidad y robustez del modelo.
* **Variación de Condiciones:** Se grabarán videos adicionales en diferentes condiciones de iluminación y con fondos más complejos para entrenar un modelo más resistente al ruido del entorno real.

**4. Siguientes Pasos**

| Tarea | Responsable(s) | Fecha Límite |
| --- | --- | --- |
| Finalizar la grabación de los 200 videos | Todos | Semana 13 |
| Desarrollar y ejecutar script de extracción de landmarks | Juan Pablo & Joshua | Semana 13 |
| Realizar el Análisis Exploratorio de Datos (EDA) | Santiago & Thomas | Semana 14 |
| Implementar funciones de preprocesamiento y *features* | Todos | Semana 14 |
| Entrenar y evaluar modelos base (versión inicial) | Todos | Semana 14 |
| Preparar y entregar el informe de avance 2 | Todos | Semana 14 |

**Análisis ético**

El primer eje ético es la privacidad. La captura de video y el seguimiento de articulaciones constituyen tratamiento de información potencialmente sensible. Por ello, el proyecto debe limitar el uso de los datos al objetivo declarado, minimizar la recolección (sólo lo necesario para entrenar y evaluar) y aplicar técnicas de anonimización o seudonimización cuando sea posible (por ejemplo, almacenar y procesar preferentemente coordenadas de landmarks en lugar de imágenes crudas cuando la experimentación lo permita). Además, corresponde implementar controles de acceso, cifrado en repositorio y políticas claras de retención y eliminación de datos una vez cumplido el propósito académico.

Un segundo pilar es el consentimiento informado y la transparencia. Antes de grabar a cualquier persona, debemos explicar con lenguaje claro qué se va a hacer, qué riesgos y beneficios existen, durante cuánto tiempo se guardarán los datos, quién tendrá acceso y con qué fines se usarán (incluida la publicación de resultados o demostraciones). La transparencia también aplica al comportamiento del sistema: se debe comunicar que las salidas del modelo son estimaciones probabilísticas y que pueden contener errores, evitando su uso para emitir juicios concluyentes sobre conducta o desempeño sin validación adicional.

Por último es importante la responsabilidad sobre el uso y las consecuencias del sistema. Aunque el proyecto es académico, el prototipo podría migrar a ámbitos sensibles (salud, seguridad, recursos humanos). En tales escenarios, su despliegue sin validación rigurosa y sin supervisión humana podría generar daños. Por lo tanto, el equipo debe establecer límites de uso, advertencias y salvaguardas: el sistema se concibe como apoyo a la toma de decisiones, no como sustituto del criterio profesional; y cualquier transferencia tecnológica futura exigirá nuevas evaluaciones éticas, técnicas y legales.