De cara al modelado a llevar a cabo a lo largo del paquete de trabajo posterior, “PT3: Sistema experto de prevención de caídas”, dicha información será la base para el entrenamiento y generación de dichos modelos. Para ello, se contemplan distintas fuentes de información para obtener las variables necesarias en el entrenamiento de los modelos supervisados.

Por un lado, es indispensable disponer de la información relativa a la variable objetivo de este sistema: la existencia conocida de riesgo de caída por parte del sujeto en estudio. De esta forma, se buscará generar un clasificador que permita determinar si nuevos usuarios tienen una mayor probabilidad de formar parte del grupo de riesgo o de aquellos que, a priori, no parecen presentar razones para pensar en tal problemática. Tal y como se indica en la memoria del proyecto, se buscará tener información de tres tipos de sujetos atendiendo a dicha variable: pacientes considerados con riesgo de caída, pacientes en proceso de rehabilitación tras una caída, y sujetos que no tienen patologías asociadas a dicha problemática en estudio.

A continuación, también será de especial relevancia tener en cuenta variables de interés propias del sujeto, tales como edad, sexo, peso o altura (según la disponibilidad de tal información). Todo esto servirá para que el modelo tenga capacidad de contextualizar al individuo, dado que el patrón de movimiento podrá considerarse distinto según dichas características.

Finalmente, y como información principal para el análisis, se presentan los datos de los sensores recogidos a través del experimento diseñado a lo largo del actual paquete de trabajo “PT2: *Wearable* de medición de la marcha humana”, y en especial, en la tarea asociada a este entregable, “T2.3: Captura de datos inicial”. Dichos datos serán recogidos para distintos individuos según su tipología con respecto al riesgo de caída, y serán obtenidos llevando a cabo un experimento homogéneo entre todos ellos. Así, se buscará detectar las cuatro fases de la marcha asociadas en función de los datos de los sensores, las cuales vienen dadas de forma esquemática en la Figura 1.

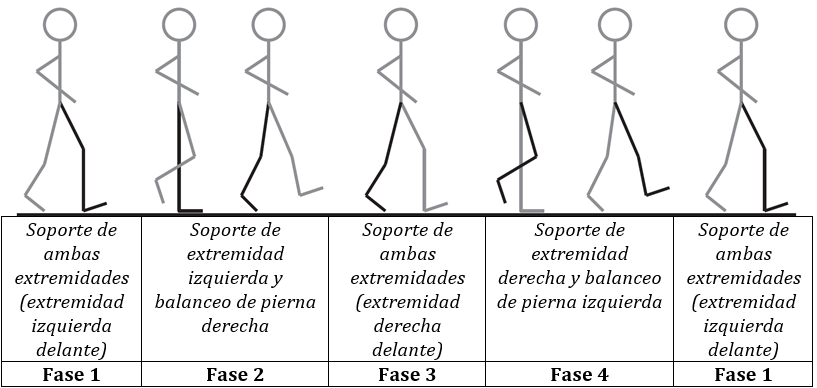


Figura 1. Descripción esquemática de las cuatro fases de la marcha

Así, una vez separadas las fases de cada recorrido, se llevará a cabo un análisis estadístico de los subconjuntos generados, de tal forma que se obtengan variables agregadas con respecto a los tres sensores que componen el sistema de medición, y que sirvan posteriormente para el entrenamiento del modelo.

Para mayor detalle en lo relativo a cómo obtener dichas fases, así como la agregación estadística asociada, se podrán consultar los entregables asociados al PT3, y en particular, se hará un mayor énfasis en su obtención en el entregable “E3.1: Procedimiento de depuración y preprocesado de los datos”.

De modo esquemático, se presenta en la Figura 2 el flujo de trabajo esperado a lo largo del PT3 previamente mencionado.

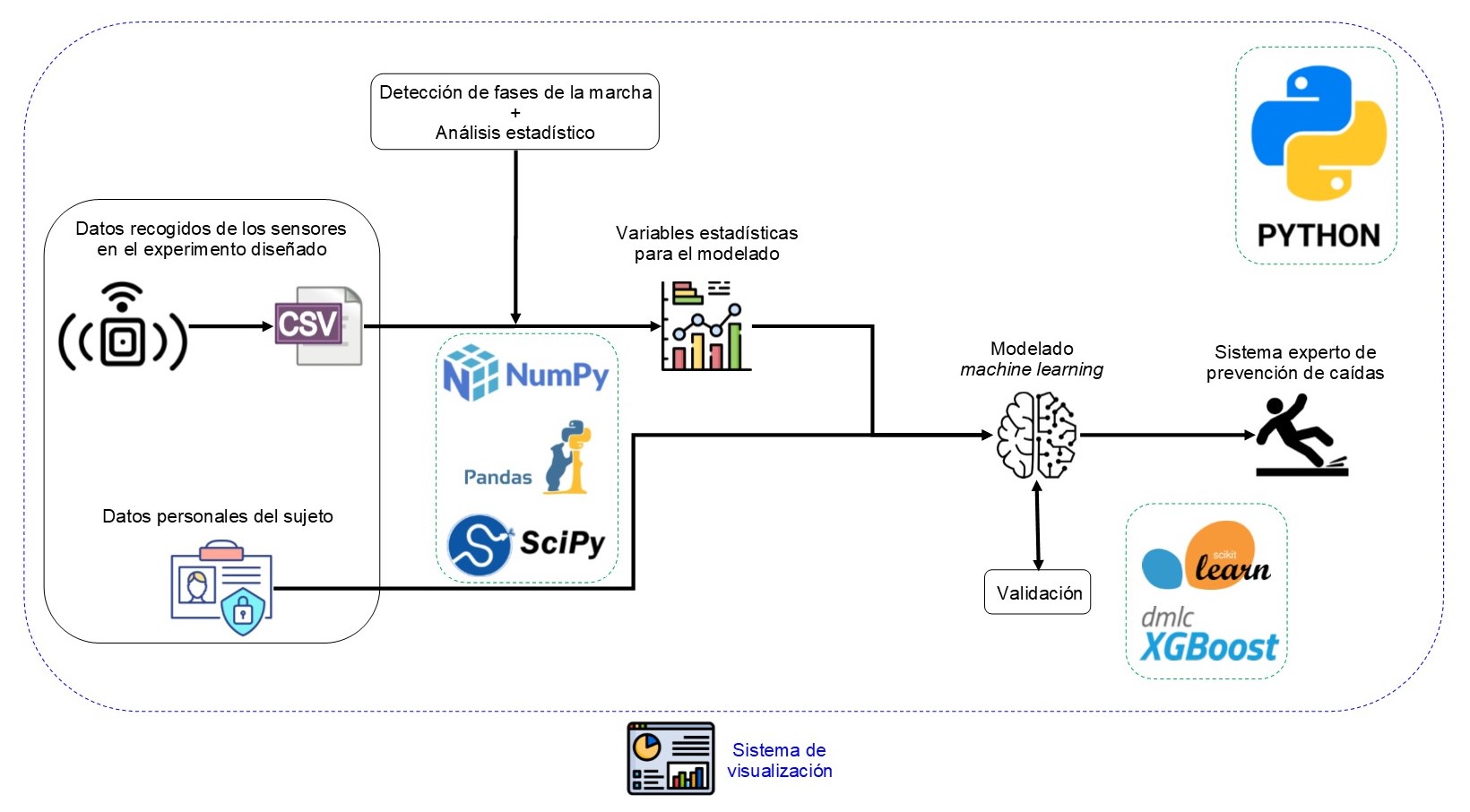


Figura . Esquema del flujo de trabajo del proceso inteligente esperado

En ésta, se pueden apreciar como punto de partida los datos del experimento recogidos a través del sensor seleccionado, así como información complementaria de los datos personales de cada sujeto. Con esto, se buscará, tal y como se menciona previamente, la forma de detectar las distintas fases de la marcha mostradas en la Figura 1, de tal modo que, junto a los datos personales iniciales, se proceda con los distintos pasos relativos al modelado inteligente. Todo esto, requerirá una fase previa exhaustiva de pre-procesado y preparación de los datos.

Dicho modelado se realizará haciendo uso de distintos modelos posibles y parametrizaciones de algoritmos de aprendizaje supervisado de clasificación, y será validado utilizando las técnicas que se consideren adecuadas mediante la selección de la métrica o métricas de relevancia.

Así, dicho modelo será la base para la generación de un sistema experto de prevención de caídas que, junto con todo lo previo, será incluido en un sistema de visualización que presente la información relevante detectada.

En cuanto a las herramientas de programación a utilizar, se basará el desarrollo en el lenguaje Python, haciendo uso a priori de la versión 3.7 de éste, y utilizando librerías asociadas al análisis de datos e inteligencia artificial, como *numpy*, *pandas*, *scipy*, *sklearn* o *xgboost*.