A picture containing text, font, screenshot, graphics

Description automatically generated

**Facultat Internacional de Comerç i Economia Digital La Salle**

Trabajo Final de Máster

Máster Universitario en Ciencia de los Datos / Data Science

TÍTULO TFM

Alumno Profesor Ponente

Bernat Sort Rufat Roger Mallol Parera

|  |
| --- |
| **ACTA DEL EXAMEN**  **DEL TRABAJO FINAL DE MÁSTER** |

Reunido el Tribunal calificador en el día de la fecha, el alumno

**Bernat Sort Rufat**

Expuso su Trabajo de Final de Máster, el cual trató sobre el tema siguiente:

**TÍTULO DEL TFM**

Acabada la exposición y contestadas por parte del alumno las objeciones formuladas por los miembros del tribunal, este valoró el mencionado Trabajo con la calificación de

|  |
| --- |
|  |

Barcelona,

VOCAL DEL TRIBUNAL VOCAL DEL TRIBUNAL

PRESIDENTE DEL TRIBUNAL

Agradecimientos

Resumen

Este proyecto consiste en desarrollar una herramienta de predicción de resultados clínicos en el manejo de los trastornos de la marcha en niños con parálisis cerebral. El objetivo principal es predecir cómo queda la cinemática de los niños después de realizar ciertas intervenciones quirúrgicas. Para lograrlo, se utilizarán datos cinemáticos previos y posteriores obtenidos a través del análisis tridimensional del movimiento, así como las intervenciones aplicadas. Se trabajará con una muestra homogénea de niños con la misma enfermedad (diplejia espástica y deformidades ortopédicas) y edad similar. El segundo objetivo es desarrollar una aplicación de Inteligencia Artificial (IA) que pueda aprender de la experiencia clínica para ayudar en la toma de decisiones desde los datos cinemáticos de la prueba inicial.

El proyecto se dividirá en varias fases. En primer lugar y una vez se hayan obtenido los datos, se realizará un exhaustivo análisis exploratorio de los datos (EDA) para comprenderlos mejor, así como un preprocesado de estos. Esto incluirá realizar estadísticas descriptivas, visualizaciones de las variables, visualizaciones que permitan ver las relaciones entra cada variable y la variable objetivo, tratamiento de valores nulos, tratamiento de outliers y balanceo de datos si fuese necesario. A continuación, se seleccionarán las variables predictoras, es decir las características más relevantes para la predicción del resultado clínico. Estas se escogerán en base a técnicas como el análisis de correlación, análisis de componentes principales (PCA), SHAP, e información mutua. Para lograr los objetivos mencionados, se utilizarán técnicas de aprendizaje automático, en particular modelos de regresión y clasificación. Se entrenarán y evaluarán varios modelos de aprendizaje automático, como la Regresión Logística, Random Forests, SVC*, gradient boosting methods* (XGBoost y LightGBM) y ensamblaje de modelos (por ejemplo, el *stacking*). La herramienta de predicción de resultados se desarrollará utilizando técnicas de regresión para predecir la cinemática posterior de los niños después de las intervenciones quirúrgicas. La aplicación de inteligencia artificial se basará en técnicas de clasificación para proporcionar recomendaciones para la toma de decisiones clínicas basadas en los datos cinemáticos de la prueba inicial.

La organización del proyecto seguirá un enfoque iterativo y colaborativo, en el que se trabajará en estrecha colaboración con el equipo clínico responsable de la gestión de los pacientes con parálisis cerebral. Se comenzará con la recopilación de datos cinemáticos y clínicos de los pacientes, que se utilizarán para entrenar y validar los modelos predictivos. Posteriormente, se desarrollará la aplicación de inteligencia artificial y se realizará una evaluación rigurosa de su capacidad para ayudar en la toma de decisiones clínicas. Finalmente, se presentarán los resultados obtenidos y se discutirán las posibles limitaciones y perspectivas futuras de la herramienta desarrollada.

**Palabras clave:** *Análisis Clínico de la Marcha, Biomecánica, Parálisis Cerebral, Cinemática de medias, Machine Learning, Deep Learning, Redes Neuronales.*

Abstract

Tabla de Contenidos

[1 Introducción 1](#_Toc143360895)

[1.1 Contexto 1](#_Toc143360896)

[1.2 Motivación y justificación 1](#_Toc143360897)

[1.3 Objetivos 1](#_Toc143360898)

[1.4 Hipótesis 1](#_Toc143360899)

[1.5 Estructura del documento 1](#_Toc143360900)

[2 Estado del arte 1](#_Toc143360901)

[3 Métodos 1](#_Toc143360902)

[3.1 Estudio de la población y recolección de datos 1](#_Toc143360903)

[3.2 Análisis exploratorio de los datos 1](#_Toc143360904)

[3.3 Preprocesamiento de los datos 1](#_Toc143360905)

[3.4 Análisis multivariante (modelos) 1](#_Toc143360906)

[4 Resultados y discusiones 5](#_Toc143360907)

[4.1 Limitations of the study 5](#_Toc143360908)

[5 Consideraciones éticas 5](#_Toc143360909)

[6 Coste económico 5](#_Toc143360910)

[7 Conclusiones 5](#_Toc143360911)

[8 Vías abiertas 5](#_Toc143360912)

[9 Referencias 6](#_Toc143360913)

[10 Apéndice 6](#_Toc143360914)

[10.1 Apéndice 1. Repositorio del proyecto 6](#_Toc143360915)

[11 APUNTES EXTRA 6](#_Toc143360916)

[11.1 Ciclo de la marcha vs paseo 6](#_Toc143360917)

[11.2 Ciclos 7](#_Toc143360918)

[11.3 Sample 8](#_Toc143360919)

[11.4 Medias 8](#_Toc143360920)

[11.5 Corte de ciclos 9](#_Toc143360921)

[11.6 Proceso 9](#_Toc143360922)

[12 Correo sergio 12](#_Toc143360923)

[13 Dudas sergio 13](#_Toc143360924)

[14 chatGPT 14](#_Toc143360925)

[14.1 MODELS 14](#_Toc143360926)

Índice de Figuras

[Figure 1. Flowchart of this study. This flowchart shows the patient selection and the different steps applied in each cohort. 12](#_Toc110619725)

[Figure 2. Daily frequency of PaO2 records per patient versus the daily frequency of SpO2 records per patient. 13](#_Toc110619726)

[Figure 3. Correlation between PaO2/FiO2 and SpO2/FiO2 ratios. 14](#_Toc110619727)

[Figure 4. SpFi distribution based on each respiratory SOFA PaFi score. 15](#_Toc110619728)

[Figure 5. Groups (0, 1, 2, 3, and 4) of the new respiratory SOFA score based on the SpO2/FiO2 ratio. 15](#_Toc110619729)

[Figure 6. Histogram showing the distance between SOFA SpFi score and SOFA PaFi score. 16](#_Toc110619730)

[Figure 7. Comparison of PaFi and SpFi AUCs obtained by the several ML algorithms for classification. 18](#_Toc110619731)

Índice de Tablas

[Table 1. The criteria for assessment of the respiratory SOFA score [12]. 2](#_Toc110594726)

[Table 2. Variable selection. The criteria for assessment of the respiratory SOFA score [12]. 6](#_Toc110594727)

[Table 3. Features and target dataset example of PaFi on subset srpafi4% > 0\* 9](#_Toc110594728)

[Table 4. Features and target dataset example of SpFi on subset srpafi4% > 0\* 9](#_Toc110594729)

[Table 5. Performance metric for the machine learning algorithms in both datasets: AUC. 17](#_Toc110594730)

# Introducción

## Contexto

La biomecánica, el análisis de la marcha y la rehabilitación constituyen componentes integrales del panorama médico, cuyo objetivo es mejorar el bienestar y la capacidad funcional de las personas con deficiencias musculoesqueléticas. Los trastornos musculoesqueléticos abarcan una amplia gama de afecciones que afectan a huesos, músculos, ligamentos y articulaciones, y que a menudo provocan alteraciones en los patrones de movimiento y reducen la movilidad. Abordar estas deficiencias mediante terapias y regímenes de rehabilitación adaptados es fundamental no sólo para recuperar la función física, sino también para mejorar la calidad de vida general de las personas afectadas.

Sin embargo, hasta la fecha no se ha encontrado una forma de garantizar que la rehabilitación sea eficaz para lograr un cambio significativo en la vida del paciente, lo que a veces supone una grave pérdida de tiempo y dinero tanto para el paciente como para el sistema sanitario. Aunque la rehabilitación es muy prometedora a la hora de restablecer las funciones y mejorar el bienestar de las personas con deficiencias musculoesqueléticas, su eficacia varía mucho debido a la compleja interacción de factores que influyen en los resultados de la terapia. Esta imprevisibilidad supone un reto importante tanto para los médicos como para los pacientes, que invierten un tiempo, un esfuerzo y unos recursos considerables en programas de rehabilitación con resultados inciertos.

Es en este contexto donde la importancia de esta investigación emerge con mayor claridad. La laguna existente entre el impacto potencial de la rehabilitación y la realidad de unos resultados inciertos exige un cambio de paradigma. La integración de técnicas de aprendizaje automático para evaluar y predecir el progreso de la terapia promete transformar el panorama de la rehabilitación. Aprovechando el poder de los conocimientos basados en datos del innovador conjunto de datos GaitRec, esta investigación se esfuerza por salvar la brecha entre la intención y el resultado, entre la esperanza y la evidencia.

Las implicaciones de evaluar y predecir eficazmente el progreso de la terapia van más allá del ámbito clínico. Dotar a los médicos de herramientas para adaptar las intervenciones en función de indicadores objetivos de progreso es una innovación que puede redefinir el panorama del tratamiento. Los pacientes, armados con una comprensión más clara de su viaje de rehabilitación, pueden embarcarse en el camino de la recuperación con un sentido de propósito y dirección. Además, la posibilidad de optimizar la asignación de recursos dentro de los sistemas sanitarios, junto con el potencial para mitigar el coste emocional, físico y económico de los pacientes, sirve de faro que guía la trayectoria de esta investigación.

## Motivación y justificación

Esta investigación encuentra su motivación en la necesidad de reducir la brecha entre los objetivos de la rehabilitación y los resultados concretos alcanzados por individuos con deficiencias musculoesqueléticas.

La incertidumbre que rodea la efectividad de las intervenciones de rehabilitación no solo plantea un desafío práctico, sino también una preocupación profundamente humanitaria. A medida que las personas emprenden el camino hacia la recuperación, invirtiendo esperanzas, energía y recursos, la incapacidad para anticipar la evolución del progreso puede resultar en desilusión y frustración.

Desde una perspectiva social, las ineficiencias derivadas de programas de rehabilitación ineficaces o mal adaptados pueden ejercer una carga excesiva sobre los sistemas de atención sanitaria. La motivación subyacente de este proyecto surge en respuesta a estas inquietudes, como un llamada a tomar medidas con el fin de trazar una ruta más sólida y fundamentada hacia el éxito en la rehabilitación.

La convergencia de la ciencia de datos y la biomecánica ofrece una oportunidad que resuena directamente con el núcleo de la motivación de esta investigación. Aprovechando el valor de los datos, este estudio busca elevar el proceso de toma de decisiones terapéuticas desde el ámbito de la intuición y la experiencia hacia la evidencia y el análisis riguroso. En última instancia, la finalidad es mejorar la calidad de vida de las personas afectadas por deficiencias musculoesqueléticas y contribuir al desarrollo de enfoques de rehabilitación más eficaces y fundamentados.

## Objetivos

El objetivo principal de este proyecto es analizar y predecir el progreso y la evolución de los pacientes con deficiencias musculoesqueléticas sometidos a terapia o rehabilitación mediante el desarrollo y la evaluación de modelos de aprendizaje automático.

Estos modelos podrían servir como herramientas de apoyo a la toma de decisiones para los médicos, ayudándoles a evaluar la eficacia de las intervenciones de rehabilitación y a tomar decisiones informadas sobre la continuación, adaptación o finalización de la terapia.

Es por ello por lo que esta iniciativa busca mejorar la atención médica de las personas con deficiencias musculoesqueléticas al proporcionar a los profesionales sanitarios una herramienta adicional basada en datos que puede enriquecer su juicio clínico y, en última instancia, contribuir a resultados más exitosos en el proceso de rehabilitación.

Es por ello que, en última instancia, el objetivo es mejorar la atención médica de las personas con deficiencias musculoesqueléticas al proporcionar a los médicos una herramienta adicional basada en datos que puede enriquecer su juicio clínico y contribuir a resultados más exitosos en el proceso de rehabilitación.

## Hipótesis

Se parte de la hipótesis de que los modelos de aprendizaje automático desarrollados y entrenados utilizando un amplio conjunto de datos anotados, que contiene ensayos de marcha con fuerzas de reacción del suelo (GRF) bilaterales de pacientes que presentan diversas discapacidades musculoesqueléticas [1], demostrarán la capacidad de predecir de manera precisa el progreso y la evolución de los pacientes sometidos a terapia o rehabilitación.

## Estructura del documento

En la introducción se presenta el contexto necesario para la investigación y se sientan las bases de nuestro estudio. Se comienza introduciendo el campo de la biomecánica, el análisis de la marcha y la rehabilitación. Se discute la relevancia de comprender y evaluar los patrones de la marcha en individuos con deficiencias musculoesqueléticas y el impacto potencial en su calidad de vida. Se explican los retos que plantea la incertidumbre de los resultados de la rehabilitación y las posibles ineficiencias de los sistemas sanitarios. Se plantean los objetivos de nuestra investigación, así como hipótesis que guiará la investigación.

En el marco teórico, se profundiza en el conjunto de conocimientos existentes relacionados con nuestro tema de investigación. Se realiza una revisión de estudios similares relacionados con la evaluación de terapias e intervenciones de individuos con deficiencias musculoesqueléticas, así como de las aplicaciones del aprendizaje automático en la predicción del progreso y evolución de los pacientes. Finalmente, se destacan algunas de las limitaciones del panorama actual.

En el apartado de métodos, se presentan los procedimientos adoptados para la realización de la investigación. El primer apartado aborda la descripción exhaustiva del conjunto de datos empleado, incluyendo el número de participantes involucrados, los diversos tipos de deficiencias musculoesqueléticas abarcadas y el proceso riguroso de recolección de datos. A continuación, se presenta el análisis exploratorio de los datos, seguido de la exposición de los pasos de preprocesamiento de datos y la ingeniería de características o *feature engineering*. Posteriormente, se detalla el proceso de división de los datos para a la fase de entrenamiento, validación y pruebas. Se destaca además la elección y la justificación de los algoritmos de aprendizaje automático seleccionados. Para finalizar, se abordan en detalle las métricas de evaluación empleadas para valorar el desempeño y rendimiento de los modelos de aprendizaje automático.

En el apartado de resultados y discusión, se evalúa el rendimiento de los distintos modelos de aprendizaje automático utilizados empleando métricas de evaluación como el error cuadrático medio (RMSE) [] y el coeficiente de determinación (R2) []. Se compara los resultados de los distintos modelos y se seleccionan los que mejor rendimiento hayan tenido a la hora de predecir el progreso y la evolución de los pacientes con deficiencias musculoesqueléticas sometidos a terapia o rehabilitación. Se interpretan las implicaciones de los resultados en relación con la hipótesis y objetivos, se contextualizan los resultados, y se analiza su importancia en el ámbito de la evaluación del progreso de la rehabilitación.

En el siguiente apartado, se discuten las consideraciones éticas en torno al uso de los datos de los pacientes, haciendo hincapié en las medidas que se tomaron para garantizar la privacidad del paciente y la seguridad de los datos. Además, se reflexiona sobre el impacto potencial de nuestra investigación en la atención al paciente y las implicaciones éticas del uso de modelos predictivos en la toma de decisiones clínicas.

En la sección de coste económico, se analizan las implicaciones económicas relacionadas con nuestro proyecto de investigación. Se realiza una estimación de los recursos económicos requeridos para la ejecución del proyecto, abarcando tanto los costos de los materiales necesarios para llevar a cabo la investigación como el cálculo del salario por hora del científico de datos encargado de su realización. Esta evaluación económica proporciona una visión completa de los recursos financieros necesarios para el desarrollo y la ejecución exitosa del proyecto de investigación en cuestión.

En la sección de conclusiones, se presentan de manera concisa las conclusiones más destacadas y las contribuciones significativas derivadas de nuestra investigación. Se revisan los objetivos e hipótesis planteadas en el proyecto con el propósito de evaluar si se han alcanzado satisfactoriamente, y se destaca la importancia del trabajo realizado en el avance del campo de la evaluación del progreso de la rehabilitación y su posible impacto en la atención al paciente y los sistemas sanitarios.

Finalmente, en el apartado de vías abiertas, se proponen posibles vías de investigación y desarrollo en este campo. Se reflexiona cómo nuestro trabajo puede ampliarse, refinarse o adaptarse a diferentes contextos. También se aportan ideas sobre cómo podrían integrarse nuestros modelos predictivos en la práctica clínica, contribuyendo a la atención personalizada de los pacientes y a la optimización de las estrategias de rehabilitación.

# Estado del arte

Mirar articles que hagin utilizat el tipus de dades que he fet servir jo

# Métodos

Escribirlo en pasado

## Estudio de la población y recolección de datos

## Análisis exploratorio de los datos

## Preprocesamiento de los datos

## Análisis multivariante (modelos)

**Evaluación inicial:**

El niño acude a la consulta y el especialista lo evalúa comprender su historial médico, síntomas, limitaciones funcionales y objetivos de tratamiento. También se realizan pruebas físicas y evaluaciones clínicas para obtener información sobre la función motora y la marcha actual del niño.

**Preparación del paciente:**

Para poder realizar el registro de la marcha, se le pide al niño que use ropa ajustada y cómoda, y se le colocan los marcadores en puntos anatómicos clave del cuerpo. Estos marcadores ayudan a rastrear y registrar los movimientos y la posición de las articulaciones durante la marcha.

**Registro de la marcha:**

El niño camina en un área designada mientras se graba su forma de andar utilizando sistemas de captura de movimiento o fotogrametría. Estos sistemas pueden incluir cámaras de video, sensores inerciales, sensores de presión en los pies u otras tecnologías avanzadas que capturan datos cinemáticos y cinéticos de la marcha.

**Datos y su procesamiento:**

Los datos se capturan del sistema de análisis de movimiento (nombre del sistema ¿SMARTanayzer -BTS S.p.a.?) que funciona a 250 Hz con 8 cámaras. El software utilizado es ¿nombre del software?.

El archivo inicial (.tdf) contiene gran cantidad de datos no anonimizados.

Se etiquetan los marcadores para pasar de tener trayectorias en x,y,z y unas matrices de datos brutos a tener datos etiquetados.

A continuación, el técnico del laboratorio corta los datos etiquetados en ciclos. Es decir, se define algo que clínicamente es relevante.

Posteriormente, el médico o especialista selecciona los ciclos que considera más importantes o significativos, y los demás se descartan. A partir de los ciclos seleccionados se obtienen los valores promedios de los ángulos de cada articulación en cada muestra de los ciclos de la marcha registrados. En este punto se han excluido los datos de carácter personal, por tanto, tenemos los datos anonimizados.

A partir de los ciclos procesados, se obtiene la cinemática en tres dimensiones de las articulaciones, que se utilizará en el proceso de toma de decisiones.

**Descripción de los datos cinemáticos:**

A picture containing receipt, text, white, algebra

Description automatically generated

Figura 1. Fichero emt correspondiente a los datos cinemáticos preoperatorios del miembro inferior derecho de un paciente.

Cada paciente realizó múltiples caminatas o paseos durante la sesión de evaluación o grabación de la marcha. Las caminatas son el conjunto de pasos que realiza un individuo durante una sesión de evaluación o grabación de la marcha y pueden estar compuestas por varios ciclos de la marcha, y el número exacto de ciclos por caminata puede variar dependiendo de la duración de la caminata y las características individuales del paciente. Como observamos en la Figura X, se han registrado un total de 40 ciclos de la marcha. El ciclo de la marcha se refiere al período completo desde el momento en que un pie toca el suelo hasta el siguiente contacto del mismo pie. Se divide en dos fases principales: la fase de apoyo y la fase de balanceo. Durante la fase de apoyo, el pie está en contacto con el suelo y soporta el peso del cuerpo, mientras que, durante la fase de balanceo, el pie se levanta del suelo y se mueve hacia adelante para preparar el próximo contacto.

A picture containing silhouette, dance, footwear

Description automatically generated

Figura 2. Ciclo de la marcha.

Los datos cinemáticos pre y postoperatorios de los se encuentran en los ficheros emt, que son ¿descripción ficheros emt y de donde provienen?.

La columna *sample* representa el número de muestras dentro de los ciclos de la marcha registrados. El número de muestras indica la secuencia de cada punto de datos tomado durante el movimiento de la marcha. Para cada paciente tenemos 101 muestras en total, numeradas desde 0 hasta 100. Cada muestra corresponde a un punto específico en el tiempo durante los 40 ciclos de la marcha registrados. ¿Estas muestras se han tomado en intervalos de tiempo uniformes o en momentos específicos durante la caminata para capturar la información cinemática en diferentes fases del ciclo de la marcha? Esta información es útil para identificar y hacer un seguimiento de cada punto de datos específico en relación con el ciclo de la marcha.

Las columnas restantes, representan los valores promedio de los ángulos de las diferentes articulaciones de las extremidades inferiores derecha (R) e izquierda (L) en cada muestra de los ciclos de la marcha registrados. Cada ángulo está asociado con una articulación específica, como la rodilla (K), la cadera (H), el tobillo (A), pelvis (P) y ¿hombro? (SH). Los valores de los ángulos se registran en grados y varían en cada muestra del ciclo de la marcha.

Los datos cinemáticos se componen por grados – tiempo. Por tanto, vamos a tener 11 gráficas con 101 valores cada gráfica (101 valores de cada miembro inferior). Tenemos 2 piernas representadas en cada gráfica con 101 valores por pierna. Esos 101 valores suelen ser la media de un par de estudios o 3 que hemos incluido. Es un conjunto bastante grande de datos.

# Resultados y discusiones

Resultados en pasado

Discusiones en presente

# Discusiones

## Limitaciones del estudio

# Consideraciones éticas

# Coste económico

# Conclusiones

# Vías abiertas

# Referencias

1. Horsak, B., Slijepcevic, D., Raberger, AM. *et al.* GaitRec, a large-scale ground reaction force dataset of healthy and impaired gait. *Sci Data* **7**, 143 (2020). <https://doi.org/10.1038/s41597-020-0481-z>
2. Marshall, J. C., Bosco, L., Adhikari, N. K., Connolly, B., Diaz, J. V., Dorman, T., … Zimmerman, J. (2017). What is an intensive care unit? A report of the task force of the World Federation of Societies of Intensive and Critical Care Medicine. *Journal of Critical Care*, *37*, 270–276. <https://doi.org/10.1016/j.jcrc.2016.07.015>

# Apéndice

## Apéndice 1. Repositorio del proyecto

El código fuente desarrollado para este proyecto está disponible en el repositorio GitHub: https://github.com/bernatsort/Therapy-Progress-Gait-Prediction.git

# APUNTES EXTRA

## Ciclo de la marcha vs paseo

Sí, hay una diferencia entre "gait cycle" y "walk" (o paseo) en el contexto del análisis de la marcha.

El "gait cycle" (ciclo de la marcha) se refiere al período completo desde el momento en que un pie toca el suelo hasta el siguiente contacto del mismo pie. El ciclo de la marcha se divide en dos fases principales: la fase de apoyo y la fase de balanceo. Durante la fase de apoyo, el pie está en contacto con el suelo y soporta el peso del cuerpo, mientras que durante la fase de balanceo, el pie se levanta del suelo y se mueve hacia adelante para preparar el próximo contacto.

El "walk" o "paseo" se refiere a la acción de caminar de manera general. Es un término más amplio que se utiliza para describir el acto de desplazarse a pie. El caminar implica el uso repetitivo del ciclo de la marcha, pero también puede incluir variaciones en la velocidad, el estilo, la postura y otros factores específicos de cada individuo.

En resumen, el "gait cycle" se refiere específicamente al ciclo completo de la marcha, mientras que "walk" o "paseo" es un término más amplio que abarca la acción general de caminar e incluye el uso del ciclo de la marcha, pero con variaciones individuales.

No, las caminatas y los ciclos de la marcha son conceptos diferentes en el análisis de la marcha.

- Caminatas: Se refiere a los pasos o caminatas individuales que realiza un individuo durante una sesión de evaluación o grabación de la marcha. Cada caminata se compone de una serie de pasos consecutivos y se considera una unidad discreta de movimiento. Durante una caminata, se registran datos sobre la marcha, como la posición de las articulaciones, el movimiento de los segmentos corporales y otros parámetros relevantes.

- Ciclos de la marcha: Un ciclo de la marcha se refiere a un ciclo completo de movimiento desde un evento inicial hasta el mismo evento en la siguiente ocasión. En el análisis de la marcha, se suelen considerar dos ciclos de la marcha para obtener mediciones precisas y representativas. Estos eventos clave que marcan el comienzo y el final del ciclo de la marcha pueden variar, pero comúnmente se definen como el momento en que el talón toca el suelo (inicio del ciclo) hasta el próximo momento en que el mismo talón toca el suelo (final del ciclo).

En resumen, una caminata se compone de múltiples pasos, mientras que un ciclo de la marcha se refiere a un ciclo completo de movimiento desde un evento clave hasta el mismo evento en la siguiente ocasión. En el análisis de la marcha, los ciclos de la marcha se utilizan para obtener mediciones y realizar análisis detallados de los parámetros cinemáticos y cinéticos, mientras que las caminatas representan el conjunto de pasos realizados durante una sesión de evaluación o grabación de la marcha.

En el análisis de la marcha, una caminata generalmente se considera como un conjunto de ciclos de la marcha. Sin embargo, es importante tener en cuenta que el número de ciclos puede variar dependiendo del contexto y la definición utilizada.

En situaciones típicas, una caminata completa o una pasada de caminata se compone de varios ciclos de la marcha consecutivos. Cada ciclo de la marcha representa un ciclo completo de movimiento desde un evento clave hasta el mismo evento en la siguiente ocasión, como se mencionó anteriormente.

Si tienes 40 ciclos registrados, es posible que representen múltiples caminatas realizadas por el individuo durante la sesión de evaluación o grabación de la marcha. Cada caminata puede estar compuesta por varios ciclos de la marcha, y el número exacto de ciclos por caminata puede variar dependiendo de la duración de la caminata y las características individuales del paciente.

En resumen, una caminata se considera un conjunto de ciclos de la marcha, y el número de ciclos puede variar dependiendo de la duración y las características de la caminata. Si tienes 40 ciclos, es posible que representen varias caminatas realizadas durante la sesión de evaluación o grabación de la marcha.

## Ciclos

En el contexto de los datos cinemáticos que has proporcionado, el término "cycles" se refiere al número de ciclos de la marcha que se han registrado. Según los datos, se han registrado un total de 40 ciclos de la marcha.

Cada ciclo de la marcha representa un ciclo completo de los movimientos y posiciones de las articulaciones durante una caminata. El ciclo de la marcha incluye tanto la fase de apoyo (cuando el pie está en contacto con el suelo) como la fase de balanceo (cuando el pie se levanta del suelo y se mueve hacia adelante).

En tus datos, cada fila corresponde a una muestra específica durante los 40 ciclos registrados. Los valores en las columnas representan los ángulos de diferentes articulaciones en cada muestra. Cada ángulo está asociado con una articulación específica, como la rodilla (K), la cadera (H), el tobillo (A), etc. Los valores de los ángulos se registran en grados (deg) y varían en cada muestra del ciclo de la marcha.

## Sample

En el contexto de los datos que proporcionaste, la columna "sample" representa el número de muestra dentro de los ciclos de la marcha registrados. El número de muestra indica la secuencia de cada punto de datos tomado durante el movimiento de la marcha.

En tu caso, los datos muestran 101 muestras en total, numeradas desde 0 hasta 100. Cada muestra corresponde a un punto específico en el tiempo durante los 40 ciclos de la marcha registrados. Es posible que estas muestras se hayan tomado en intervalos de tiempo uniformes o en momentos específicos durante la caminata para capturar la información cinemática en diferentes fases del ciclo de la marcha.

La columna "sample" es útil para identificar y hacer un seguimiento de cada punto de datos específico en relación con el ciclo de la marcha y puede ser utilizado para el análisis posterior y la interpretación de los datos cinemáticos registrados.

## Medias

La columna "aRAFE.M" en tus datos representa la media de los ángulos de flexión-extensión de la articulación del tobillo derecho (Right Ankle Flexion-Extension) en cada muestra de los ciclos de la marcha registrados.

La media (o promedio) en este contexto se refiere al valor promedio de los ángulos de flexión-extensión del tobillo derecho en un conjunto de datos específico. Para calcular la media, se suman todos los valores de los ángulos de flexión-extensión del tobillo derecho en una muestra y luego se divide entre el número total de muestras.

La media puede proporcionar una medida representativa o central de los ángulos de flexión-extensión del tobillo derecho durante los ciclos de la marcha registrados. Puede ser útil para analizar la tendencia general o el comportamiento promedio de la articulación del tobillo derecho en el conjunto de datos proporcionado.

## Corte de ciclos

En el análisis de la marcha, los ciclos de la marcha pueden ser identificados y segmentados por un especialista en base a criterios específicos. En general, el proceso implica la identificación de eventos clave en la marcha, como el contacto inicial del pie con el suelo (inicio del ciclo) y el siguiente contacto del mismo pie (final del ciclo).

El especialista puede utilizar diferentes métodos y tecnologías para identificar estos eventos, como el análisis visual de vídeos de la marcha, el uso de sensores de presión en los pies, la captura de movimiento tridimensional o el uso de sistemas de análisis de la marcha.

Una vez identificados los eventos clave, los ciclos de la marcha pueden ser segmentados de manera precisa. Estos segmentos son utilizados posteriormente para realizar mediciones y análisis detallados de los parámetros de la marcha, como los ángulos articulares, las fuerzas de reacción del suelo, la duración de las fases de la marcha, entre otros.

Es importante destacar que el proceso de segmentación de los ciclos de la marcha requiere experiencia y conocimiento del especialista para garantizar una identificación correcta de los eventos y una segmentación adecuada de los ciclos. Esto permite un análisis preciso de los datos cinemáticos y cinéticos de la marcha.

## Proceso

genial! Podría describirme el procese des de que el niño con paralisis cerebral llega a la consulta, se le colocan los marcadores por el cuerpo y se graba su forma de andar, para luego extraer los datos cinematicos?

ChatGPT:

Por supuesto, aquí te describo el proceso general desde que un niño con parálisis cerebral llega a la consulta hasta que se obtienen los datos cinemáticos de su forma de andar:

1. **Evaluación inicial:**

El especialista en marcha o el terapeuta evalúa al niño para comprender su historial médico, síntomas, limitaciones funcionales y objetivos de tratamiento. También se pueden realizar pruebas físicas y evaluaciones clínicas para obtener información sobre la función motora y la marcha actual del niño.

1. **Preparación del paciente:**

Para el registro de la marcha, se le pide al niño que use ropa ajustada y cómoda, y se le colocan marcadores o reflectores en puntos anatómicos clave del cuerpo. Estos marcadores ayudan a rastrear y registrar los movimientos y la posición de las articulaciones durante la marcha.

1. **Registro de la marcha:**

El niño camina en un área designada mientras se graba su forma de andar utilizando sistemas de captura de movimiento. Estos sistemas pueden incluir cámaras de video, sensores inerciales, sensores de presión en los pies u otras tecnologías avanzadas que capturan datos cinemáticos y cinéticos de la marcha.

1. **Procesamiento de los datos:**

Una vez completada la sesión de grabación, los datos registrados se procesan utilizando software especializado. Este software rastrea los marcadores en el cuerpo del niño y extrae información cinemática, como los ángulos articulares en diferentes articulaciones durante los ciclos de la marcha.

1. **Análisis e interpretación de los datos:**

Los datos cinemáticos obtenidos se analizan para evaluar la calidad y las características de la marcha del niño. Los especialistas pueden examinar los patrones de movimiento, la amplitud de movimiento de las articulaciones, la coordinación, el equilibrio y otros aspectos relevantes para comprender las limitaciones y las necesidades de tratamiento del niño.

1. **Plan de tratamiento:**

Con base en los datos cinemáticos y la evaluación completa, se desarrolla un plan de tratamiento individualizado para el niño. Este plan puede incluir terapia física, ejercicios específicos, dispositivos de asistencia, ortesis u otras intervenciones destinadas a mejorar la función motora y la marcha del niño.

You:

es correcto el siguiente proceso? añadirías algo más?

Evaluación inicial:

El niño acude a la consulta y el especialista lo evalúa comprender su historial médico, síntomas, limitaciones funcionales y objetivos de tratamiento. También se realizan pruebas físicas y evaluaciones clínicas para obtener información sobre la función motora y la marcha actual del niño.

Preparación del paciente:

Para poder realizar el registro de la marcha, se le pide al niño que use ropa ajustada y cómoda, y se le colocan los marcadores en puntos anatómicos clave del cuerpo. Estos marcadores ayudan a rastrear y registrar los movimientos y la posición de las articulaciones durante la marcha.

Registro de la marcha:

El niño camina en un área designada mientras se graba su forma de andar utilizando sistemas de captura de movimiento o fotogrametría. Estos sistemas pueden incluir cámaras de video, sensores inerciales, sensores de presión en los pies u otras tecnologías avanzadas que capturan datos cinemáticos y cinéticos de la marcha.

Datos y su procesamiento

Los datos se capturan del sistema de análisis de movimiento (nombre del sistema ¿SMARTanayzer -BTS S.p.a.?) que funciona a 250 Hz con 8 cámaras.

El archivo inicial (.tdf) contiene gran cantidad de datos no anonimizados.

Se etiquetan los marcadores para pasar de tener trayectorias en x,y,z y unas matrices de datos brutos a tener datos etiquetados.

A continuación, el técnico del laboratorio corta los datos etiquetados en ciclos. Es decir, se define algo que clínicamente es relevante.

Posteriormente, el médico o especialista selecciona los ciclos que considera más importantes o significativos, y los demás se descartan. A partir de los ciclos seleccionados se obtienen los valores promedios de los ángulos de cada articulación en cada muestra de los ciclos de la marcha registrados.

En este punto se han excluido los datos de carácter personal, por tanto, tenemos los datos anonimizados.

A partir de los ciclos procesados, se obtiene la cinemática en tres dimensiones de las articulaciones, que se utilizará en el proceso de toma de decisiones.

Los datos cinemáticos se componen por grados – tiempo. Por tanto, vamos a tener 11 gráficas con 101 valores cada gráfica (101 valores de cada miembro inferior). Tenemos 2 piernas representadas en cada gráfica con 101 valores por pierna. Esos 101 valores suelen ser la media de un par de estudios o 3 que hemos incluido. Es un conjunto bastante grande de datos.

ChatGPT:

El proceso descrito parece ser una secuencia adecuada para la evaluación y registro de la marcha. Aquí hay algunos puntos adicionales que podrían considerarse:

1. Calibración y configuración del sistema: Antes de comenzar el registro de la marcha, es importante realizar una calibración precisa del sistema de captura de movimiento y configurar correctamente los parámetros del equipo utilizado. Esto garantiza una captura de datos precisa y confiable.

2. Análisis de la marcha: Una vez que se han obtenido los datos cinemáticos y cinéticos de la marcha, se realiza un análisis detallado de los patrones de movimiento, las características de la marcha y cualquier anomalía o alteración en los parámetros medidos. Esto puede implicar el uso de software especializado para analizar los datos y generar informes.

3. Interpretación de los resultados: El médico o especialista interpreta los resultados obtenidos del análisis de la marcha, teniendo en cuenta el historial médico del paciente, los objetivos de tratamiento y las características individuales. Esto puede incluir la identificación de patrones anormales de movimiento, áreas de mejora o intervención específica requerida.

4. Plan de tratamiento: Con base en los resultados y la interpretación del análisis de la marcha, se desarrolla un plan de tratamiento individualizado para el paciente. Esto puede implicar intervenciones terapéuticas, ejercicios de rehabilitación, dispositivos de asistencia o cualquier otro enfoque adecuado para mejorar la función y la calidad de vida del paciente.

Es importante tener en cuenta que el proceso puede variar ligeramente dependiendo del sistema de análisis de movimiento utilizado, las prácticas clínicas específicas y las necesidades del paciente. Por lo tanto, es fundamental seguir los protocolos y pautas establecidas por el equipo médico y los especialistas involucrados en el análisis de la marcha.

# Correo sergio

Buenas tardes, Sergio.

Te envío un documento Word con varías dudas que me han surgido. Además, te adjunto un párrafo describiendo el proceso de cómo se obtienen los datos y su descripción para que corrijas lo que consideres que está mal o añadas información faltante importante.

Muchas gracias.

Saludos,

Bernat

# Dudas sergio

1. Los ficheros .emt que tipo de ficheros son? ¿Como los podría describir? ¿Los proporciona algún software en específico?
2. Desde que el sistema obtiene los datos más brutos hasta los archivos .emt que envías, cuál ha sido el procesamiento de los datos?
3. En el listado de intervenciones hay varias intervenciones, pero al principio dijiste que la operación era la misma para todos los pacientes: osteotomía extensora de rodilla. ¿Significa que ahora hay los pacientes se han sometido a distintas intervenciones?
4. La columna sample de los datos: ¿Estas muestras se han tomado en intervalos de tiempo uniformes o en momentos específicos durante la caminata para capturar la información cinemática en diferentes fases del ciclo de la marcha?
5. ¿Los valores promedios de los ángulos de las diferentes articulaciones, son los valores promedios de los paseos o de los ciclos?
6. En los primeros ficheros que enviaste había 11 ángulos. En los últimos datos que me enviaste, hay 20. De esos 20 ángulos diferentes, ¿cuáles nos interesan para nuestro estudio?
7. ¿Están bien las siguientes definiciones? ¿Podrías completar las que faltan, ya que no he conseguido saber que son? (solo he incluido las de la pierna derecha para no repetirme).

* **Tobillo:** 
  + aRAFE.M: promedio de los ángulos de flexión y extensión del tobillo derecho.
  + aRAIE.M: promedio de los ángulos de inversión y eversión del tobillo derecho.
* **Rodilla:** 
  + aRKFE.M: promedio de los ángulos de flexión y extensión de la rodilla derecha.
  + aRKAA.M: promedio de los ángulos de abducción y aducción de la rodilla derecha.
  + aRKIE.M: promedio de los ángulos de inversión y eversión de la rodilla derecha.
* **Cadera:**

* + aRHPFE.M: promedio de los ángulos de flexión y extensión de la cadera derecha.
  + aRHPAA.M: promedio de los ángulos de abducción y aducción de la cadera derecha.
  + aRHPIE.M: promedio de los ángulos de rotación interna y externa de la cadera derecha.
* **Pelvis:** 
  + aRPTILT.M: promedio de los ángulos de inclinación pélvica de la cadera derecha.
  + aRPOBLI.M: promedio de los ángulos de oblicuidad pélvica de la cadera derecha.
  + aRPROT.M: promedio de los ángulos de rotación pélvica de la cadera derecha.
* **Hombro:** 
  + aRSHROT.M: promedio de los ángulos de rotación del hombro.
  + aRSPML.M: promedio de los ángulos de flexión y extensión del hombro derecho.
  + aRSPIE.M: promedio de los ángulos de abducción y aducción del hombro derecho.
  + aRSHTILT.M: promedio de los ángulos de inclinación de la articulación del hombro derecho.
  + aRSHOBLI.M: promedio de los ángulos de oblicuidad de la articulación del hombro derecho.
  + aRSHTILTOFF.M: ¿?
  + aRSHOBLIOFF.M: ¿?
  + aRSHROTOFF.M: ¿?
  + aRSPFE.M: ¿?

# chatGPT

## MODELS

I am trying to build a model with multiple inputs and multiple outputs. I have 101 inputs and 101 outputs, as I try to predict the last session of a patient's rehabilitation from the first session. The 101 inputs and outputs are the same variables but time-normalized to 100% stance (i.e. 101 points).

Description of the data stored in the “GRF\_\*.csv” files. “\*” for the associated file name is a placeholder for “right” and “left”. n is a time-normalized vector of 101 points for the post-processed (“PRO”) data. Note that the first three columns of each file hold the SUBJECT\_ID, SESSION\_ID, and TRIAL\_ID.

I want to apply 4 models for each algorithm. Let me explain this. For each algorithm, I want to make a first model with the 101 input and output (base model), a second model using PCA on X (reducing the X), a third model reducing the "y" (PCA on y), and a fourth model reducing X and y (PCA on X and y). Then, for the reduced y in models 3 and 4, I return to the original feature space.

So I used the following algorithms:

The first thing I did was a multiple linear regression with leave one out cross-validation, and for the 4 models, I got good results! Now, apart from performing a decision tree, then a random forest and finally an xgboost, I though to also perform a PLS.