



BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

COMPLEJO REGIONAL CENTRO SEDE SAN JOSÉ CHIAPA

Título de la tesis:

**Aplicación de técnicas de análisis de datos en terapias a
pacientes que han sufrido EVC**

Tesis que para obtener el título de:

**LICENCIADO EN INGENIERÍA EN SISTEMAS Y TECNOLOGÍAS DE
LA INFORMACIÓN INDUSTRIAL**

Presenta:

Juan Galicia Sánchez

Director de Tesis:

María Claudia Delicia Carral

San José Chiapa, Puebla, 05 de enero del 2023

Agradecimientos

A mi querida familia, no hay palabras suficientes para agradecerles todo lo que han hecho por mí. Gracias por su amor incondicional, paciencia, apoyo y sacrificio para hacer posible mi educación. Ustedes son los pilares de mi vida y esta tesis es su logro tanto como el mío.

Índice

Índice de Tablas	III
Índice de Figuras	IV
Resumen	1
1 Introducción.....	2
1.1 Antecedentes del Proyecto	2
1.2 Planteamiento del Problema.....	5
1.3 Objetivos General y Específicos del Proyecto	5
1.3.1 Objetivo General.	5
1.3.2 Objetivos Específicos.	5
1.3.3 Preguntas de investigación	6
1.4 Alcances y limitaciones.....	6
1.5 Organización de la tesis.....	6
2 Marco Histórico	8
2.1 Introducción	8
2.2 Estado del arte.....	9
3 Marco Teórico.....	12
3.1 Metodología CRISP-DM	12
3.2 Técnicas de preparación de los datos	13
3.3 Algoritmos de minería de datos	14
3.3.1 Algoritmo de regresión lineal.....	16
3.4 Complicaciones de la enfermedad vascular cerebral	17
3.5 Rehabilitación de la enfermedad vascular cerebral.....	18

3.6	Juegos serios.....	20
3.7	Motivación.....	21
4	Análisis de datos de rehabilitación por medio de minería de datos	23
4.1	Introducción	23
4.2	Comprensión del dominio del estudio y establecimiento de objetivos.....	23
4.2.1	Herramientas utilizadas.....	24
4.3	Recolección, limpieza y preprocesamiento de los datos	25
4.3.1	Descripción del conjunto de datos	25
4.4	Análisis de los datos	32
4.5	Proceso de minado a través de regresión lineal.....	50
5	Conclusiones.....	52
5.1	Conclusiones	52
5.2	Trabajo a futuro.....	53
	Referencias.....	54

Índice de Tablas

Tabla 1. Estado del arte.....	10
Tabla 2. Diccionario de datos	26
Tabla 3. Resumen de datos.....	27
Tabla 4. Paciente P06.....	28
Tabla 5. Paciente P07.....	28
Tabla 6. Paciente P08.....	28
Tabla 7. Paciente P09.....	29
Tabla 8. Paciente P10.....	29
Tabla 9. Paciente P11.....	29
Tabla 10. Cálculo de variables.....	30
Tabla 11. Resumen encuestas totales.....	31
Tabla 12. Encuestas por paciente	31
Tabla 13. P06 - Tabla de frecuencia.....	47
Tabla 14. P07 – Tabla de frecuencia	47
Tabla 15. P08 – Tabla de frecuencia	48
Tabla 16. P09 – Tabla de frecuencia	48
Tabla 17. Matriz de correlación.....	51
Tabla 18. Coeficiente e intersección de regresión lineal	51
Tabla 19. Matriz de correlación.....	51

Índice de Figuras

Figura 1 Gesture Therapy	4
Figura 2. Etapas de CRISP-DM, imagen obtenida de (Takeshi, s.f.)	12
Figura 3. Algoritmo de clasificación Imagen obtenida de (Buitrago, 2020).....	15
Figura 4. Algoritmo de asociación Imagen obtenida de (Alcalde, 2018).....	15
Figura 5. Algoritmo de agrupamiento Imagen obtenida de (Data Science Team, 2020).....	16
Figura 6. Ejemplo de regresión lineal	17
Figura 7. Juegos serios, imagen obtenida de (MedicalExpo, s.f.)	21
Figura 8. Escala de motivación intrínseca	31
Figura 9. Preparación de los datos de encuestas en RapidMiner	33
Figura 10. Gráfica de dolor por género	35
Figura 11. Gráfica de dolor por paciente	36
Figura 12. Gráfica de esfuerzo/importancia e interés/disfrute por género	37
Figura 13. Gráfica esfuerzo/importancia por paciente	38
Figura 14. Gráfica de interés/disfrute por paciente	39
Figura 15. Gráfica de valor/utilidad por paciente	40
Figura 16. Gráfica de presión/tensión por paciente	41
Figura 17. Dolor vs interés/disfrute por paciente	42
Figura 18. Valor/utilidad vs interés/disfrute por paciente	42
Figura 19. Gráficas del paciente P06, a) Distancia recorrida, b) Presión ejercida, c) Velocidad	43
Figura 20. Gráficas del paciente P07. a) Presión b) Distancia por sesión c) Distancia promedio	44
Figura 21. Gráfica de variables del paciente P07 a) Velocidad total por sesión b) Presión por sesión c) Distancia de presión por sesión	45
Figura 22. P08 – Promedio de presión por sesión.....	46
Figura 23. Distancia total por paciente	49
Figura 24. Varianza de la distancia por paciente	49
Figura 25. Diseño de regresión lineal	50

Resumen

En este trabajo de investigación se ha realizado un proceso de análisis de datos empleando la metodología CRISP-DM, la cual abarca la interpretación y entendimiento de los datos obtenidos. Para ello se ha realizado una investigación exhaustiva del tema, una vez comprendiendo los alcances y objetivos del estudio, se hizo una limpieza profunda de los datos para poder procesarlos, empleando diversas técnicas y herramientas dedicadas a este proceso como lo es rapidminer, se han modelado y graficado los datos haciendo uso de distintos tipos de gráficas apoyándose de herramientas como weka, powerbi y rapidminer, facilitando la visualización e interpretación de la información obtenida, algunos de los algoritmos y técnicas más relevantes aplicadas en esta investigación es la matriz de correlación, para confirmar y cuantificar la relación entre las variables involucradas en el tratamiento de los pacientes y el algoritmo de regresión lineal, permitiendo mostrar de manera más definida la relación existente entre estas variables.

1 Introducción

1.1 Antecedentes del Proyecto

En la actualidad, la Enfermedad Vascul ar Cerebral (EVC) se ha convertido en una de las primeras causas de muertes en el mundo, de hecho, de acuerdo con la (Organización Mundial de la Salud (OMS), 2015) es considerada la segunda causa global de muerte en el mundo, siendo los países de ingresos medios y bajos, los más afectados.

Se clasifica en 2 tipos, la isquémica aparece cuando el flujo sanguíneo en el cerebro se interrumpe por una obstrucción, que puede ser causada por grasa o sangre en un vaso sanguíneo, por otro lado, la hemorrágica se presenta cuando un vaso sanguíneo se rompe, causando un sangrado dentro del cerebro.

La (Organización Mundial de la Salud (OMS), 2015) también menciona los síntomas más recurrentes de EVC, así como los factores de riesgo que se presentan a continuación:

- Alteración repentina de la visión en un ojo o ambos
- Pérdida repentina de la fuerza en un brazo, una pierna o ambos
- Sensación de hormigueo en la cara, brazo o pierna
- Aparición repentina de:
- Problemas para hablar y/o entender lo que se escucha, acompañada por balbuceo
- Desequilibrio o inestabilidad
- Dolor de cabeza

Con el paso de la edad, el riesgo de padecer EVC aumenta, presentándose mayoritariamente en hombres. Algunas de las acciones que contribuyen al aumento de probabilidades de sufrir un daño de esta naturaleza es el tabaquismo, el alcoholismo, la inactividad física y la obesidad. Por otra parte, padecer presión

arterial alta, diabetes mellitus y niveles de colesterol elevados también forman de los factores para favorecer el padecimiento de EVC. Cabe mencionar que el riesgo aumenta considerablemente mientras mayor número de factores de riesgo se presenten.

De acuerdo con (Mayo Clinic, 2021), si el accidente cerebrovascular afectó el lado derecho del cerebro, el movimiento y la sensibilidad del lado izquierdo del cuerpo pueden verse afectados. Si el accidente cerebrovascular dañó el tejido cerebral del lado izquierdo del cerebro, el movimiento y la sensibilidad del lado derecho del cuerpo pueden verse afectados. El daño cerebral en el lado izquierdo del cerebro también puede causar trastornos del habla y del lenguaje.

También denota que la mayoría de los sobrevivientes de un derrame cerebral van a rehabilitación y, después del tratamiento de emergencia, el cuidado del accidente cerebrovascular se enfoca en ayudar a recuperar las mayores funciones posibles, esto con terapias físicas dentro del tratamiento.

Existen diferentes tipos de terapias de rehabilitación apoyadas por la tecnología para un mejor desempeño. Una de ellas, conocida como Gesture Therapy, fue desarrollada por especialistas del Instituto Nacional de Astrofísica, Óptica y Electrónica para rehabilitar los movimientos de las extremidades superiores en pacientes que han padecido EVC (INAOE, 2018). Gesture Therapy fue uno de los 20 proyectos, de diferentes partes del mundo, exhibidos en un foro que organiza la fundación de la Academia Francesa de Medicina. El evento se realizó en la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (Unesco), en París (INAOE, 2018).

Gesture Therapy, consta de una computadora, una cámara y una manija o gripper, ver Figura 1, para que el paciente realice sus terapias a través de juegos por computadora que emulan situaciones cotidianas como cocinar un alimento, cambiar una llanta o ahuyentar un insecto, reveló la Agencia Informativa del Consejo Nacional de Ciencia y Tecnología de México (INAOE, 2018).



Figura 1 Gesture Therapy

El encargado del proyecto Dr. Luis Enrique Sucar Succar comenta que *“Gesture Therapy consiste en una serie de juegos en las cuales el paciente interactúa a través de una manija que se desarrolló especialmente para medir el movimiento en tres dimensiones y la presión que se ejerce sobre ella. El sistema está enfocado en ejercitar los movimientos de las manos y de los brazos, e incluye ejercicios de abducción, prensión y rotación”* (Campos, 2016). Además, expresa que mediante la interacción del paciente con el videojuego se mejoran las técnicas de rehabilitación porque hacen énfasis en movimientos determinados que permiten una recuperación más efectiva y rápida.

Gesture Therapy registra los movimientos de los pacientes en tres dimensiones además de la presión ejercida, estos datos serán aprovechados en este trabajo para analizarlos desde un enfoque de minería de datos que permitirá obtener patrones que proporcionen información sobre el avance de los pacientes cuando utilizan el sistema.

Al igual que en otros ámbitos, el objetivo del análisis de datos es examinar información con el propósito de obtener conocimiento. Este conocimiento puede tener un valor científico o investigador que ayude a determinar causas o patrones de determinadas patologías o a identificar poblaciones de riesgo, y así ayudar en la detección o diagnóstico de distintas enfermedades afirma (STIMULUS, 2017).

En este proyecto, la aplicación de técnicas de análisis de datos servirá como apoyo en las terapias de rehabilitación de pacientes que han sufrido EVC.

1.2 Planteamiento del Problema

Los datos recabados por Gesture Therapy en las terapias de los pacientes que sufrieron EVC se encuentran almacenados en un gran repositorio y se requiere aprovechar al máximo la información almacenada. Estos datos son: la posición de la manija en 3 ejes y en cada frame durante las terapias y la presión que se ejerce sobre el dispositivo, además de una serie de encuestas que se realizan al finalizar cada terapia a los pacientes.

Dado el gran volumen de datos acumulado en él, y la incapacidad de los especialistas de identificar patrones de comportamiento y extraer conocimiento oculto en los datos almacenados para apoyar sus decisiones, surge la necesidad de aplicar análisis de datos.

En este trabajo se expone la explotación de datos recabados por Gesture Therapy durante las terapias mediante el uso de técnicas de minería de datos para contribuir al análisis de los datos generados en las terapias de pacientes que han sufrido EVC.

1.3 Objetivos General y Específicos del Proyecto

1.3.1 Objetivo General.

Aplicar técnicas de análisis de datos para identificar patrones, describir comportamientos y en general, sacar el máximo provecho a los datos almacenados y así brindar información de apoyo en las terapias de rehabilitación de pacientes que han sufrido EVC.

1.3.2 Objetivos Específicos.

- Describir a partir de la información recolectada por Gesture Therapy y los terapeutas los factores que influyen tanto positiva como negativamente en el

tratamiento de los pacientes para así mejorar la toma de decisiones en su rehabilitación.

- Selección y aplicación de algoritmos estadísticos que de acuerdo con las necesidades del proyecto y a su vez, promuevan una mejor explotación de los datos.
- Evaluar el avance de los pacientes de acuerdo con las características que tuvieron durante su tratamiento para relacionar los comportamientos con el resultado de sus terapias.

1.3.3 Preguntas de investigación

¿Qué información puede ser descubierta a partir de los registros del sistema Gesture Therapy que sea de apoyo en la rehabilitación de los pacientes con EVC?

1.4 Alcances y limitaciones

Aplicación de técnicas de análisis de datos que permita mejorar la explotación de la información que se obtenga a partir de los datos recopilados por Gesture Therapy en las terapias de los pacientes que han sufrido EVC. A partir de este análisis apoyar la toma de decisiones en ambas partes, médicos terapeutas y los pacientes.

Esta investigación se limita a la cantidad de los datos obtenidos de 6 pacientes y al número de juegos practicados durante sus terapias.

1.5 Organización de la tesis

La tesis se encuentra organiza de la siguiente manera:

En el capítulo 2 el marco histórico y el estado del arte, en este capítulo se repasan los trabajos e investigaciones relacionados a la presente investigación, de igual manera se resaltan las características más importantes y relacionadas entre la investigación y los proyectos mencionados en este capítulo.

En el capítulo 3 se presenta el marco teórico donde se describen conceptos y herramientas importantes para desarrollar esta investigación.

En el capítulo 4 se presenta el marco metodológico, en este capítulo se encuentra la metodología empleada para el proyecto y sus fases que la componen, además de los resultados obtenidos en cada una.

En el capítulo 5 se presentan las conclusiones, donde a partir de la información obtenida ejecutando la metodología descrita en el capítulo 4, se pueden obtener diversas conclusiones acerca de cada uno de los casos de los pacientes tratados con Gesture Therapy.

2 Marco Histórico

2.1 Introducción

El accidente cerebrovascular es la principal causa de discapacidades motoras y cognitivas que requieren terapia en el mundo, por lo tanto, es importante desarrollar tecnología de rehabilitación que permita a las personas que han sufrido un accidente cerebrovascular practicar un entrenamiento de movimiento intensivo sin el gasto de un terapeuta siempre presente.

Se ha desarrollado un sistema de visión de bajo costo que permite a los sobrevivientes de un accidente cerebrovascular practicar ejercicios de movimiento de brazos en casa o en la clínica con interacciones periódicas con ayuda de terapeuta. El sistema integra un entorno virtual para facilitar el entrenamiento de movimientos repetitivos, con algoritmos de visión por computadora que rastrean la mano del paciente, utilizando una cámara económica y una computadora personal. Este sistema, llamado Gesture Therapy, incluye una pinza con sensor de presión para incluir rehabilitación de manos y dedos; y rastrea la cabeza del paciente para detectar y evitar la compensación del tronco, además, sido evaluado en un ensayo clínico controlado en el Instituto Nacional de Neurología y Neurocirugía de la Ciudad de México, llegando a compararse con la terapia ocupacional convencional.

La investigación y el desarrollo de herramientas para analizar grandes volúmenes de datos se han hecho cada vez más necesarios. Como resultado, surgieron dos metodologías o técnicas relacionadas entre sí, aunque con diferentes objetivos: “data warehousing” y minería de datos. El “data warehousing” es un proceso de almacenamiento de datos históricos para su posterior análisis mediante reportes y tableros de comando que permite ver los movimientos a través del tiempo de las principales variables de un dominio. El “data warehouse”, si bien no de manera indispensable, se convierte en una buena fuente de datos para realizar minería. La minería de datos puede realizarse a partir de un fichero pequeño. No obstante, las ventajas aumentan con grandes volúmenes de datos, siendo afortunadamente un número considerable. La minería de datos puede definirse como el procesamiento

de los datos para encontrar patrones de comportamiento que sean de utilidad para la toma de decisiones. Se relaciona de manera estrecha con la estadística, usando técnicas de muestreo y visualización de datos, depuración y cálculo de indicadores.

Aunque los objetivos de minería en las bases de datos comerciales varían con el área de negocio, en general se orientarán a determinar qué o quién realiza o posee qué cosa bajo qué circunstancia. A nivel de investigación básica se trabaja en el mejoramiento de los algoritmos de minería de datos existentes y en el desarrollo de nuevos algoritmos, mientras que a nivel de investigación aplicada se trabaja en descubrir patrones de comportamiento de datos en diferentes áreas. Por ejemplo, en el área de medicina, actividad conocida como informática biomédica, como lo es en este caso en particular.

2.2 Estado del arte

Para el estado del campo se encontraron trabajos relacionados con el proyecto que se realizará considerando diferentes perspectivas de desarrollo como las metodologías utilizadas, las técnicas de minería, los datos utilizados, los resultados obtenidos, etc. A continuación, se presentan algunos trabajos encontrados y una breve descripción de cada uno.

Artículo con citado APA	Datos utilizados	Metodología	Resultados obtenidos	Algoritmos aplicados
(Devesa Gutiérrez et al., 2014)	Edad Antecedentes médicos Tipo de EVC Tipo de tratamiento aplicado Partes del cuerpo afectadas Patrones de movimiento	No aplica	Es un hecho que la rehabilitación afecta positivamente a la supervivencia y a la capacidad funcional y recuperación de los pacientes. Este artículo sirve como guía sobre las distintas deficiencias que se provocan debido a la EVC, de los tipos de tratamientos que hay y los objetivos que se persiguen con cada uno.	No aplica
(Sánchez-Villavicencio et al., 2009)	Número de pacientes desconocidos, datos recolectados de 25 sesiones de tratamiento, distribuidas tres veces por semana y con una duración aproximada de una hora por sesión. Variables de velocidad y patrones de movimiento.	No aplica	Los resultados obtenidos mostraron que los pacientes del grupo en estudio estuvieron más satisfechos con su desempeño durante el entrenamiento, percibiéndose mayor concentración durante la actividad; asimismo, este grupo refirió mayor interés por el sistema de realidad virtual que los pacientes a los que se les realizaba el programa de terapia ocupacional en casa.	No aplica
(Antonio Gutiérrez Pérez et al., n.d.)	Se analizaron 288 datos correspondiente a 22 expedientes clínicos de pacientes. Las variables evaluadas fueron género, año, edad, diagnóstico y síntoma.	Metodología utilizada (KDD) para minería de datos	Índice de mortalidad con el tipo de cáncer. Intensidad de dolor con respecto a la edad de los pacientes. Índice de supervivencia en cada grupo de edad. La intensidad de dolor por tipo de cáncer.	Algoritmos aplicados para el estudio algoritmo j48 correspondiente e al algoritmo C.4.5
(Dávila Hernández et al., n.d.)	Expediente clínico de los pacientes Número de registros desconocido Antecedentes Familiares con enfermedades Crónicas Etnia	Metodología utilizada (CRISP-DM: Cross Industry Standard Process for Data Mining) para minería de datos	En el 40% de los casos de los pacientes que padecen hipertensión arterial están entre 45 y 65 años, son de sexo masculino y de raza mestiza. - El 34% de las personas que padecen hipertensión arterial tienen antecedentes patológicos familiares de la enfermedad y consumen tabaco. - El 66 % de los casos con hipertensión arterial fueron asintomáticos.	Árboles de decisión J48 y simple K means. Combinación de modelos matemáticos
(Sucar et al., 2014)	Para evaluar el proyecto se realizó una prueba piloto con seis pacientes durante un mes y recolectaron sus datos para su evaluación. Los datos utilizados fueron patrones de movimiento y variables de velocidad.	Obtención de los datos Procesamiento de los datos Interpretación de los datos Evaluación de la información y obtención de conocimiento	La tecnología utilizada proporciona un valor enorme al proyecto, pues mejora la motivación de los pacientes y el uso del sistema, lo que conlleva a una rehabilitación más eficaz y satisfactoria. Tal como se esperaba, variable como la velocidad, la precisión y la motivación de los pacientes incrementó mucho respecto a sus terapias de rehabilitación convencionales.	POMDP para adaptar el nivel de dificultad de la velocidad y desviación de trayectorias de movimiento suave en los juegos serios.

Tabla 1. Estado del arte

Diversos estudios como los de (Antonio Gutiérrez Pérez et al., n.d.), (Dávila Hernández et al., n.d.) o el doctor (De Cáncer et al., n.d.), han empleado metodologías o partes de éstas para el desarrollo de sus proyectos, específicamente haciendo referencia a los proyectos antes mencionados, quedando en evidencia la relevancia y utilidad que las metodologías aplicadas en procesos de minería y análisis de datos, así como en otras áreas, considerando que los datos obtenidos en estos estudios son de pacientes que han tenido que pasar por un proceso de rehabilitación similar al que tuvieron que pasar los pacientes de los cuales se obtuvieron los datos.

Entre los algoritmos utilizados en estos estudios se encuentran los árboles de decisión y simple K' means, que en el proyecto de (Dávila Hernández et al., n.d.) hace una buena combinación de ambos para su obtención de resultados, clasificando las edades y padecimientos que acontecen a los datos de las personas analizadas. Por otra parte (Antonio Gutiérrez Pérez et al., n.d.) obtuvo el índice de mortalidad del cáncer por edades, así como el índice de supervivencia aplicando el algoritmo j48 correspondiente al algoritmo C.4.5.

3 Marco Teórico

3.1 Metodología CRISP-DM

El estudio de esta investigación se hizo con la metodología CRISP-DM, ya que permite la extracción de conocimiento a partir de los datos que pasan a través de etapas conformadas por esta metodología.

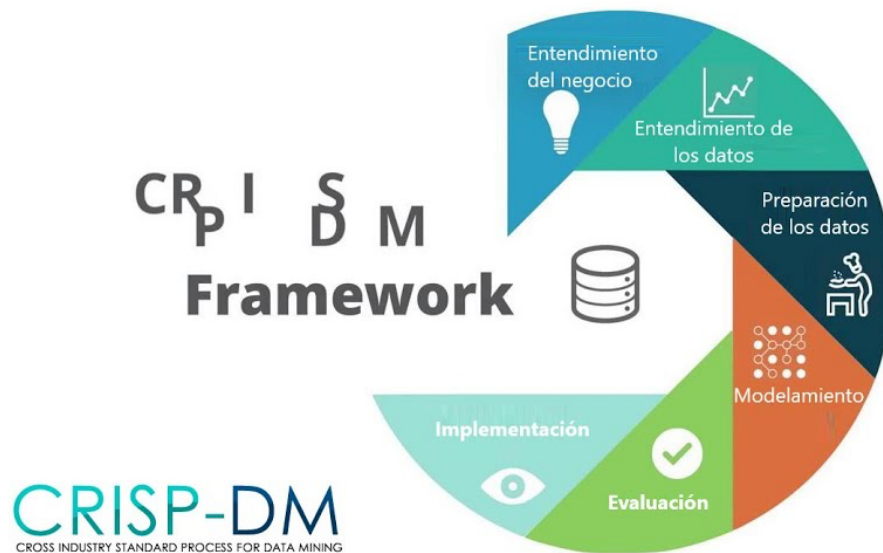


Figura 2. Etapas de CRISP-DM, imagen obtenida de (Takeshi, s.f.)

Como se muestra en la Figura 2, la metodología CRISP-DM está compuesta por 6 fases, a continuación, se define cada una de estas.

- **Entendimiento del negocio y comprensión de los datos:** Esta fase inicial se enfoca en la comprensión de los objetivos de proyecto. Después se convierte este conocimiento de los datos en la definición de un problema de minería de datos y en un plan preliminar diseñado para alcanzar los objetivos. La fase de entendimiento de datos comienza con la colección de datos inicial y continúa con las actividades que permiten familiarizarse con los datos, identificar los problemas de calidad, descubrir conocimiento preliminar sobre los datos, y/o descubrir subconjuntos interesantes para formar hipótesis en cuanto a la información oculta (Azevedo y Santos, 2008).
- **Preparación de los datos:** La fase de preparación de datos cubre todas las actividades necesarias para construir el conjunto final de datos (los datos que

se utilizarán en las herramientas de modelado) a partir de los datos en bruto iniciales. Las tareas incluyen la selección de tablas, registros y atributos, así como la transformación y la limpieza de datos para las herramientas que modelan (Azevedo y Santos, 2008).

- **Modelado:** En esta fase, se seleccionan y aplican las técnicas de modelado que sean pertinentes al problema (cuantas más mejor), y se calibran sus parámetros a valores óptimos. Típicamente hay varias técnicas para el mismo tipo de problema de minería de datos. Algunas técnicas tienen requerimientos específicos sobre la forma de los datos. Por lo tanto, casi siempre en cualquier proyecto se acaba volviendo a la fase de preparación de datos (Azevedo y Santos, 2008).
- **Evaluación:** En esta etapa en el proyecto, se han construido uno o varios modelos que parecen alcanzar calidad suficiente desde la una perspectiva de análisis de datos. Antes de proceder al despliegue final del modelo, es importante evaluarlo a fondo y revisar los pasos ejecutados para crearlo, comparar el modelo obtenido con los objetivos de negocio. Un objetivo clave es determinar si hay alguna cuestión importante de negocio que no haya sido considerada suficientemente. Al final de esta fase, se debería obtener una decisión sobre la aplicación de los resultados del proceso de análisis de datos (Azevedo y Santos, 2008).
- **Despliegue:** Generalmente, la creación del modelo no es el final del proyecto. Incluso si el objetivo del modelo es de aumentar el conocimiento de los datos, el conocimiento obtenido tendrá que organizarse y presentarse para que el cliente pueda usarlo. Dependiendo de los requisitos, la fase de desarrollo puede ser tan simple como la generación de un informe o tan compleja como la realización periódica y quizás automatizada de un proceso de análisis de datos en la organización (Azevedo y Santos, 2008)

3.2 Técnicas de preparación de los datos

La preparación de datos es un paso vital en el proceso de análisis de datos, ya que garantiza la calidad y confiabilidad de los datos para el modelado y la toma

de decisiones especialmente en este tipo de estudios donde se involucra el tratamiento de los pacientes y se toman decisiones sumamente importantes.

Las técnicas de preprocesamiento que se describen a continuación ayudarán a crear un buen conjunto de datos. Una gran parte del preprocesamiento de datos es el aspecto de la transformación. Cuando se trata de datos sin procesar nunca se sabe lo que se va a obtener. Normalizar datos mediante el proceso de transformación es una de las maneras más rápidas y eficientes para lograr el objetivo de tener datos limpios y utilizables. Para ello se emplearon distintos procesos que se describen a continuación:

- Limpieza de datos: la limpieza de datos elimina ruido y resuelve las inconsistencias en los datos, que pueden perjudicar los resultados del análisis.
- Integración: con la Integración de datos se migran datos de varias fuentes a una fuente coherente como un Data Warehouse, de esta manera será más fácil procesarlos desde un mismo almacén de datos.
- Transformación: la transformación de datos sirve para normalizar datos de cualquier tipo, de esta forma se pueden procesar de una manera más sencilla, manteniendo la información más importante y en el formato en que se requiere.
- Reducción: la reducción de datos reduce el tamaño de los datos, manteniendo solo los datos importantes a procesar y transformando o incluso, si se considera necesario, desechar los datos que no serán útiles.

3.3 Algoritmos de minería de datos

Un algoritmo en minería de datos es un conjunto de heurísticas y cálculos que permiten crear un modelo a partir de datos. Para crear un modelo, el algoritmo analiza primero los datos proporcionados, en busca de tipos específicos de patrones o tendencias.

El algoritmo usa los resultados de este análisis en un gran número de iteraciones para determinar los parámetros óptimos para crear el modelo de minería de datos. (Minewiskan, 2018).

Algoritmos de clasificación, los algoritmos enfocados a la clasificación trabajan sobre la información entregada por muestras, patrones, ejemplos o prototipos de entrenamiento tomados como representantes de las clases, y que conservan una etiqueta correcta.

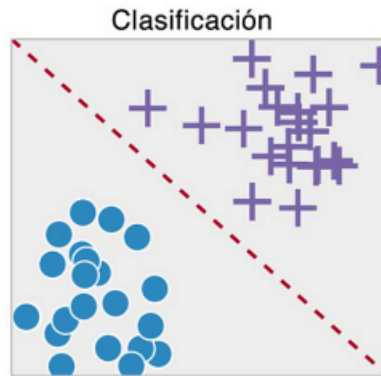


Figura 3. Algoritmo de clasificación Imagen obtenida de (Buitrago, 2020)

Algoritmos de asociación, que buscan correlaciones entre diferentes atributos de un conjunto de datos. La aplicación más común de esta clase de algoritmo es la creación de reglas de asociación, se hace énfasis en este tipo de algoritmos, ya que será de utilidad en este estudio particularmente, pues se busca relacionar las distintas variables que influyen en el tratamiento y recuperación de los pacientes.

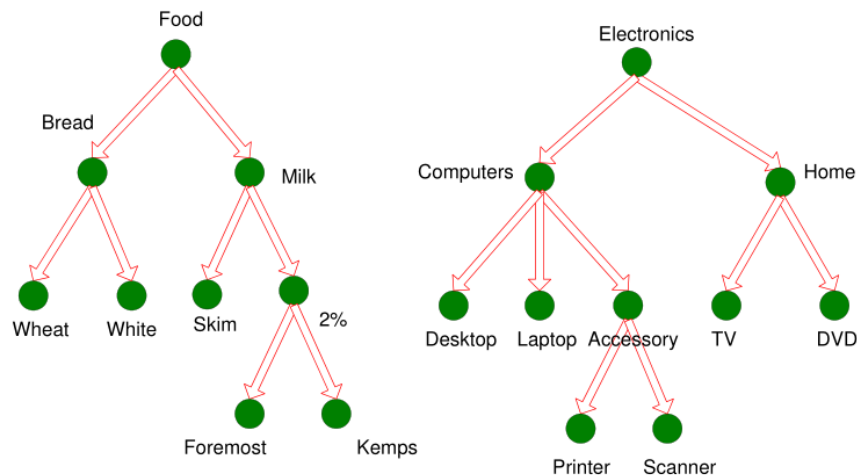


Figura 4. Algoritmo de asociación Imagen obtenida de (Alcalde, 2018)

Los algoritmos de agrupamiento, un algoritmo de agrupamiento es un procedimiento de agrupación de una serie de vectores de acuerdo con un criterio. Esos criterios son por lo general son distancia o similitud.

(Data Science Team, 2020)

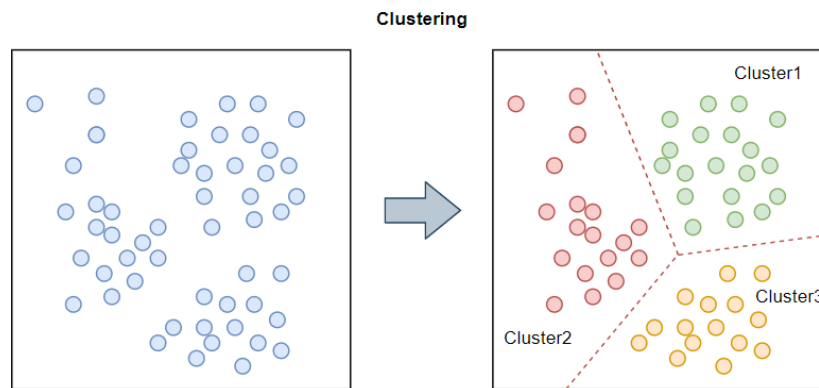


Figura 5. Algoritmo de agrupamiento Imagen obtenida de (Data Science Team, 2020)

3.3.1 Algoritmo de regresión lineal

Este algoritmo de asociación permite describir la relación entre dos o más variables, una vez definida la relación existente entre éstas, es posible “predecir” el comportamiento futuro.

Últimamente se ha utilizado en aprendizaje automático siendo un algoritmo supervisado, además es importante mencionar que el algoritmo de regresión lineal también ha demostrado su eficacia no sólo con fines de predicción, también al describir sistemas.

Se busca definir la relación entre dos variables con una línea recta, de ahí su nombre del algoritmo “Regresión Lineal”. Y el proceso básicamente es que una vez canalizados todos los puntos que hay en un plano, definir una línea que cruce por todos estos puntos y así describir la relación existente entre estas variables, aunque en la mayoría de los casos estos puntos están distribuidos de manera caótica y no es posible establecer una línea recta que pase por todos los puntos, el objetivo, es trazar la línea lo más cercana a todos los puntos que se representan; la distancia entre la línea y los puntos se conoce como función de pérdida.

La fórmula que define a la regresión lineal es la siguiente: $y = \beta_0 + \beta_1 x + \varepsilon$ donde “y” representa a la variable dependiente, “x” la variable independiente, β_0 y β_1 son dos constantes desconocidas que representan el punto de intersección y la pendiente respectivamente, finalmente ε es la función de pérdida.

A continuación, un ejemplo de cómo se puede observar la regresión lineal:

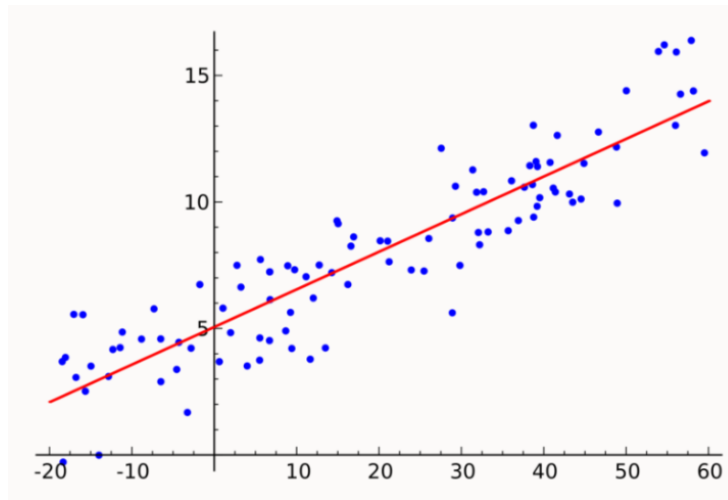


Figura 6. Ejemplo de regresión lineal

3.4 Complicaciones de la enfermedad vascular cerebral

Definiendo a lo que se conoce como enfermedad vascular cerebral (EVC) refiriéndose como a una alteración en las neuronas que provoca disminución de flujo sanguíneo en el cerebro acompañada de alteraciones cerebrales de manera momentánea o permanente. Es momento de hablar sobre las complicaciones que esta trae consigo, de acuerdo con (Mayo Clinic, 2021), las complicaciones pueden ser las siguientes:

- **Parálisis o pérdida del movimiento muscular.** Puedes quedar paralizado de un lado del cuerpo o perder el control de ciertos músculos, como los de un lado de la cara o de un brazo.
- **Dificultad para hablar o tragar.** Un accidente cerebrovascular puede afectar el control de los músculos de la boca y la garganta, dificultando que puedas hablar con claridad, tragar o comer. También puedes tener dificultad con el lenguaje, incluido hablar o entender el habla, leer o escribir.
- **Pérdida de memoria o dificultades para pensar.** Muchas personas que han tenido accidentes cerebrovasculares experimentan algo de

pérdida de memoria. Otras pueden tener dificultades para pensar, razonar, opinar y comprender conceptos.

- **Problemas emocionales.** Las personas que han tenido accidentes cerebrovasculares pueden tener más dificultad para controlar sus emociones o pueden sufrir depresión.
- **Dolor.** Se puede presentar dolor, entumecimiento u otras sensaciones inusuales en las partes del cuerpo afectadas por el accidente cerebrovascular. Por ejemplo, si un accidente cerebrovascular hace que pierdas la sensibilidad en el brazo izquierdo, puedes desarrollar una sensación de hormigueo incómoda en ese brazo.
- **Cambios en la conducta y en la capacidad de cuidado personal.** Las personas que han tenido accidentes cerebrovasculares pueden volverse más retraídas. Es posible que necesiten ayuda con el aseo personal y las tareas diarias.

3.5 Rehabilitación de la enfermedad vascular cerebral

Existen muchos enfoques para la rehabilitación del accidente cerebrovascular. El plan de rehabilitación dependerá de la parte del cuerpo o del tipo de capacidad afectada por el accidente cerebrovascular.

(Mayo Clinic, 2021) Indica que las actividades de rehabilitación físicas pueden incluir:

- **Ejercicios de motricidad.** Estos ejercicios pueden ayudar a mejorar la fuerza muscular y la coordinación.
- **Entrenamiento para la movilidad.** Se aprende a usar ayudas para la movilidad, como un andador, bastones, una silla de ruedas o una tobillera.
- **Terapia inducida por restricción.** Se restringe una extremidad no afectada mientras se practica el movimiento con la extremidad afectada para ayudar a mejorar su función.

- **Terapia de amplitud de movimiento.** Ciertos ejercicios y tratamientos pueden aliviar la tensión muscular (espasticidad) y ayudar a recuperar la amplitud de movimiento

Además, existen las terapias físicas asistidas por la tecnología, que podrían incluir:

- **Estimulación eléctrica funcional.** Se aplica electricidad a los músculos debilitados, y esto hace que se contraigan. La estimulación eléctrica puede ayudar a reeducar los músculos.
- **Tecnología robótica.** Los dispositivos robóticos pueden ayudar a que las extremidades dañadas realicen movimientos repetitivos, para que recuperen su fuerza y función.
- **Realidad virtual.** El uso de videojuegos y otras terapias basadas en computadoras implica interactuar con un entorno simulado y en tiempo real.

Las terapias cognitivas y emocionales pueden incluir:

- **Terapia para trastornos cognitivos.** La terapia ocupacional y la terapia del habla pueden ayudar con las capacidades cognitivas perdidas, como la memoria, el procesamiento, la resolución de problemas, las habilidades sociales, el juicio y la conciencia de seguridad.
- **Terapia para los trastornos de comunicación.** La terapia del habla puede ayudar a recuperar las habilidades perdidas en el habla, la audición, la escritura y la comprensión.
- **Evaluación y tratamiento psicológico.** La adaptación emocional podría ser puesta a prueba.
- **Medicamentos.** El médico puede recomendar un antidepresivo o un medicamento que afecte la lucidez mental, la agitación o el movimiento.

Las terapias experimentales incluyen:

- Estimulación cerebral no invasiva. Técnicas como la estimulación magnética transcraneal se han utilizado con cierto éxito en un entorno de investigación para ayudar a mejorar una variedad de habilidades motoras.
- Se están investigando las terapias biológicas, como las células madre, pero solo se deben utilizar como parte de un ensayo clínico.
- Medicina alternativa. Se están evaluando tratamientos como el masaje, la terapia herbal, la acupuntura y la oxigenoterapia.
- Terapias haciendo uso de juegos serios. El uso de programas de entrenamiento y rehabilitación cognitiva asistido por computadora facilita abordar dificultades del funcionamiento cognitivo con precisión y consistencia. Este tipo de asistencia ha estado basado en los últimos años en juegos serios de tecnologías web, Realidad Virtual y Realidad Aumentada.

3.6 Juegos serios

Los juegos serios responden a la necesidad de ofrecer herramientas para realizar actividades formativas mediante medios lúdicos tecnológicos y entretenidos. Estas herramientas han demostrado que el desarrollo de habilidades cognitivas y físicas de sus usuarios se incrementan debido a las novedosas y constantes estrategias para la motivación del aprendizaje no pasivo a la que estos juegos los exponen. (Cano, Peñeñory, Collazos, fardoun & Alghazzawi, 2015).

En la figura 3 se muestra un claro ejemplo de un juego serio y cómo ayuda a los pacientes a mejorar la rehabilitación.



Figura 7. Juegos serios, imagen obtenida de (MedicalExpo, s.f.)

3.7 Motivación

De acuerdo con (Santander, 2021), la motivación se define como el “conjunto de factores internos o externos que determinan en parte las acciones de una persona”. En otras palabras, se dice que cuando alguien está motivado, sus acciones y energías están dirigidas a alcanzar una meta concreta.

Actualmente, la motivación suele identificarse como la clave del éxito precisamente porque es la impulsora del resto de habilidades que conducen a un objetivo y que son la voluntad, la perseverancia, el optimismo y la actitud positiva, entre muchas otras. Es necesario señalar que la motivación puede venir dada por factores internos o externos. Por este motivo, se diferencia entre la motivación intrínseca y la motivación extrínseca.

- La motivación intrínseca es aquella que nace del interior de cada uno y es independiente a cualquier tipo de estímulo externo. Las recompensas que se tienen alimentando este tipo de motivación son puramente íntimas: amor propio, satisfacción personal, independencia, fuerza interior, confianza, etc.
- La motivación extrínseca, en cambio, es el conjunto de acciones que se llevan a cabo para conseguir una recompensa externa. El ejemplo más común es el económico, la motivación extrínseca busca la validación personal por parte de un tercero: los elogios o los refuerzos psicológicos son aspectos clave de la motivación extrínseca.

Una de las hipótesis más célebres sobre la motivación intrínseca es la teoría de la autodeterminación (Edward Deci y Richard Ryan, 1975). En esta teoría, psicólogos y profesores de la Universidad de Rochester sostienen que la motivación intrínseca aparece en todo ser humano cuando se reúnen diversas necesidades:

- Necesidad de ser autónomo
- Necesidad de relacionarse
- Necesidad de ser competente.

En otras palabras, cuando el propio comportamiento y los cambios de conducta están alimentados por un deseo personal de mejorar y de alcanzar estos sentimientos de satisfacción, sin la intención de recibir una recompensa externa, se puede afirmar que la motivación es intrínseca, y este tipo de motivación suele resultar en la segregación de serotonina. A su vez, tener altos niveles de serotonina en el organismo comporta una sensación de felicidad, serenidad y optimismo, además de aliviar la ansiedad y mejorar los procesos cognitivos.

En resumen, la motivación intrínseca y un buen estado de ánimo van de la mano, multiplicando así las probabilidades de que se alcancen las metas que se tengan en mente.

El empleo de tecnologías innovadoras juega un papel importante en la motivación extrínseca, pues los pacientes consideran que el empleo de nuevas tecnologías y métodos de rehabilitación para su pronta recuperación es sumamente gratificante, lo que aumenta sus niveles de motivación y, por lo tanto, de mejora en el proceso de rehabilitación y seguimiento en sus terapias.

4 Análisis de datos de rehabilitación por medio de minería de datos

4.1 Introducción

En este capítulo se describe la metodología utilizada en el desarrollo de la investigación, así como las técnicas e instrumentos aplicados. Empezando con el enfoque de la investigación la cual es cuantitativa, pues se busca definir los avances que han tenido los pacientes que toman terapias con Gesture Therapy, por otra parte, también se busca perseguir el grado de relación existente entre las distintas variables que influyen en los resultados obtenidos, por lo que también es una investigación correlacional.

Basándose en las metodologías empleadas en estudios anteriores similares al que se presenta en este caso, la investigación se ha desarrollado empleando la metodología de CRISP-DM, que se define a continuación con cada una de sus fases adaptadas a este trabajo de investigación.

El motivo de la elección de esta metodología se centra en gran medida en la primera fase, la cual comprende el entendimiento de los datos y el dominio del estudio, esto considerando que es un trabajo del área de informática, pero inherentemente a la medicina, pues para la realización de esta investigación es necesario realizar un estudio profundo de temas relacionados con la enfermedad vascular cerebral.

4.2 Comprensión del dominio del estudio y establecimiento de objetivos

Para este apartado de la metodología se ha hecho una investigación sobre las rehabilitaciones de pacientes con EVC, los tipos de terapias existentes para las secuelas que deja este fenómeno en las personas, todo esto ha sido necesario porque para realizar un estudio primero debe haber un entendimiento del dominio y así llegar a definir los objetivos de la investigación. Dentro de esta fase se hicieron numerosas entrevistas hacia los médicos especialistas, ayudando en gran medida la comprensión del proceso de rehabilitación de los pacientes, así como la interpretación de los datos obtenidos por Gesture Therapy, de la misma forma la

interpretación de las entrevistas de motivación intrínseca, datos sumamente importantes para el análisis.

4.2.1 Herramientas utilizadas

RapidMiner.¹ Versión 9.1. Es un software de licencia gratuita para el análisis y minería de datos. Permite el desarrollo de procesos de análisis de datos mediante el almacenamiento de más de 500 operadores de entrada y salida, preprocesamiento de datos y visualización. Alguna de sus características principales son las siguientes:

- Desarrollado en Java.
- Multiplataforma.
- Puede usarse de diversas maneras:
- A través de un GUI.
- En línea de comandos.
- En batch (lotes).
- Desde otros programas a través de llamadas a sus bibliotecas.
- Incluye gráficos y herramientas de visualización de datos.
- Dispone de módulos de integración con R y Python.

PowerBI.² Versión 2.124.1052.0 Es un servicio de análisis de datos de Microsoft orientado a proporcionar visualizaciones interactivas y capacidades de inteligencia empresarial con una interfaz lo suficientemente simple como para que los usuarios finales puedan crear por sí mismos sus propios informes y paneles.

Weka.³ Versión 3.8.6 Es una plataforma de software destinada al aprendizaje automático y la minería de datos. Algunas de las aplicaciones de WEKA son las siguientes:

Preprocesamiento de datos: Multitud de herramientas para el preprocesamiento de los datos (como por ejemplo discretización de variables).

· Clasificación: Algoritmos de clasificación, distribuidos por paquetes, como por ejemplo ID3 o C4.5

· Racimos: Diferentes algoritmos de segmentación como el simple k-means.

¹ Rapidminer - <https://rapidminer.com/>

² PowerBI - <https://powerbi.microsoft.com/es-mx/>

³ Weka - <https://www.cs.waikato.ac.nz/ml/weka/>

- Asociación: Algoritmos para encontrar relaciones de asociación entre variables (Apriori entre otros).
- Selección de atributos: Weka es capaz de buscar por nosotros las mejores variables del modelo

4.3 Recolección, limpieza y preprocesamiento de los datos

En la etapa del preprocesamiento se analiza la calidad de los datos de acuerdo con los objetivos del tratamiento, se aplican estrategias para el manejo de datos ruidosos, como pueden ser datos nulos, duplicados o incoherentes y técnicas estadísticas para su reemplazo.

Para el preprocesamiento se buscan características útiles para representar los datos dependiendo de la meta del proceso.

4.3.1 Descripción del conjunto de datos

Los datos con los que se cuentan se organizan en dos tipos, por un lado, se registran los movimientos del paciente, a través de la manija de la figura 4. Por otro lado, los datos recabados por las encuestas que se realizan al finalizar cada sesión de terapia. A continuación, se describen estos conjuntos de datos.

Datos registrados de movimiento

Los datos que contiene el archivo generado por el dispositivo de la figura 4 se describen a continuación, en la tabla 1:

Dato	Descripción	Tipo	Rango
Fecha	Fecha de inicio de sesión de terapia	Fecha	05.11.2018 – 13.05.2019
Hora	Hora de inicio de la sesión	Hora	14:03:34.892-18:24:38.289
Ubicación	Ubicación de almacenamiento del archivo	String	N/A
Frame	Unidad de frame	Numérico	0 – 6468
Timestamp	Marca de tiempo	Hora	14:03:34.892-18:24:38.289
Real_x	Se describe el movimiento del eje “x”	Valor real	0.000000 – 0.999999
Real_y	Se describe el movimiento del eje “y”	Valor real	0.000000 – 0.999999
Real_z	Se describe el movimiento del eje “z”	Valor real	0.000000 – 0.999999
Signal_1	Presión ejercida sobre la manija con el pulgar	Numérico	0.000000000 – 0.999999999

Tabla 2. Diccionario de datos

En la figura 5, se muestra un ejemplo de los datos en crudo que se han generado durante una sesión.

Cada renglón del archivo representa un movimiento del paciente, durante una sesión en promedio se registran 7,000 movimientos aproximadamente y el intervalo promedio entre cada registro de movimiento es de 164 milisegundos.

```

## [Header]
#
# Date: lunes, 5 de noviembre de 2018
# Start time: 16:24:45.410
# Video File: C:\Users\biossen-19966\Documents\INAOE\gesture\Capture\lunes, 5 de noviembre de 2018 - 04-24-45 p. m..avi
#
# Signals:
#   signal_1: Pressure/Finger1
#   signal_2: Finger2
#   signal_3: Finger3
#   signal_4: Finger4
#   signal_5: Finger5
#   signal_6: Finger6
#   signal_7: Finger7
#
## [Data]
Frame; Timestamp; Real_x; Real_y; Real_z; Real_w; Real_h; signal_1; signal_2; signal_3; signal_4; signal_5; signal_6; signal_7
0; 16:24:45.688; 0.4757926; 0.4705267; 0.03484094; 0.04645459; 0.03484094; 79.6; 0; 0; 0; 0; 0; 0
1; 16:24:45.852; 0.4745542; 0.4635573; 0.03641084; 0.04854779; 0.03641084; 79.7; 0; 0; 0; 0; 0; 0
2; 16:24:46.9; 0.4775727; 0.4429882; 0.03492033; 0.04656044; 0.03492033; 79.6; 0; 0; 0; 0; 0; 0
3; 16:24:46.157; 0.4618929; 0.4395712; 0.04223715; 0.0563162; 0.04223715; 79.6; 0; 0; 0; 0; 0; 0
4; 16:24:46.326; 0.4703383; 0.4329314; 0.03405781; 0.04541042; 0.03405781; 79.7; 0; 0; 0; 0; 0; 0
5; 16:24:46.480; 0.4792348; 0.4193146; 0.03729784; 0.04973045; 0.03729784; 79.6; 0; 0; 0; 0; 0; 0
6; 16:24:46.758; 0.4847324; 0.4129026; 0.03856986; 0.05142648; 0.03856986; 79.7; 0; 0; 0; 0; 0; 0
7; 16:24:47.31; 0.4706413; 0.4224602; 0.04698724; 0.06264965; 0.04698724; 79.6; 0; 0; 0; 0; 0; 0
8; 16:24:47.186; 0.4831648; 0.4618256; 0.03307243; 0.04409657; 0.03307243; 79.7; 0; 0; 0; 0; 0; 0
9; 16:24:47.465; 0.4854752; 0.482657; 0.0367981; 0.04906414; 0.0367981; 79.7; 0; 0; 0; 0; 0; 0
10; 16:24:47.646; 0.4965762; 0.4776543; 0.03775138; 0.05033517; 0.03775138; 79.6; 0; 0; 0; 0; 0; 0
11; 16:24:47.781; 0.5084479; 0.4873204; 0.02986509; 0.03982012; 0.02986509; 79.6; 0; 0; 0; 0; 0; 0
12; 16:24:47.932; 0.5097151; 0.4844434; 0.03721801; 0.04962402; 0.03721801; 79.7; 0; 0; 0; 0; 0; 0
13; 16:24:48.112; 0.5141001; 0.4793912; 0.03149401; 0.04199201; 0.03149401; 79.7; 0; 0; 0; 0; 0; 0
14; 16:24:48.257; 0.5159485; 0.4827429; 0.03964007; 0.05285343; 0.03964007; 79.7; 0; 0; 0; 0; 0; 0
15; 16:24:48.403; 0.5108488; 0.4690521; 0.04026508; 0.05368678; 0.04026508; 79.7; 0; 0; 0; 0; 0; 0
16; 16:24:48.554; 0.492775; 0.4654675; 0.0322611; 0.0430148; 0.0322611; 79.7; 0; 0; 0; 0; 0; 0
17; 16:24:48.698; 0.5008108; 0.4553503; 0.03085126; 0.04113501; 0.03085126; 79.7; 0; 0; 0; 0; 0; 0
18; 16:24:48.842; 0.4834252; 0.4529293; 0.03601755; 0.0480234; 0.03601755; 79.7; 0; 0; 0; 0; 0; 0
19; 16:24:48.993; 0.5004058; 0.4437962; 0.02952107; 0.03936143; 0.02952107; 79.7; 0; 0; 0; 0; 0; 0
20; 16:24:49.84; 0.4996238; 0.4355839; 0.03538185; 0.04717579; 0.03538185; 79.7; 0; 0; 0; 0; 0; 0
21; 16:24:49.228; 0.4808802; 0.4303954; 0.03123257; 0.04164343; 0.03123257; 79.7; 0; 0; 0; 0; 0; 0
22; 16:24:49.370; 0.4755751; 0.433394; 0.03936284; 0.05248378; 0.03936284; 79.7; 0; 0; 0; 0; 0; 0
23; 16:24:49.508; 0.4810091; 0.4456908; 0.0307355; 0.04098067; 0.0307355; 79.7; 0; 0; 0; 0; 0; 0
24; 16:24:49.642; 0.4762706; 0.4529794; 0.03220294; 0.04293726; 0.03220294; 79.7; 0; 0; 0; 0; 0; 0

```

Figura 1. Datos en crudo recolectados por Gesture Therapy

Hay una colección de 411 archivos de las sesiones de cada paciente, en conjunto son 418,364 datos aproximadamente, es decir, cada uno registra un movimiento del paciente. Cabe mencionar que se genera un archivo por cada juego hecho por el paciente, en una sesión hay varios juegos, por lo tanto, en cada sesión se generan varios archivos, en la tabla 2 se detalla esta información.

Pacientes	Sesiones	Movimientos	Archivos
6 pacientes	65 sesiones	418,364	411

Tabla 3. Resumen de datos

En total, se cuentan con datos de 6 pacientes que han tomado terapias usando Gesture Therapy. En las tablas 3 a la 8 se describen a detalle los datos que se almacenan.

No. Paciente	Sexo	Movimientos por sesión	Archivos por sesión
P06	Femenino	Sesión 1: 5,706 movimientos Sesión 2: 5,654 movimientos Sesión 3: 6,513 movimientos Sesión 4: 6,412 movimientos Sesión 5: 3,790 movimientos	Sesión 1: 6 archivos Sesión 2: 5 archivos Sesión 3: 6 archivos Sesión 4: 5 archivos Sesión 5: 6 archivos

Tabla 4. Paciente P06

No. Paciente	Sexo	Movimientos por sesión	Archivos por sesión
P07	Masculino	Sesión 1: 3,974 movimientos Sesión 2: 7,720 movimientos Sesión 3: 6,336 movimientos Sesión 4: 7,361 movimientos Sesión 5: 9,158 movimientos Sesión 6: 7,889 movimientos Sesión 7: 7,004 movimientos Sesión 8: 8,246 movimientos Sesión 9: 8,532 movimientos Sesión 10: 8,474 movimientos Sesión 11: 7,024 movimientos Sesión 12: 8,143 movimientos Sesión 13: 9,158 movimientos	Sesión 1: 5 archivos Sesión 2: 8 archivos Sesión 3: 5 archivos Sesión 4: 5 archivos Sesión 5: 6 archivos Sesión 6: 7 archivos Sesión 7: 5 archivos Sesión 8: 7 archivos Sesión 9: 5 archivos Sesión 10: 6 archivos Sesión 11: 5 archivos Sesión 12: 6 archivos Sesión 13: 5 archivos

Tabla 5. Paciente P07

No. Paciente	Sexo	Movimientos por sesión	Archivos por sesión
P08	Femenino	Sesión 1: 3,974 movimientos Sesión 2: 7,720 movimientos Sesión 3: 6,336 movimientos Sesión 4: 7,361 movimientos Sesión 5: 9,158 movimientos Sesión 6: 7,889 movimientos Sesión 7: 7,004 movimientos Sesión 8: 8,246 movimientos Sesión 9: 8,532 movimientos Sesión 10: 8,474 movimientos Sesión 11: 7,024 movimientos	Sesión 1: 5 archivos Sesión 2: 8 archivos Sesión 3: 5 archivos Sesión 4: 5 archivos Sesión 5: 6 archivos Sesión 6: 7 archivos Sesión 7: 5 archivos Sesión 8: 7 archivos Sesión 9: 5 archivos Sesión 10: 6 archivos Sesión 11: 5 archivos

Tabla 6. Paciente P08

No. Paciente	Sexo	Movimientos por sesión	Archivos por sesión
P09	Femenino	Sesión 1: 3,974 movimientos Sesión 2: 7,720 movimientos Sesión 3: 6,336 movimientos Sesión 4: 7,361 movimientos Sesión 5: 9,158 movimientos Sesión 6: 7,889 movimientos Sesión 7: 7,004 movimientos Sesión 8: 8,246 movimientos Sesión 9: 8,532 movimientos Sesión 10: 8,474 movimientos Sesión 11: 7,024 movimientos Sesión 12: 9,158 movimientos	Sesión 1: 5 archivos Sesión 2: 8 archivos Sesión 3: 5 archivos Sesión 4: 5 archivos Sesión 5: 6 archivos Sesión 6: 7 archivos Sesión 7: 5 archivos Sesión 8: 7 archivos Sesión 9: 5 archivos Sesión 10: 6 archivos Sesión 11: 5 archivos Sesión 12: 6 archivos

Tabla 7. Paciente P09

No. Paciente	Sexo	Movimientos por sesión	Archivos por sesión
P10	Masculino	Sesión 1: 3,974 movimientos Sesión 2: 7,720 movimientos Sesión 3: 6,336 movimientos Sesión 4: 7,361 movimientos Sesión 5: 9,158 movimientos Sesión 6: 7,889 movimientos Sesión 7: 7,004 movimientos Sesión 8: 8,246 movimientos Sesión 9: 8,532 movimientos Sesión 10: 8,474 movimientos Sesión 11: 7,024 movimientos	Sesión 1: 5 archivos Sesión 2: 8 archivos Sesión 3: 5 archivos Sesión 4: 5 archivos Sesión 5: 6 archivos Sesión 6: 7 archivos Sesión 7: 5 archivos Sesión 8: 7 archivos Sesión 9: 5 archivos Sesión 10: 6 archivos Sesión 11: 5 archivos

Tabla 8. Paciente P10

No. Paciente	Sexo	Movimientos por sesión	Archivos por sesión
P11	Masculino	Sesión 1: 3,974 movimientos Sesión 2: 7,720 movimientos Sesión 3: 6,336 movimientos Sesión 4: 7,361 movimientos Sesión 5: 9,158 movimientos Sesión 6: 7,889 movimientos Sesión 7: 7,004 movimientos Sesión 8: 8,246 movimientos Sesión 9: 8,532 movimientos Sesión 10: 8,474 movimientos Sesión 11: 7,024 movimientos	Sesión 1: 5 archivos Sesión 2: 8 archivos Sesión 3: 5 archivos Sesión 4: 5 archivos Sesión 5: 6 archivos Sesión 6: 7 archivos Sesión 7: 5 archivos Sesión 8: 7 archivos Sesión 9: 5 archivos Sesión 10: 6 archivos Sesión 11: 5 archivos

Tabla 9. Paciente P11

Datos obtenidos de encuestas de motivación intrínseca

Otra parte de los datos a procesar son los generados en las encuestas de motivación intrínseca, estas se realizan al finalizar cada sesión de terapia.

Estas encuestas tienen como encabezado el nombre del paciente, y la fecha de la sesión terapéutica. Posteriormente podremos observar una escala del 1 al 7, donde 1 significa “No es cierto”, 4 “más o menos cierto” y 7 “totalmente cierto”.

17 preguntas hechas al paciente que tendrá que contestar según la escala mencionada, tras el encabezado. Además de las 17 preguntas, se hacen cálculos para obtener ciertas características que experimentó el paciente durante su sesión de terapia que se describen en la tabla 9. La (R) frente a algunas preguntas significa que se calculará inversa, es decir, la escala 1 será la 7 y la 7 será la 1 para calcular, porque estas preguntas son negativas.

Característica por calcular	Respuestas clave para el cálculo
Interés / disfrute	1, 15 ®, 8
Precepción de competencia	2 ®, 7, 14
Esfuerzo / importancia	3, 10, 13
Presión / tensión	4, 9, 16
Valor / utilidad	5, 12, 17
Dolor	6, 11 ®

Tabla 10. Cálculo de variables

En la figura 6 se muestra un ejemplo de las encuestas de motivación intrínseca con los datos antes mencionados.

ESCALA DE MOTIVACIÓN INTRÍNSECA

Paciente: P06 (1)
 Día de la sesión: 05-11-2018

OTÓRGUE, EN BASE A SU PERCEPCIÓN, UNA CALIFICACIÓN QUE RESPONDA LAS SIGUIENTES PREGUNTAS:

1	2	3	4	5	6	7
NO ES CIERTO			MÁS O MENOS CIERTO	TOTALMENTE CIERTO		

CUESTIONARIO

1. El entrenamiento durante la terapia guiada con el computador, fue una agradable experiencia 7
2. Yo no fui capaz de realizar una adecuada ejecución de los ejercicios 6
3. Yo estaba plenamente involucrado con el tratamiento 5
4. Me sentí inquieto o incómodo durante la ejecución de los ejercicios 0
5. Creo que la terapia guiada con computador es muy importante para obtener mayor movimiento en mi brazo 5
6. Durante el entrenamiento sentí dolor importante en mi brazo 0
7. Estoy satisfecho de mi rendimiento durante las sesiones de entrenamiento 4
8. Creo que la terapia guiada con el computador es muy interesante 7
9. Sentí ansiedad durante las sesiones de entrenamiento 4
10. Para mí era muy importante realizar de forma adecuada los ejercicios 7
11. No experimenté dolor en mi brazo durante las sesiones de tratamiento 0
12. Considero este tipo de entrenamiento motor muy importante en la rehabilitación de los pacientes 7
13. He puesto mucho esfuerzo durante las sesiones de tratamiento 7
14. Me considero muy hábil para llevar a cabo los ejercicios del entrenamiento 6
15. Sentí presión durante la ejecución del entrenamiento 0
16. Considero que los ejercicios son muy aburridos 3
17. Creo que obtuve un beneficio con la terapia guiada con el computador 7

Respuesta clave:
 Interés / Disfrute: 1, 15 (R), 8.
 Percepción de Competencia: 2 (R), 7, 14.
 Esfuerzo / Importancia: 3, 10, 13.
 Presión / Tensión: 4, 9, 16.
 Valor / Utilidad: 5, 12, 17.
 Dolor: 6, 11 (R).

Figura 8. Escala de motivación intrínseca

En la tabla 10 se muestra un resumen de los datos obtenidos por las encuestas.

Encuestas	Preguntas
64	1088

Tabla 11. Resumen encuestas totales

En la tabla 11 se detalla el número de encuestas por paciente, además del número de preguntas a procesar.

Paciente	Número de encuestas	Número de preguntas a procesar
P06	5	85
P07	13	221
P08	10	170
P09	12	204
P10	12	204
P11	12	204

Tabla 12. Encuestas por paciente

4.4 Análisis de los datos

En este apartado el objetivo es preparar los datos obtenidos para su procesamiento, en esta fase se ha optado la utilización de la herramienta de RapidMiner ya que ofrece un abanico muy variado para la preparación de los datos, algunas de las técnicas utilizadas fueron la integración de los datos de los pacientes en un mismo lote de datos, también se empleó la reducción, ya que se eliminaron atributos innecesarios y que sólo generarán ruido al momento de procesar los datos, por otra parte la transformación, pues se hizo cálculo de nuevas variables a partir de las ya existentes y finalmente se obtuvieron algunas gráficas estadísticas de los datos, además se hizo la exportación de los datos ya preparados para poder ser utilizados en otras herramientas tales como Weka y PowerBI y así aplicar las distintas herramientas y ventajas que ofrece cada una.

En la figura 8, se muestra un modelo de integración de los datos de las encuestas de motivación intrínseca en RapidMiner, además se pueden observar las funciones que ofrece la herramienta antes mencionada, como la selección de atributos, ordenación de los datos y su integración.

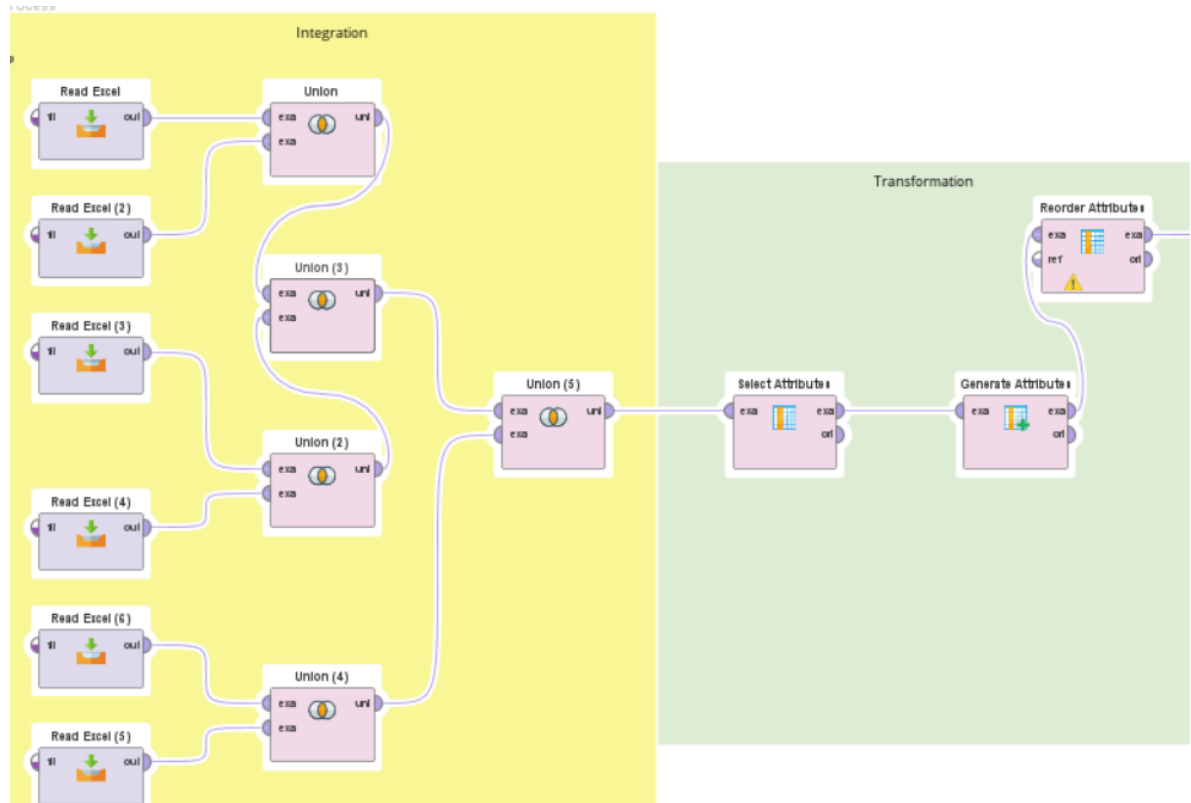


Figura 9. Preparación de los datos de encuestas en RapidMiner

Por otra parte, para los datos de los movimientos registrados por Gesture Therapy se ha hecho el cálculo de nuevas variables considerando el objetivo de estudio y los datos con los que se cuentan, estas nuevas variables se describen a continuación:

- Tiempo: Diferencia de tiempo en milisegundos entre cada frame.
- Distancia (x, y, z): Distancia recorrida de cada eje.
- Velocidad (x, y, z): Velocidad obtenida en cada eje por unidad de tiempo.
- Distancia total: Distancia total recorrida por cada frame.
- Velocidad total: Velocidad alcanzada en cada uno de los frame.
- Diferencia de presión entre cada frame

La forma de obtención de estas:

- Tiempo: Diferencia de tiempo en milisegundos entre cada registro.

$$t = f_1 - f$$

donde “f” es el tiempo en el primer y “f₁” el tiempo en el segundo frame.

- Distancia (x, y, z): Diferencia de la posición x, y, z entre cada frame.

$$d = x_1 - x$$

donde x es la posición consecutiva de " x_1 " y a su vez, " x_1 " es el valor de la posición inicial del aparato, aplicando esta ecuación para los tres ejes.

- Velocidad (x, y, z): Resultado de la división de la distancia (d) entre los milisegundos en cada frame (t).

$$v = \frac{d}{t}$$

- Distancia total: Se obtuvo la raíz cuadrada de la suma de la diferencia elevada al cuadrado de la distancia recorrida entre cada frame ($x_2 - x_1$), obtenida por cada uno de los ejes (x, y, z), es decir, se usó la distancia euclidiana.

$$Dt = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

- Velocidad total: Resultado de la división de la distancia total entre los milisegundos de cada frame.

$$Vt = \frac{Dt}{t}$$

- Diferencia de presión entre cada frame

$$P = p_2 - p_1.$$

Tras limpiar los datos de las encuestas, se pueden implementar modelos gráficos estadísticos para obtener información de interés, como se observa en las figuras 7-16.

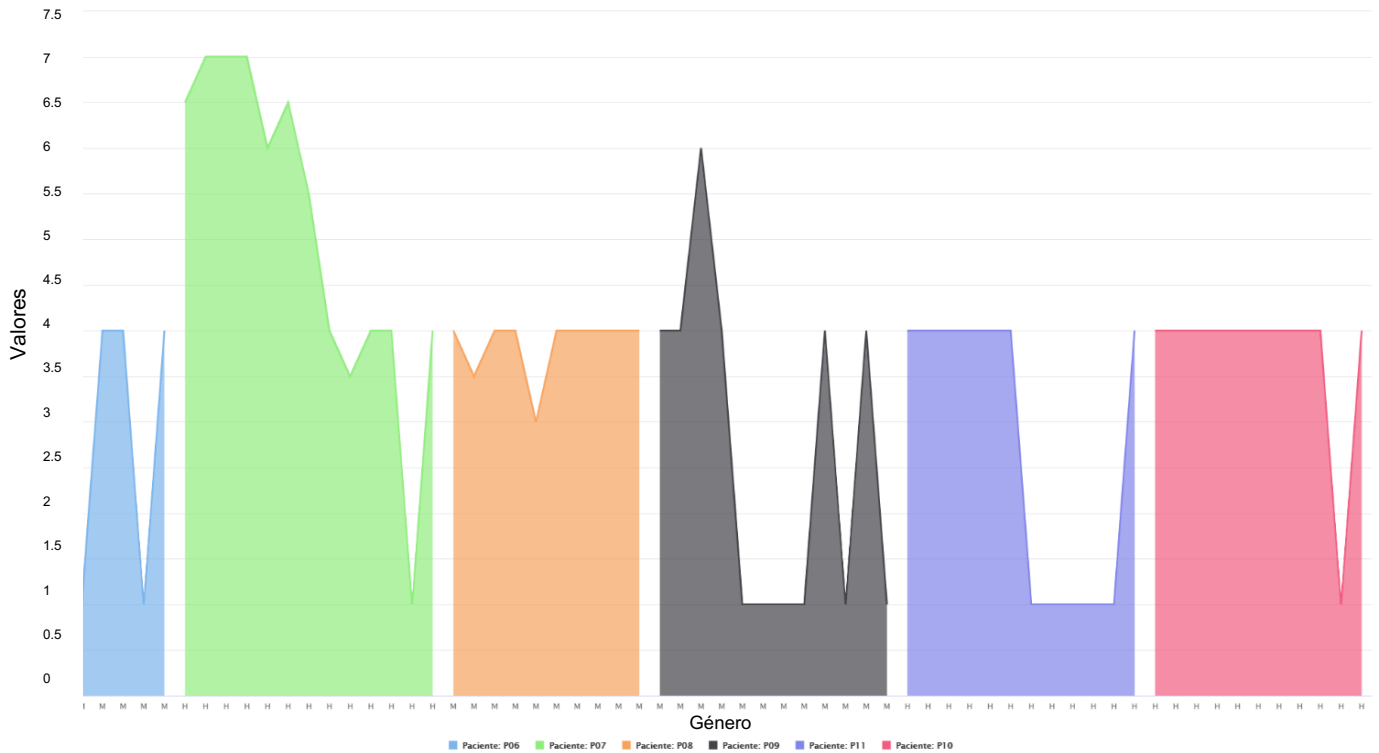


Figura 11. Gráfica de dolor por paciente

Existen más combinaciones interesantes que se han realizado en las gráficas, por ejemplo en la figura 12, se ha representado el esfuerzo/importancia frente al interés/disfrute y manteniendo la información por género, de la cual podemos inferir que las personas que más tienen interés y disfrutan de la rehabilitación le han puesto más esfuerzo e importancia a sus terapias y viceversa, pues la mayoría de los reactivos verdes, pertenecientes a los hombres se encuentran en los valores más altos de ambos ejes de la gráfica, concluyendo así que los hombres son los que más disfrutan la rehabilitación y por consecuente, le han puesto más importancia.

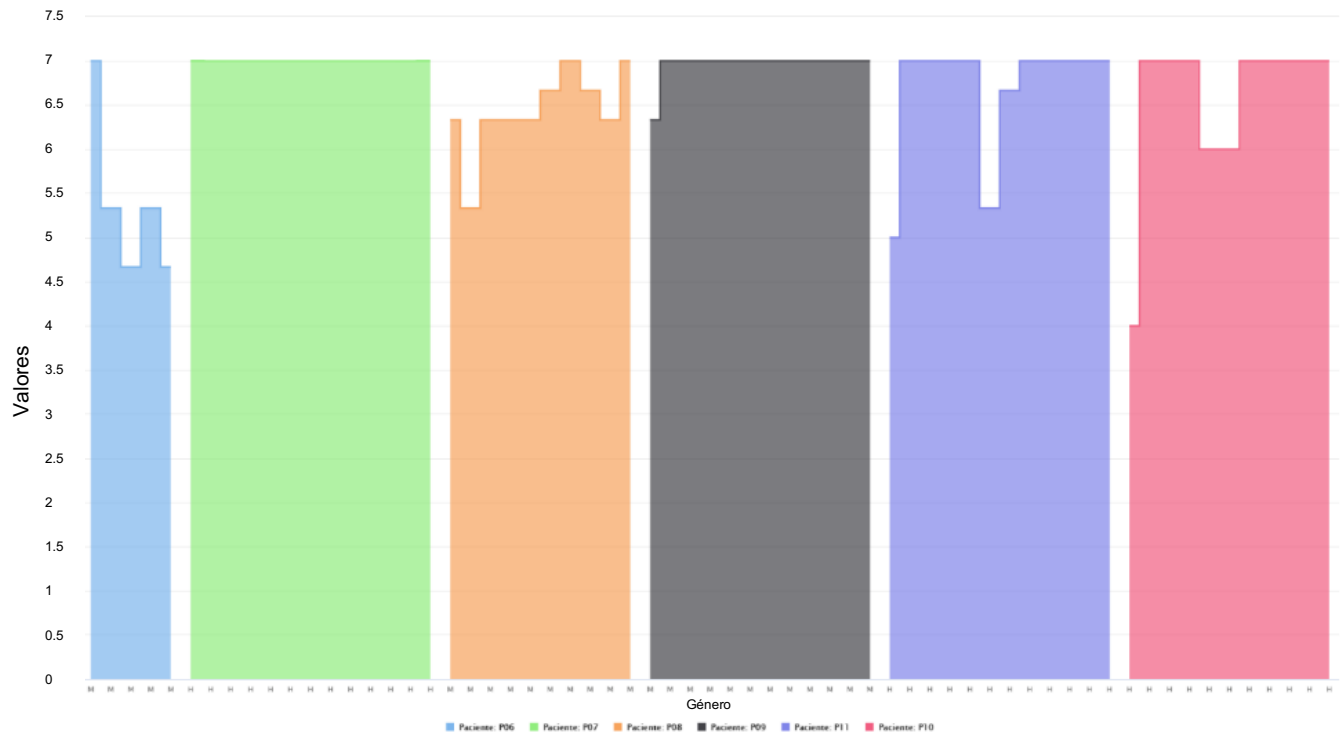


Figura 13. Gráfica esfuerzo/importancia por paciente

En la gráfica de la figura 14 se observa que el promedio más bajo del interés en la rehabilitación lo tiene el paciente 6, desde aquí, se empiezan a hacer notables las diferencias entre el paciente que menos tuvo sesiones con respecto a los demás,

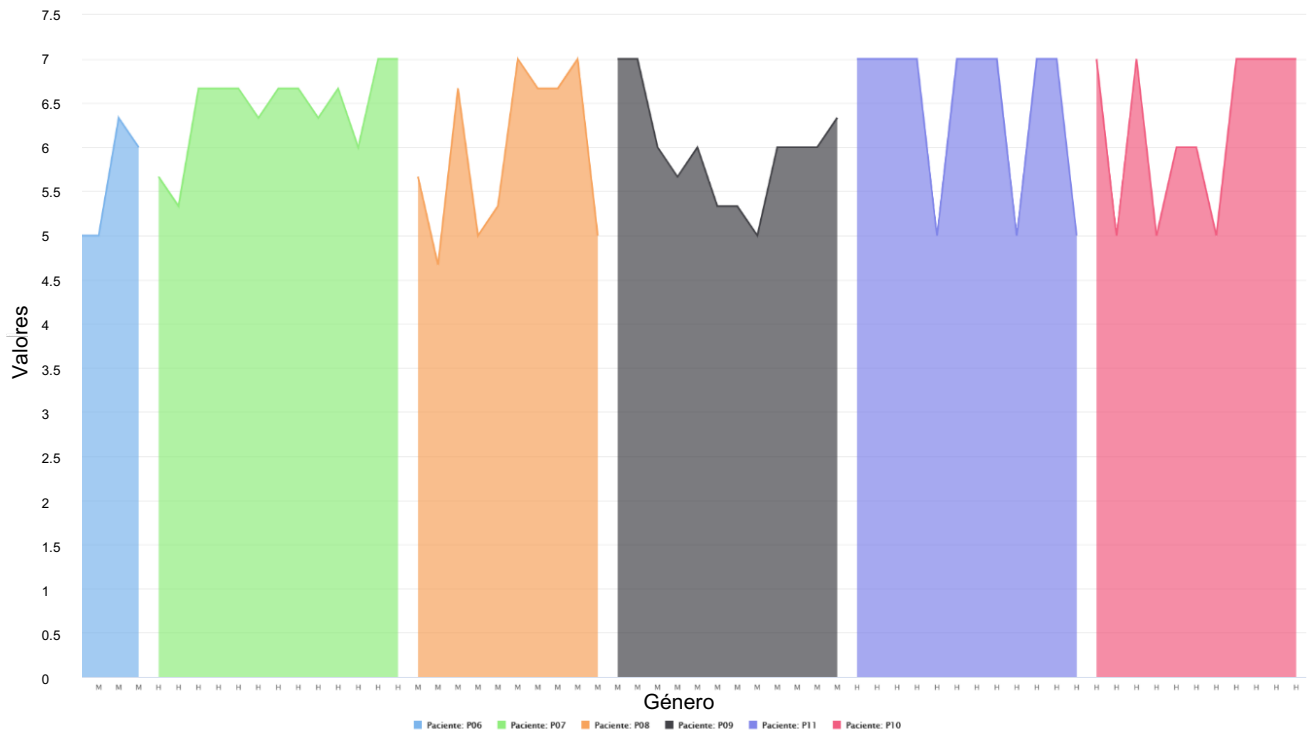


Figura 14. Gráfica de interés/disfrute por paciente

En la figura 15 se ha graficado el valor y la utilidad de cada paciente a su tratamiento, y nuevamente se nota una diferencia del paciente 6 con los demás, aunque también se ve que el paciente 10 empezó con un nivel bajo, pero que esto aumentó con el avance en sus terapias, así que es una posibilidad que el paciente 6 haya ganado más interés y valor a sus terapias si seguía con sus sesiones, pues como se observa, estos valores aumentaron en otros pacientes conforme al paso de sus terapias.

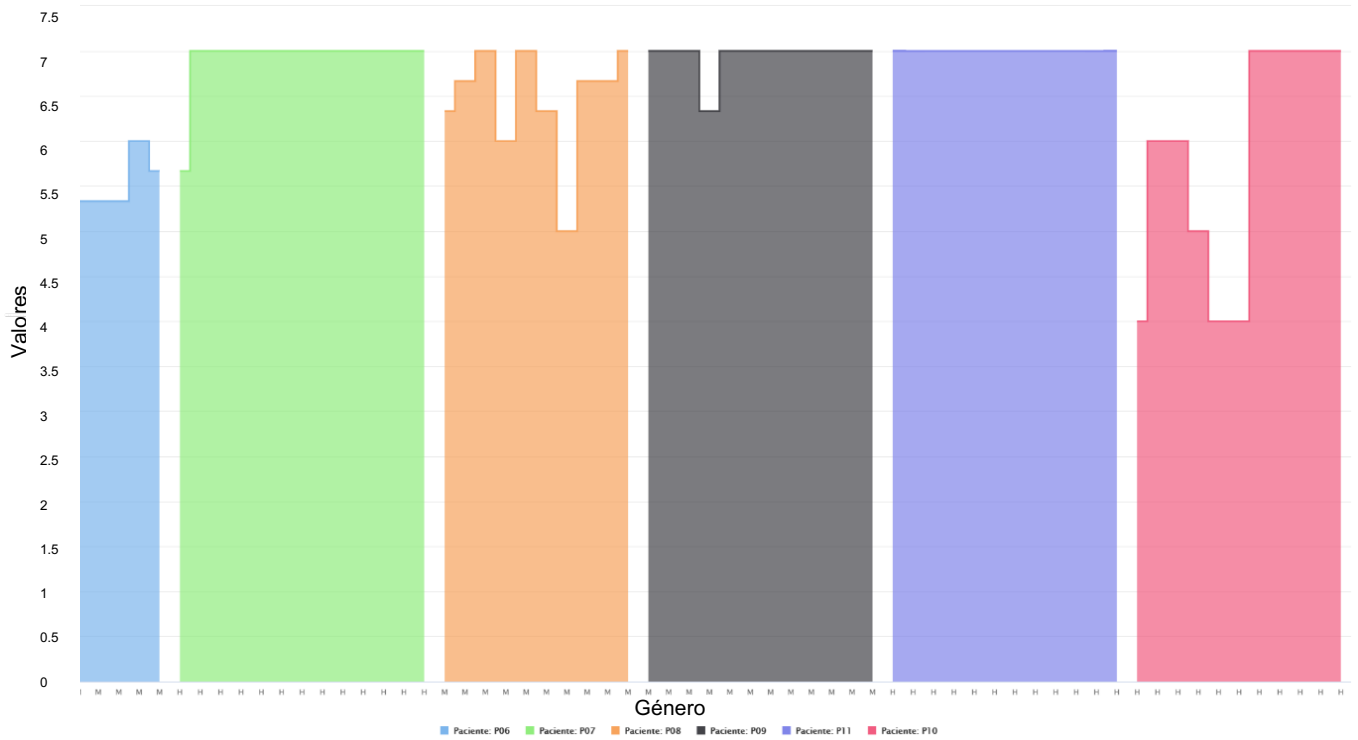


Figura 15. Gráfica de valor/utilidad por paciente

En la figura 16 se aprecia una tendencia a la baja de la tensión y presión de los pacientes con el paso del tratamiento, al menos para 3 de los 6, 2 de los cuales se mantienen estables y el paciente 6, el que sufrió más presión en su tratamiento, por lo que también influyó en la decisión de abandonarlo antes que los demás.

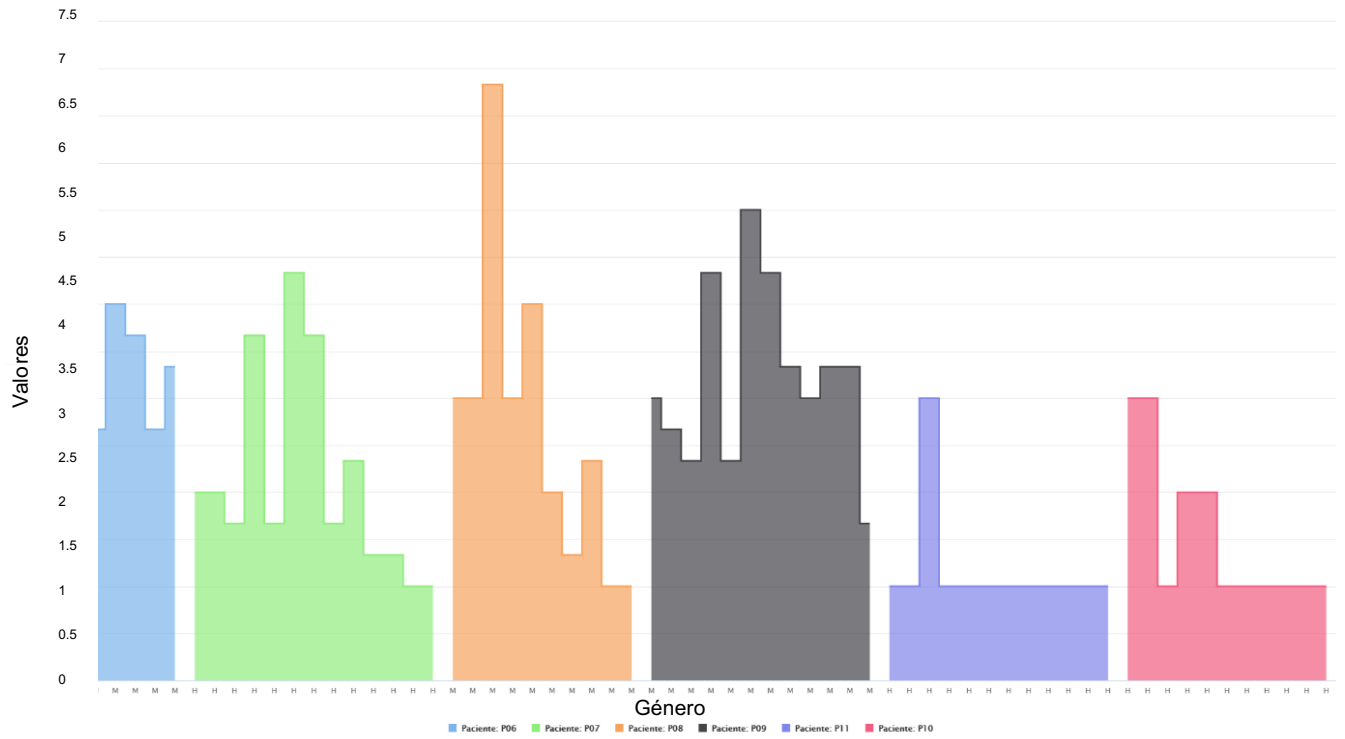


Figura 16. Gráfica de presión/tensión por paciente

Finalmente, las figuras 17 y 18 representan el contraste entre dos de las características de los pacientes, que son el dolor contra el interés y disfrute de los pacientes, se puede observar una clara proporcionalidad inversa entre las variables del dolor y del interés/disfrute y una proporcionalidad directa con el valor y la utilidad, lo que puede ayudar en la mejora de la toma de decisiones.

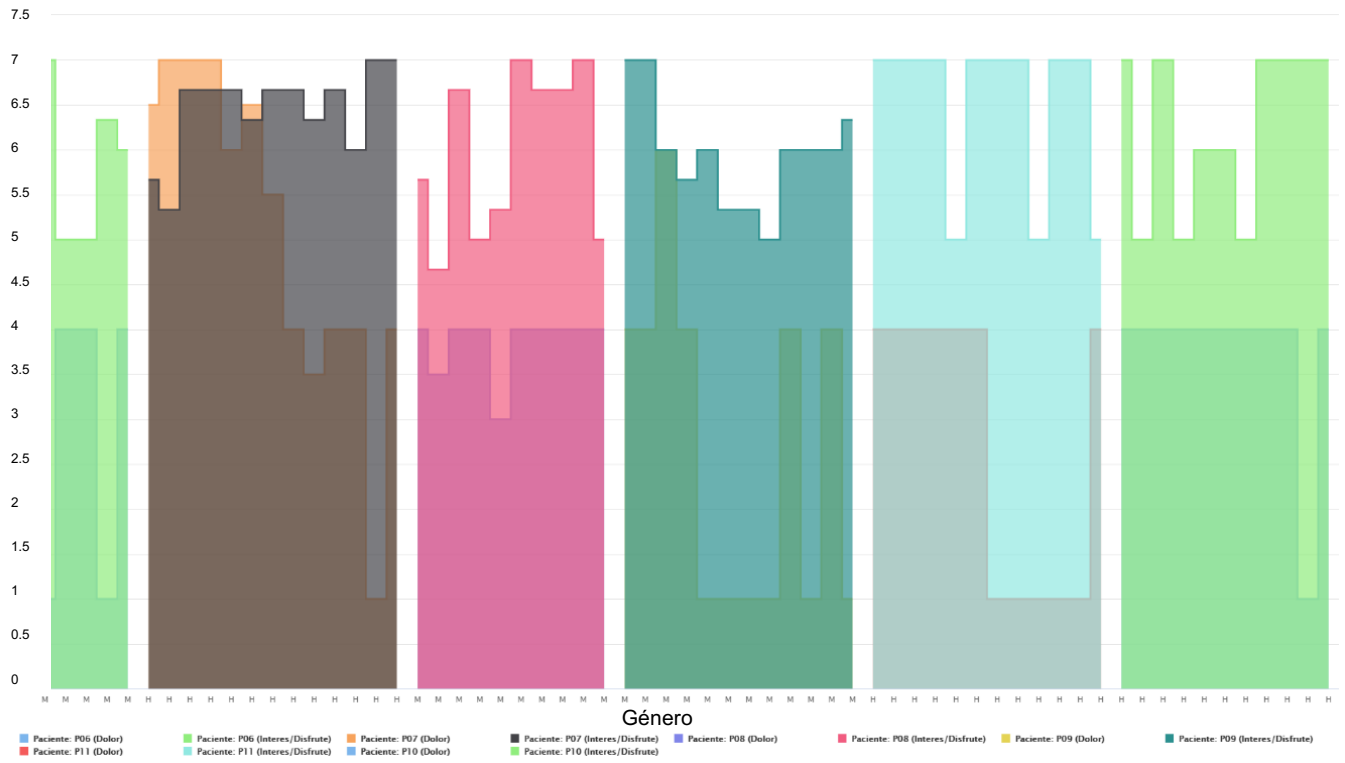


Figura 17. Dolor vs interés/disfrute por paciente

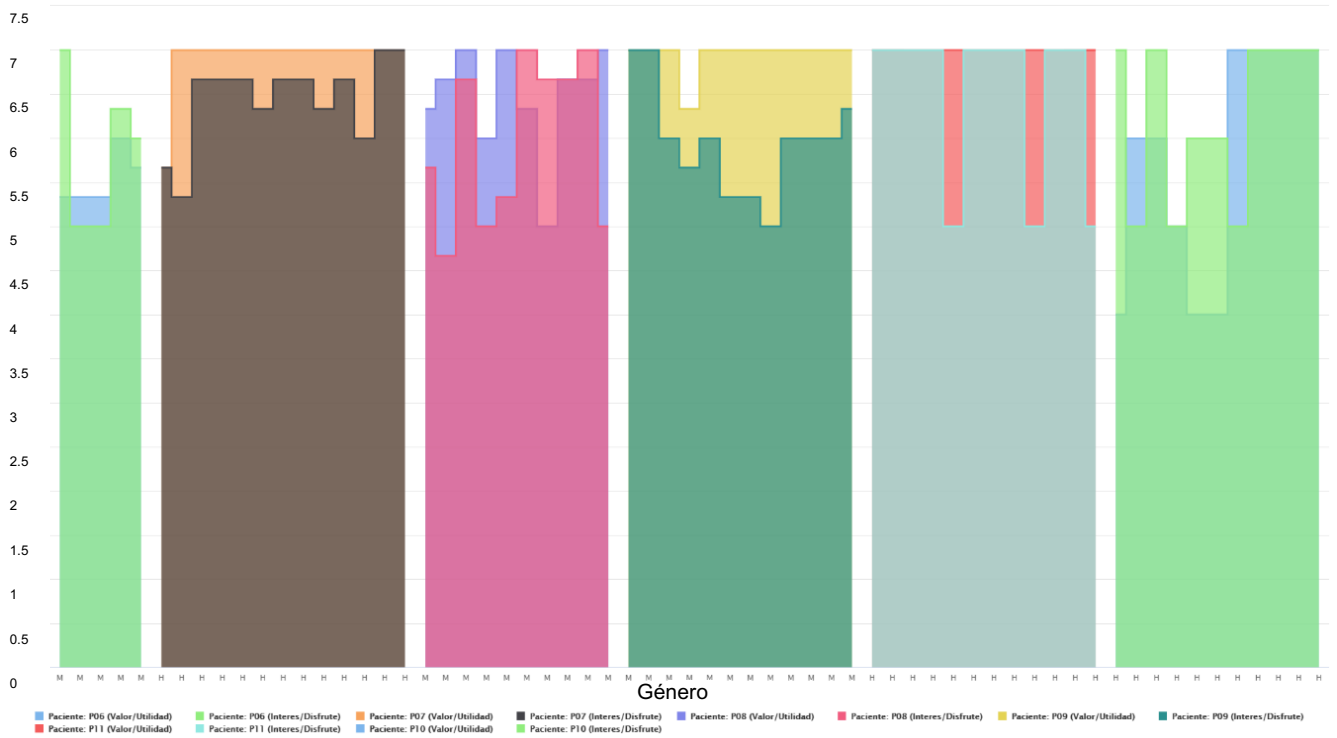


Figura 18. Valor/utilidad vs interés/disfrute por paciente

En cuanto al análisis de los datos recolectados por Gesture Therapy, se ha graficado en cada uno de los pacientes las variables de distancia total, velocidad total y presión, que son las variables que indican el avance que han tenido los pacientes comparando sus gráficos desde su primera hasta su última sesión. Estas gráficas se presentan de la figura 19 a la 22.

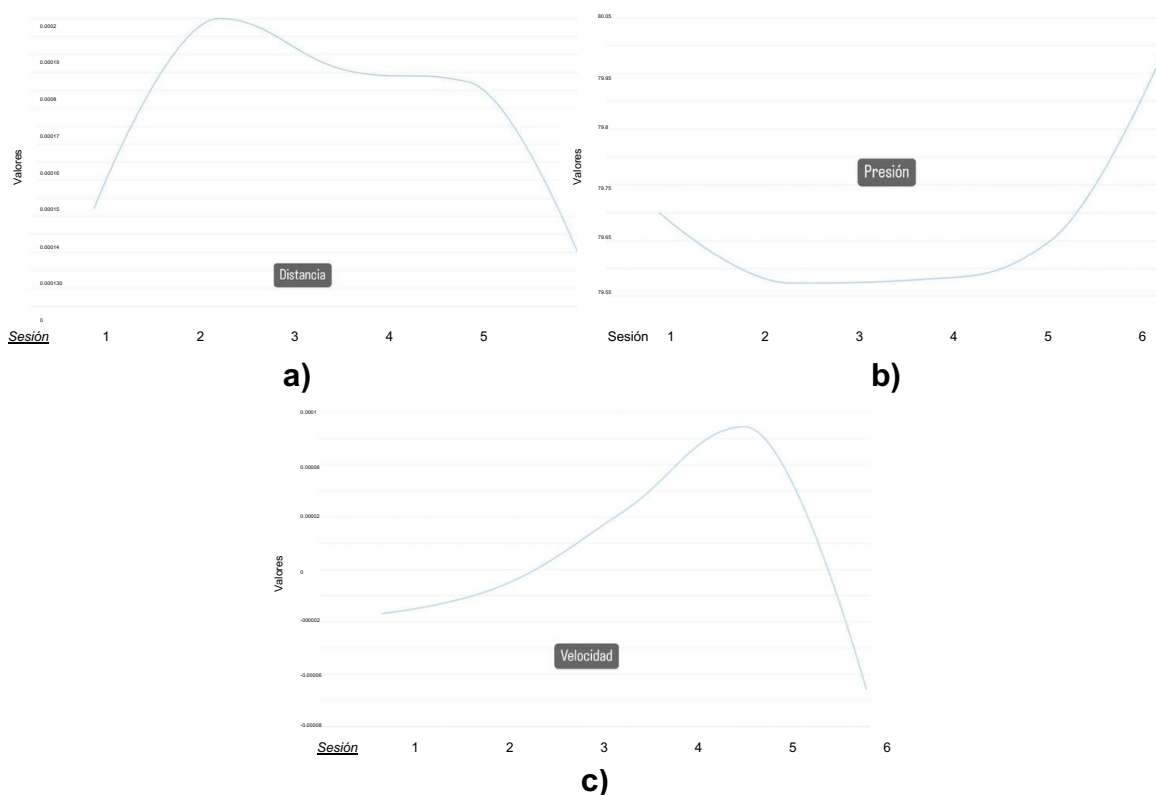


Figura 19. Gráficas del paciente P06, a) Distancia recorrida, b) Presión ejercida, c) Velocidad

En la gráfica a) de la figura 20 se observa la distancia promedio recorrida por el paciente (eje y) en cada sesión (eje x), en este caso el paciente P06 tuvo 5 sesiones y se observa que a partir de la sesión número 4 tuvo una pérdida considerable del movimiento.

En la gráfica b) se observa que la presión (eje y) va en aumento con cada sesión (eje x), contrario a lo que sucede con la variable de la distancia, es por esto, que se ha considerado que la presión se comporta muy diferente a las variables de

velocidad y distancia, pues esto se ve reflejado en los resultados de todos los pacientes.

Finalmente, en la gráfica c), y como se mencionó antes, la variable de velocidad (eje y) se comporta de manera similar a la de distancia, pues el paciente perdió velocidad en sus extremidades y en la última sesión (eje x) volvió a tener resultados similares a los del inicio de su tratamiento.

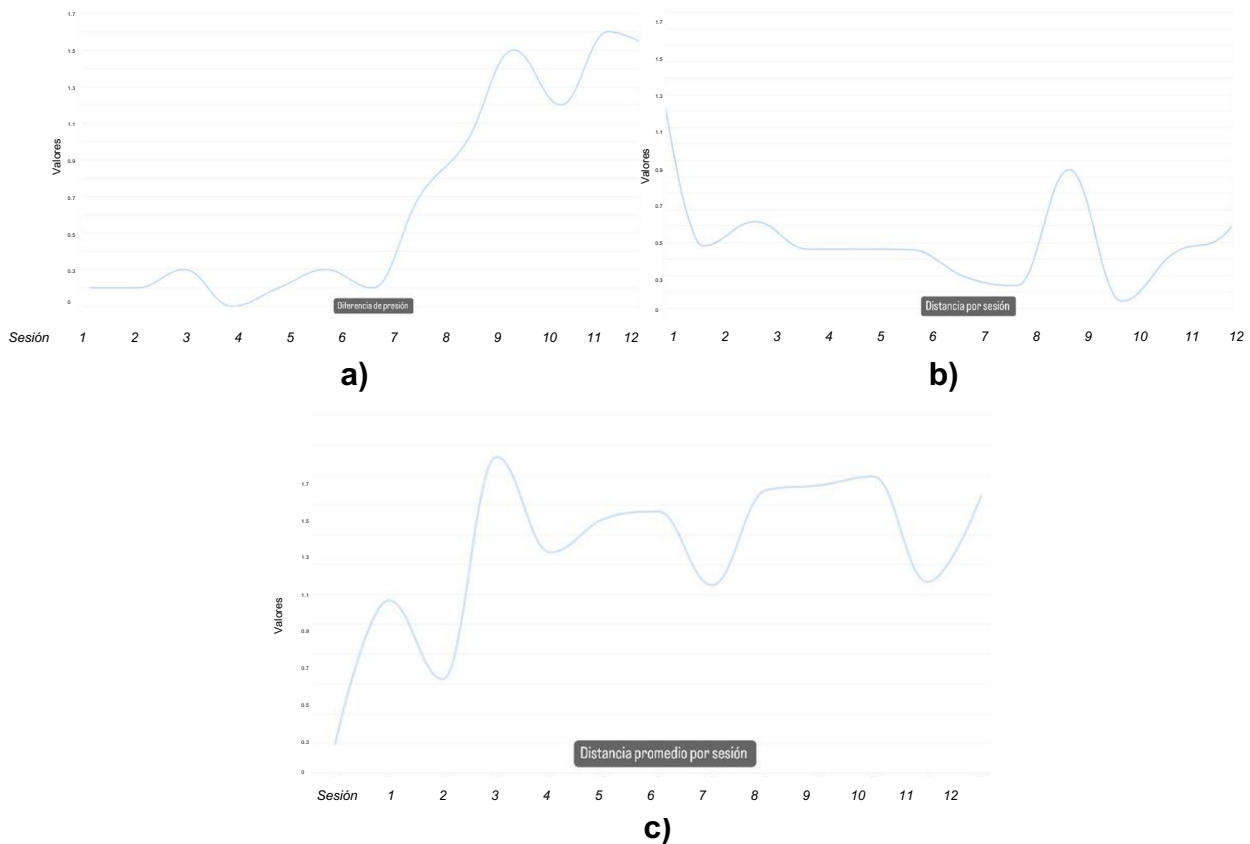


Figura 20. Gráficas del paciente P07. a) Presión b) Distancia por sesión c) Distancia promedio

En la gráfica a) de la figura 21 se presenta la diferencia de presión (eje y) de cada sesión (eje x) del paciente P07, que aumentó considerablemente a partir de la sesión 7 y esta se mantuvo en aumento constante durante el tratamiento.

Este paciente tuvo una falta de 4 días de la sesión 3 a la 4 y se observa que hubo una tendencia bajista entre estas sesiones, particularmente en la gráfica b).

En la gráfica c) de la figura 21 se muestra el promedio de la distancia recorrida por el paciente (eje y) en cada sesión (eje x), en este caso se observa un aumento considerable, a excepción de que se puede observar que en todas sus variables se desmejora entre la sesión 7 y la 8, tomando el tiempo entre estas sesiones hay una distancia de 4 días, siendo una distancia máxima de 2 días entre cada sesión, se asume que esta desmejora es debido a ese aumento de tiempo entre las sesiones.

La gráfica a) de la figura 22 muestra la velocidad (eje y) por sesión (eje x) y como se esperaba nuevamente, hubo una desmejora de entre la sesión 7 y 8, sin embargo, esta variable pudo recuperarse de manera casi inmediata, por lo menos en este paciente pues para el paciente P06 no fue la misma situación.

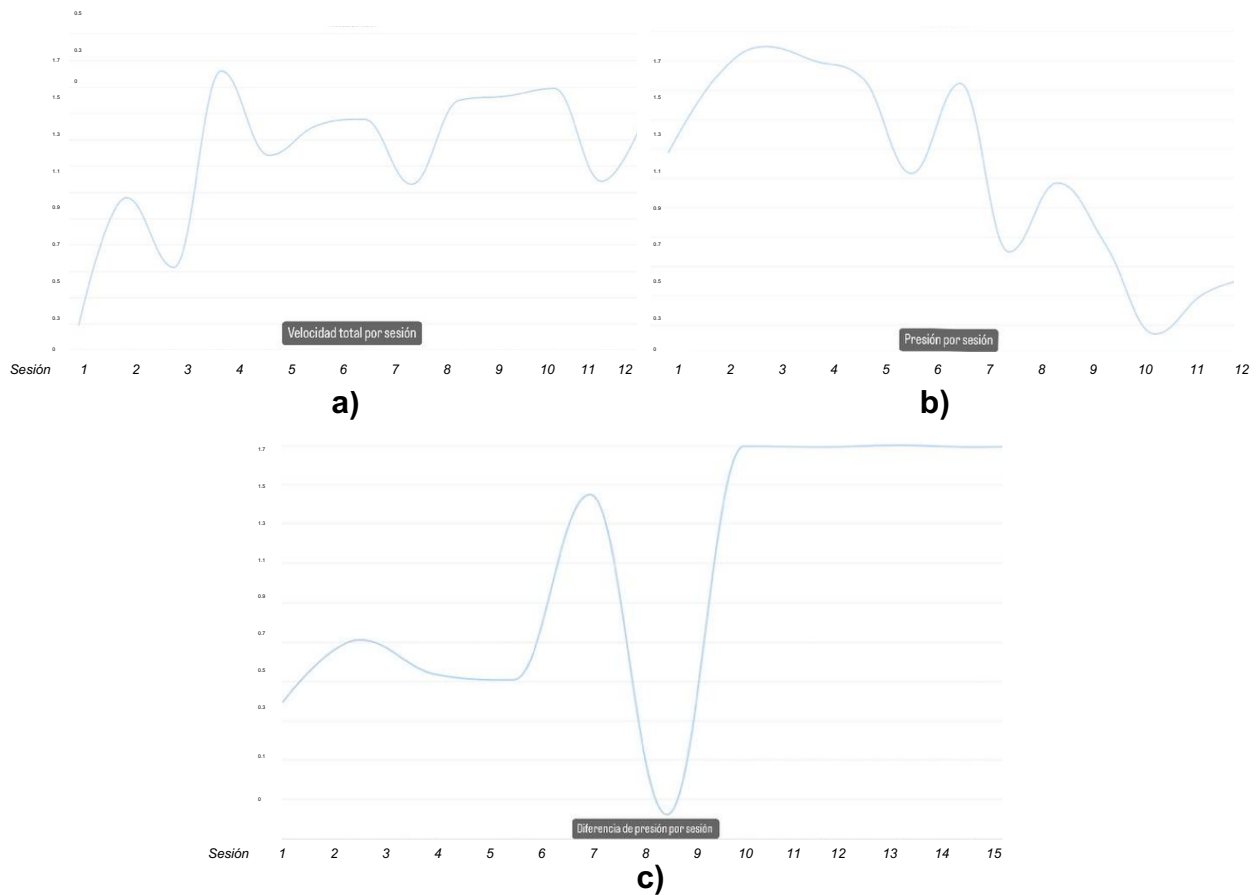


Figura 21. Gráfica de variables del paciente P07 a) Velocidad total por sesión b) Presión por sesión c) Distancia de presión por sesión

La variable de presión se comporta de manera diferente a la velocidad y la distancia, pues un paciente puede mejorar la presión y no necesariamente la velocidad y la distancia, como se observa en las últimas 3 gráficas del paciente P07 mostradas en la figura 22.

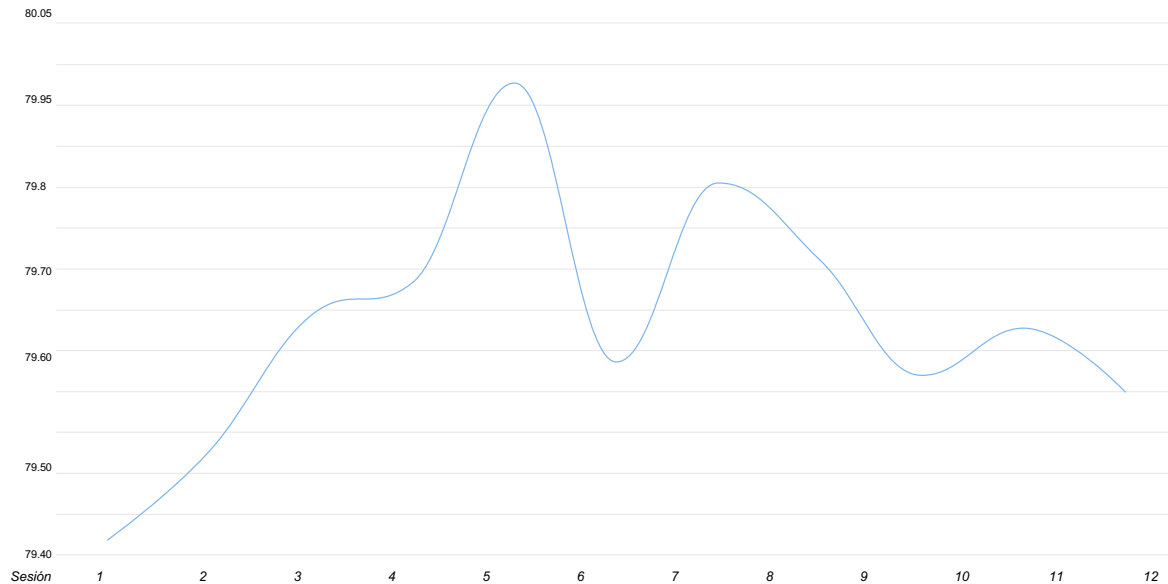


Figura 22. P08 – Promedio de presión por sesión

Otra de las variables importantes que se consideraron fue la del promedio de la presión (eje y) por sesión (eje x), la figura 23 muestra que entre la sesión 5 y 6 del paciente hubo una variación, entre estas sesiones no pasaron más de dos días, pero sí hubo una variación en la cantidad de archivos entre estas sesiones, recordemos que cada archivo es un juego por lo que se puede decir que esta variable depende también del tipo de juego que se hace en las sesiones de los pacientes.

En otra etapa para el análisis, se estudiaron los días de cada sesión de los pacientes y, como se esperaba, esta variable sí que es relevante para obtener buenos resultados en el tratamiento. A continuación, se muestra esta relación de mejoría o desmejora entre cada sesión y los días que pasaron mostrados por cada paciente.

Básicamente se obtuvo el porcentaje de mejora considerando la relación entre el último valor obtenido en el movimiento y la velocidad, es decir, si en la primera

sesión un paciente tuvo un valor de 5 en velocidad y en la segunda, un valor de 7.5, hubo una mejora del 50%, para el cálculo del siguiente porcentaje de mejora o desmejora, se tomará como referencia el valor de la última sesión (7.5).

	Sesión	Fecha	Días desde la última sesión	Movimiento	Velocidad
P06	1	05-nov	0	0	0
	2	07-nov	2	31%	34%
	3	09-nov	2	95%	-3%
	4	12-nov	3	82%	-8%
	5	14-nov	2	-125%	-71%

Tabla 13. P06 - Tabla de frecuencia

	Sesión	Fecha	Días desde la última sesión	Movimiento	Velocidad
P07	1	03-dic	0	0	0
	2	05-dic	2	-76%	-68%
	3	06-dic	1	30%	-27%
	4	10-dic	4	-27%	45%
	5	12-dic	2	1%	-22%
	6	13-dic	1	-2%	31%
	7	17-dic	4	-36%	-50%
	8	19-dic	1	-19%	51%
	9	20-dic	1	297%	200%
	10	26-dic	6	-85%	-29%
	11	27-dic	1	87%	0%
	12	02-ene	2	21%	-80%
	13	03-ene	1	12%	46%

Tabla 14. P07 – Tabla de frecuencia

Sesión	Fecha	Días desde la última sesión	Movimiento	Velocidad
1	27-dic	0	0	0
2	29-dic	2	31%	34%
3	31-dic	2	95%	20%
4	02-ene	2	82%	15%
5	04-ene	2	-7%	30%
6	06-ene	2	25%	5%
7	08-ene	2	36%	15%
8	11-ene	2	30%	14%
9	13-ene	2	23%	20%
10	15-ene	2	14%	5%
11	17-ene	2	10%	8%

Tabla 15. P08 – Tabla de frecuencia

Sesión	Fecha	Días desde la última sesión	Movimiento	Velocidad
1	07-ene	0	0	0
2	09-ene	2	38%	58%
3	11-ene	2	10%	-1%
4	14-ene	3	9%	-21%
5	16-ene	2	12%	39%
6	18-ene	2	44%	13%
7	21-ene	3	-35%	-30%
8	23-ene	2	3%	30%
9	25-ene	2	-9%	3%
10	28-ene	3	6%	-25%
11	30-ene	2	31%	8%
12	01-feb	2	10%	10%

Tabla 16. P09 – Tabla de frecuencia

Por otra parte, se ha graficado el avance que han tenido todos los pacientes en la figura 27. En ella se observa que el paciente 07 tiene bastante avance, sin embargo, por la frecuencia de sus terapias ese avance resultó de poca ayuda, sobre todo en las variables de velocidad y distancia.

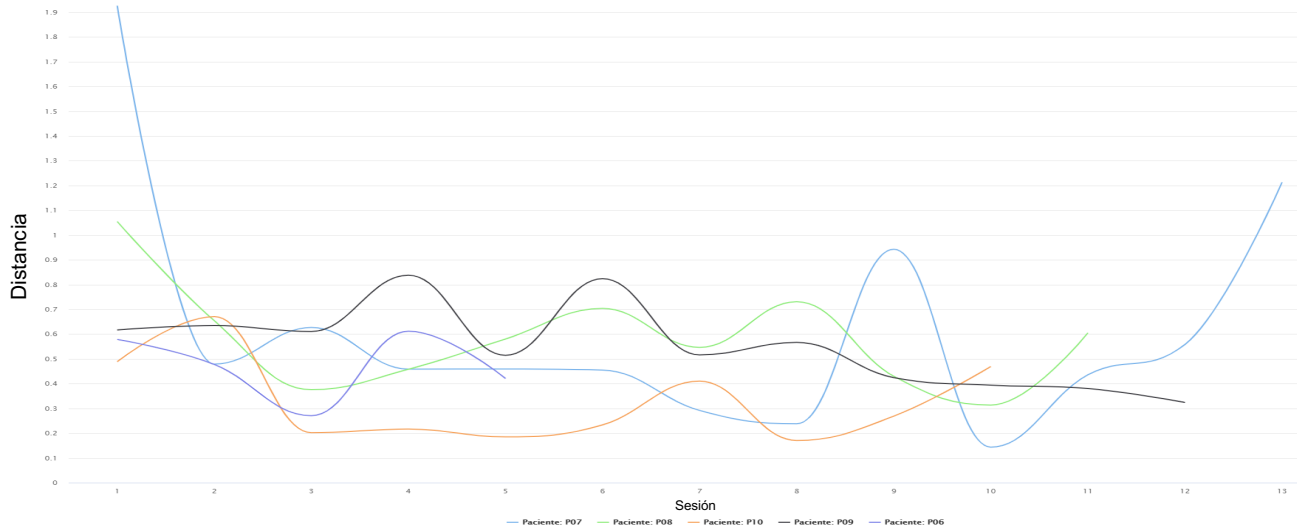


Figura 23. Distancia total por paciente

Varianza de Distancia_total por Paciente

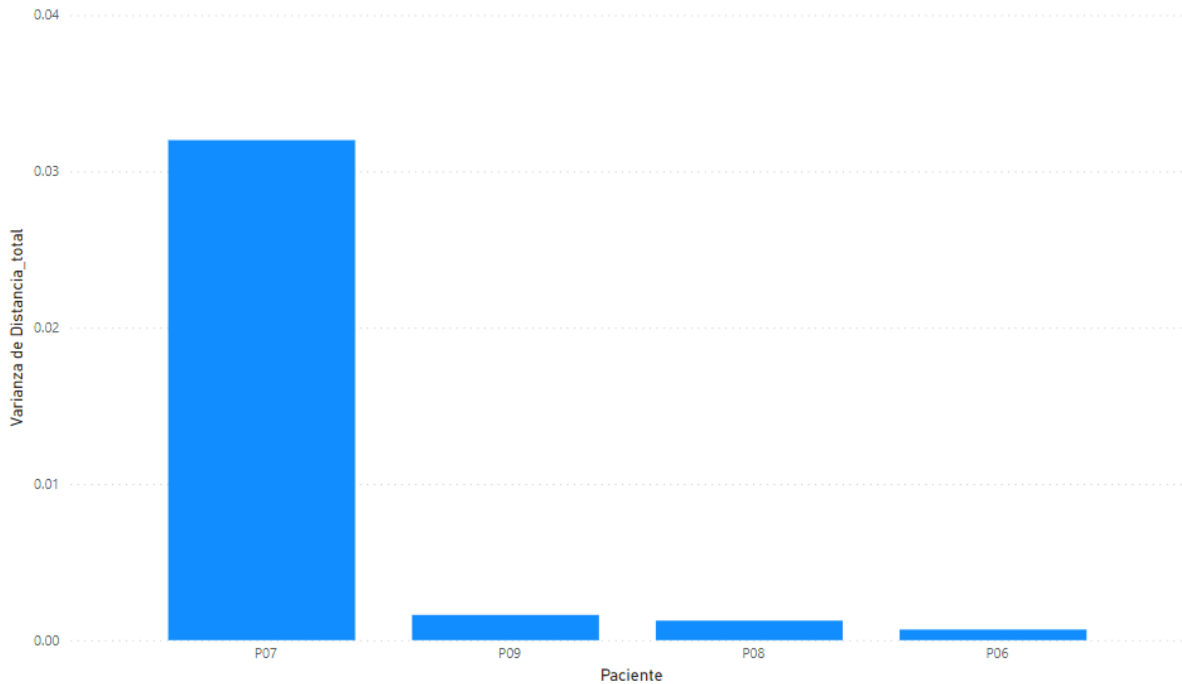


Figura 24. Varianza de la distancia por paciente

Ahora en la figura 28 se puede observar que el P07 es el que tiene una varianza bastante grande entre sus sesiones, esto pasa con la variable de la velocidad, tomando en consideración la tabla de frecuencia esta gran diferencia entre sus sesiones se debe a que este paciente llega a faltar hasta 4 días, y después va cada

día y enseguida vuelve a faltar 4 días, y para “reponer” sus terapias a las que faltó acude cada día a sus sesiones, esto ocasiona una gran diferencia de las variables entre sus sesiones, pero de poco ha funcionado este tipo de recurrencia a su tratamiento pues finalmente vuelve como si iniciara desde cero su tratamiento, así que a pesar de que tuvo un tratamiento con bastantes sesiones de por medio, no sirvieron casi de nada ya que no tuvo la consistencia que debía.

4.5 Proceso de minado a través de regresión lineal

Se ha aplicado el algoritmo de regresión lineal para ver qué tanta relación hay entre las variables de velocidad y distancia, para ello se tomaron los datos antes mencionados, en la figura 29 se muestra la carga de los datos y la selección de estos atributos del conjunto del dataset.

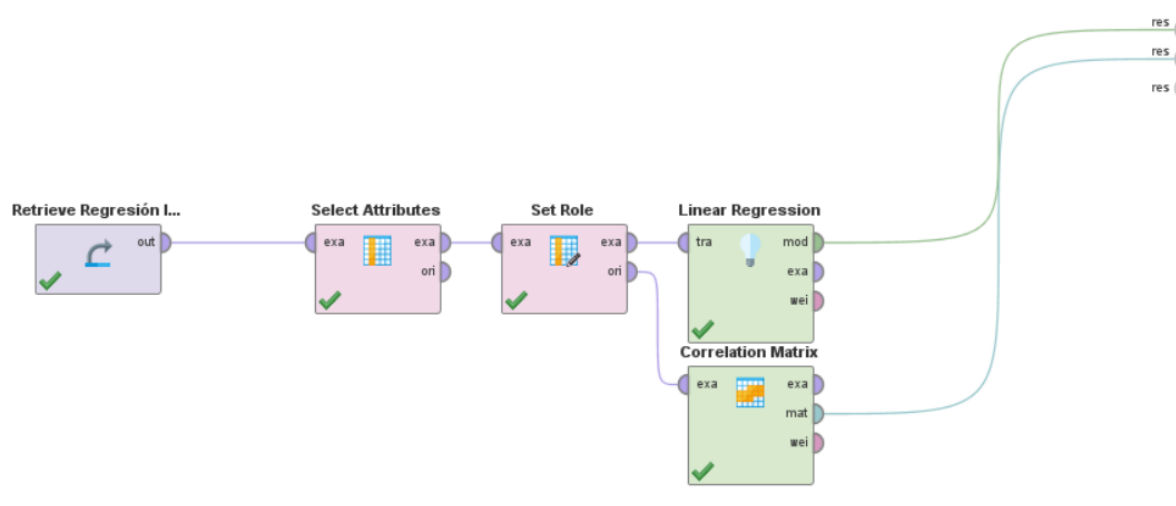


Figura 25. Diseño de regresión lineal

En el apartado de “set role” se ha escogido la variable dependiente que en este caso podría ser cualquiera de las dos, finalmente se coloca el proceso de regresión lineal para obtener los resultados y por último y no menos importante la matriz de correlación que arroja los resultados que se buscan en el estudio. Primero se obtuvo de la matriz de correlación el nivel de relación que tienen las variables, con este resultado se puede comprobar que efectivamente, sí existe una correlación positiva y directa entre estas dos variables, y que además es muy fuerte, este nivel va de 0 a 1 y cuanto más se acerque a 1 la correlación es más fuerte.

Primer atributo	Segundo atributo	Correlación
Distancia_total	Velocidad_total	0.875

Tabla 17. Matriz de correlación

Atributo	Coefficiente
Distancia_total	0.004
Intercepción	-0.000

Tabla 18. Coeficiente e intersección de regresión lineal

En este caso basados en la fórmula de regresión lineal el valor de B_0 de la intersección es .004 y la pendiente de -0.000. Haciendo hincapié en que el propósito de la aplicación del algoritmo fue buscar el nivel de relación existente entre dos variables, no el de predecir en sí.

Analizando los resultados de la matriz de correlación entre la variable de distancia y la presión se puede observar claramente que no existe relación alguna entre estas dos variables.

Primer atributo	Segundo atributo	Correlación
Distancia_total	Presión	-0.000

Tabla 19. Matriz de correlación

5 Conclusiones

5.1 Conclusiones

Definitivamente el empleo del análisis de datos juega un papel de suma importancia en la mejora de la toma de decisiones por parte de los terapeutas y los pacientes, pues se han encontrado puntos a considerar por parte de ambos elementos que afectan directamente a los resultados obtenidos de los pacientes en sus terapias. A continuación, los puntos encontrados:

Basándose en las tablas de mejora, los pacientes que fueron constantes han tenido mejores resultados que los que no lo hicieron. Un paciente constante pasa más de dos días entre sus sesiones de terapia, los pacientes más constantes fueron: P08 y P09 con un 303% de mejora de promedio, considerablemente mayor que los que pasaron más días entre sus terapias o no completaron su tratamiento, como el P06, que apenas mejoró un 35 %, además, se ha hecho una comparación entre las gráficas de interés/disfrute y la tabla de mejoras, esta variable está directamente relacionada con los pacientes con mejores resultados; gracias a los resultados obtenidos por la matriz de correlación, se pudo concluir que la presión ejercida en la terapia es independiente a las variables de velocidad y distancia recorrida por las extremidades.

Los pacientes con buenos resultados sienten menos presión, aunque no necesariamente menos dolor, como el paciente P09, que tuvo uno de los picos más altos en la gráfica de dolor, pero sus mejoras fueron de las mejores, hablando de la frecuencia de las terapias, los pacientes que llegan a faltar al menos 4 días a su sesión pueden decaer casi como si empezaran de cero, como es el paciente P07, aunque tuvo un tratamiento más completo que el paciente P06, su constancia no fue buena y en las sesiones en las que se presentaba 4 días después, su avance fue negativo, como sucedió entre la sesión 9 y 10, que pasó casi una semana, tomando en consideración la tabla de mejoras, una sesión de 15 días representa apenas un 23% de beneficios a la de un mes, además la falta de interés es una de las principales causantes de las terapias poco o nada eficientes.

5.2 Trabajo a futuro

Hay más algoritmos que son posibles de aplicar a los datos obtenidos, ejemplo de ellos es a priori, permitiendo obtener posibles escenarios dadas ciertas condiciones o resultados arrojados en las encuestas de los pacientes. De la misma forma, es posible aplicar un algoritmo de predicción si se llegara a contar con datos de más pacientes, permitiendo de esta forma poder generalizar la información y obtener mejores resultados. Dado el limitado número de pacientes analizados, no es posible poder deducir con la precisión deseada algunos factores o predicciones que sí que serían de utilidad, además de mejorar la precisión de las estadísticas presentadas, dicho de otra forma, este estudio se ha quedado en parte a una descripción del proceso de rehabilitación de cada uno de los pacientes analizados.

Además, este trabajo de investigación puede ser tomado como referencia para la aplicación y adaptación de la metodología CRISP-DM a un caso de estudio, por otra parte, también aporta al apoyo de los juegos serios y el empleo de tecnologías emergentes para la rehabilitación y mejora de las terapias que aplican los pacientes con discapacidades motoras.

Referencias

- Campos, T. (2016). Xataka.
<https://www.xataka.com/investigacion/gesture-therapy-es-un-videojuego-desarrollado-por-mexicanos-que-ayuda-a-recobrar-la-movilidad>
- INAOE. (1 de 11 de 2018). Promueven sistema mexicano para rehabilitar las extremidades superiores. El Hospital.
<https://www.elhospital.com/temas/Promueven-sistema-mexicano-para-rehabilitar-la-movilidad-de-las-extremidades-superiores+127915>
- Mayo Clinic. (09 de 02 de 2021). Mayo Clinic.org.
<https://www.mayoclinic.org/es-es/diseases-conditions/stroke/diagnosis-treatment/drc-20350119>
- Organización Mundial de la Salud (OMS). (26 de 11 de 2015). Instituto Mexicano del Seguro Social.
[http://www.imss.gob.mx/salud-en-linea/enfermedad-vascular-cerebral#:~:text=La%20enfermedad%20vascular%20cerebral%20\(EVC,y%20bajos%20los%20m%C3%A1s%20afectados.](http://www.imss.gob.mx/salud-en-linea/enfermedad-vascular-cerebral#:~:text=La%20enfermedad%20vascular%20cerebral%20(EVC,y%20bajos%20los%20m%C3%A1s%20afectados.)
- STIMULUS. (03 de 08 de 2017). Stimulus.
<https://stimuluspro.com/blog/la-mineria-de-datos-aplicada-a-la-salud/>
- Dávila Hernández, F., Yovannys, I., & Corales, S. (n.d.). Técnicas de minería de datos aplicadas al diagnóstico de entidades clínicas Data mining techniques applied to diagnosis of clinical entities. In *Revista Cubana de Informática Médica* (Vol. 2012, Issue 2). <http://scielo.sld.cu>
- Antonio Gutiérrez Pérez, L., Gutiérrez Cruz, D., Molina, R. R., Liliana, C., Páez, R., Aarón, Y., Fernández, A., Rivera, B. S., Rebeca, A., Cruz, G., & Duran López, V. M. (n.d.). Discovery of knowledge on incidence of cancer type for terminal patients through data mining.
- Rancan., Rodríguez, T., En, B., De, M., Reparaz, D., Rancan, D., Rodríguez, C., & - Martínez, G. (n.d.). determinacion de la eficacia de la braquiterapia
- Devesa Gutiérrez, I., María Elena Mazadiego González, D., Miguel Ángel Baldomero Hernández Hernández, T., & Héctor Antonio Mancera Cruz, T. (2014). Rehabilitación del paciente con enfermedad vascular cerebral (EVC). In *Rev Mex Med Fis Rehab* (Vol. 26, Issue 4).
www.medigraphic.org.mx<http://www.medigraphic.com/medicinafisica>
- Sánchez-Villavicencio, I., Hernández-Franco, J., Sucar, E., & Leder, R. S. (2009). Actividades simuladas de la vida diaria para rehabilitación motora del miembro superior en pacientes con enfermedad vascular cerebral Artículo original. In *Arch Neurocién (Mex)* (Vol. 14, Issue 4).
- Sucar, L. E., Orihuela-Espina, F., Velazquez, R. L., Reinkensmeyer, D. J., Leder, R., & Hernandez-Franco, J. (2014). Gesture therapy: An upper limb virtual reality-based motor rehabilitation platform. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 22(3), 634–643.
<https://doi.org/10.1109/TNSRE.2013.2293673>

- Alcalde, A. (2018, 04 03). El baúl del programador.
<https://elbaultdelprogramador.com/aprendizaje-nosupervisado-reglas-avanzadas/>
- Buitrago, B. (2020, 09 24). Medium.
<https://medium.com/iwannabedatadriven/machine-learning-modelos-de-regresi%C3%B3n-i-d293ae235e9a>
- Burdea, G. (2018, 07 12). CAIP Center.
http://www.ti.rutgers.edu/publications/papers/2002_vrmhr_burdea.pdf
- Data Science Team. (2020, 08 02). Data Science.
<https://datascience.eu/es/aprendizaje-automatico/algoritmos-de-agrupacion-y-su-importancia-en-el-aprendizaje-de-las-maquinas/>
- Geografía, I. N. (2021). Gobierno de México.
<https://www.gob.mx/salud/prensa/531-en2021-ictus-o-enfermedad-vascular-cerebral-ocasiono-mas-de-37-mil-decesos-en-mexico>
- Góngora, A. (2021). Gobierno de México.
<https://www.gob.mx/salud/prensa/531-en2021-ictus-o-enfermedad-vascular-cerebral-ocasiono-mas-de-37-mil-decesos-en-mexico>
- Mayo Clinic. (2021, 10 29).
<https://www.mayoclinic.org/es-es/diseases-conditions/stroke/symptoms-causes/syc-20350113>
- MedicalExpo. (n.d.).
<https://www.medicalexpo.es/prod/evolv/product-84105-533579.html>
- Montesano, A. (2018, 01 01). Multiple Sclerosis.
[https://www.msard-journal.com/article/S2211-0348\(17\)30267-5/fulltext](https://www.msard-journal.com/article/S2211-0348(17)30267-5/fulltext)
- Regalón, J. A. (2019, 12 1). Revista Cubana de Informática Médica.
http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1684-18592019000200140
- Santander. (2021, 04 14).
<https://www.becas-santander.com/es/blog/motivacion-intrinseca-y-extrinseca.html>
- Social, I. M. (2015, 11 26). Instituto Mexicano del Seguro Social.
<https://www.imss.gob.mx/salud-en-linea/enfermedad-vascular-cerebral#:~:text=Se%20conoce%20como%20enfermedad%20vascular,2%20subtipos%3A%20isku%C3%A9mica%20y%20hemorr%C3%A1gica.>
- Takeshi, A. (n.d.). Medium.
<https://medium.com/@apolotakeshi/understanding-crisp-dm-with-my-cats-weight-and-covid-19-s-prediction-e84c42309a04>
- Yoon, J.-G. (2015). Department of Physical Therapy, Cheon-An City Geriatric Hospital, Republic of Korea.
https://www.jstage.jst.go.jp/article/jpts/27/3/27_jpts-2014-625/_pdf
- Zannino, G. D. (2008, 12 4). National Library of Medicine.
<https://pubmed.ncbi.nlm.nih.gov/19331782/>
- Zucchella, C. (2014, 12 04). National Library of Medicine.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4264781/>