

Trabajo Fin de Grado

Minería de procesos en el sistema de gestión de la salud

Process mining in healthcare system management

Autor:

Daniel Sanz Valtueña

Directores

Jorge Albareda Albareda

David Ubide Alaiz

Ponente

Cristian Mahulea

Grado en Ingeniería de Tecnologías Industriales

Escuela de Ingeniería y Arquitectura

2022-2023

Agradecimientos

Me gustaría agradecer a mis directores, Jorge Albareda y David Ubide, la dedicación, apoyo y entusiasmo que han depositado en mí y en este proyecto. A Jorge por mostrarme aspectos del hospital y la vida de los médicos que sólo se ven de puertas adentro; y a David por enseñarme gran parte de lo que he aprendido durante el trabajo, pero también lecciones de vida, como a hacer las cosas de manera profesional y no antes de tiempo.

También me gustaría agradecer a mi profesor y ponente de este proyecto, Cristian Mahulea, por su colaboración y por despertar en mí el interés sobre esta materia.

Al equipo de médicos residentes formado por Juan, Carolina, Irene y Elena, por su implicación en la recogida de los datos de pacientes.

A mi familia, que han sufrido y disfrutado a la vez que yo con este trabajo, gracias por vuestro apoyo incondicional y no me refiero tan solo a estos meses.

A mi familia del CBZ, habéis sido mi vía de escape número uno y aunque las temporadas a veces acaben antes de tiempo, me quedo siempre con lo que vivimos dentro y fuera de la cancha.

A mis amigos, los de siempre, de la uni, los que viven fuera y las nuevas incorporaciones del Erasmus, porque disfrutar del trayecto es fundamental y es un gustazo compartir momentos con vosotros.

Resumen

El problema abordado en este trabajo de fin de grado surge de la necesidad del *Hospital Clínico Universitario Lozano Blesa* de Zaragoza de revisar y diseñar las vías clínicas en el área de *Cirugía Ortopédica y Traumatología*. Las vías clínicas son instrumentos que se usan para coordinar, planificar y marcar la secuencia de procedimientos a seguir sobre el paciente según una patología en concreto. Abarcan el período de tiempo que va desde el ingreso del paciente hasta que este recibe el alta. Todo para conseguir la máxima eficiencia en el proceso asistencial de los pacientes.

Ante esta necesidad se propone utilizar metodologías basadas en la minería de procesos (algoritmos de descubrimiento de procesos) para realizar un seguimiento de pacientes utilizando modelos de redes de Petri obtenidos a partir de registros de datos clínicos. Este seguimiento se comparará con las vías clínicas actuales proporcionadas por el equipo de médicos que ha colaborado durante el proyecto. A partir de estas comparaciones, se extraerán conclusiones que serán de utilidad para la gestión del hospital y la toma de decisiones del personal sanitario, con el objetivo de actualizar y diseñar vías clínicas específicamente para pacientes que requieren cirugía de artroscopia de cadera y rodilla.

Este trabajo colabora con el proyecto TED2021-130449B-I00 titulado “*Evaluación y optimización de vías clínicas mediante métodos formales (VALIDATOR)*”, que tiene como objetivo principal desarrollar una metodología apoyada en datos clínicos para la revisión y diseño de vías clínicas. Como parte del proyecto, se está desarrollando un servidor web que cumple la doble funcionalidad de servir como base de datos y aplicación de gestión de la información del Hospital Clínico Universitario. Se implementarán en este servidor web las metodologías de seguimiento de pacientes desarrolladas a lo largo del trabajo de fin de grado.

Índice

RESUMEN	5
1 Introducción	8
1.1 Dominio del problema	8
1.2 Objetivos del TFG	9
1.3 Técnicas utilizadas para su resolución	10
1.4 Alcance	11
2 Campo de estudio	12
2.1 Minería de procesos	12
2.2 Modelado de procesos con redes de Petri	14
2.3 Modelado de procesos con árboles de procesos	16
2.4 Algoritmos de minería de procesos	18
2.5 Análisis clásico de datos	21
3 Minería de procesos para el diseño de vías clínicas	21
3.1 Trayectorias de pacientes que padecen de Sepsis	22
3.2 Estudio de pacientes con prótesis total de cadera	24
3.3 Diseño de vías clínicas	26
3.4 Despliegue de metodologías en servidor web	27
4 Tratamiento de datos y colaboración con el hospital	28
4.1 Recogida de datos	28
4.2 Procesado de datos	28
4.3 Colaboración con el personal sanitario	30
5 Análisis de riesgos del proyecto	31
6 Problemas encontrados	32
7 Resultados	33
7.1 Algoritmos óptimos de minería de procesos	33
7.2 Comparación de vías clínicas con registros de pacientes	33
7.3 Implementación de metodologías en servidor web	34
7.4 Análisis clásico de datos	35
8 Conclusiones	36
9 Bibliografía	37
ANEXOS	38
Anexo 1. Vías clínicas existentes	38
Anexo 2. Análisis clásico de datos	41
Anexo 3. Interfaz del servidor web	46
Anexo 4. Resultados completos para cada caso de estudio	49
Anexo 5. Resultados del análisis clásico de datos	58
Anexo 6. Gestión del proyecto	62

1 Introducción

Durante este capítulo se realizará la introducción a este trabajo de fin de grado. Se expondrán el dominio del problema y objetivos del trabajo, así como las técnicas utilizadas para su resolución y alcance.

1.1 Dominio del problema

Este Trabajo de Fin de Grado (TFG) está relacionado con una tarea del proyecto TED2021-130449B-I00 titulado “*Evaluación y optimización de vías clínicas mediante métodos formales (VALIDATOR)*”, y se ha realizado en la Escuela de Ingeniería y Arquitectura (EINA) así como en el Hospital Clínico Universitario Lozano Blesa (HCU). El objetivo principal de este proyecto es desarrollar una metodología apoyada en datos clínicos para la revisión y diseño de vías clínicas.

Las vías clínicas (en la *Figura 1* se presenta un ejemplo de una parte de una vía clínica) consisten en un conjunto de recomendaciones y protocolos, basadas en una revisión sistemática de la evidencia y en la evaluación de los riesgos y beneficios de las diferentes alternativas de tratamientos. Abarcan el período de tiempo que va desde el ingreso del paciente hasta que este recibe el alta. Todo ello con el objetivo de aportar a los usuarios (pacientes) la mejor atención médica y de aprovechar los recursos existentes de forma óptima. Se tratan por tanto de procesos dinámicos que se deben implementar en la práctica para así poder valorar e identificar las intervenciones más efectivas.

<u>ARTROPLASTIA TOTAL DE RODILLA</u>	
Preingreso:	● Radiografía, telemétricas ● Causa de la intervención ● Consentimientos de la cirugía ● Valoración preanestésica en consultas
PrelQ:	● Comprobar valoración por anestesia/ petición anestesia intrahospitalaria ● Comprobación de CI ● Pauta de medicación + Antibiótico en inducción ● Alérgico a penicilina si / no ● Pauta de HBPM 12h preiq ● Solicitud de PC según criterios actuales

Figura 1. Fragmento de vía clínica [*Anexo 1*]

Estas vías clínicas pueden ser específicas para una enfermedad en particular o pueden ser genéricas para un grupo de enfermedades relacionadas. Es importante señalar que las vías clínicas no son transferibles de un hospital a otro debido a las diferencias en los recursos disponibles, los tipos de pacientes y otros factores. En este caso, el trabajo está centrado en patologías del área de *Cirugía Ortopédica y Traumatología* del HCU, concretamente en las intervenciones quirúrgicas de prótesis de cadera y rodilla [*Anexo 1*].

Uno de los motivos por los que surge este proyecto y por consiguiente este TFG, es que en los hospitales hay muy pocas vías clínicas diseñadas. Todos los pacientes

son atendidos con el mejor trato posible, pero no siempre están pautados rigurosamente las recomendaciones y protocolos a seguir. Además, se va a intentar actualizar la forma de representar estas vías. En el *Anexo 1* se muestra cómo ha proporcionado el personal médico las vías clínicas actuales para las patologías citadas.

La creación de nuevas vías clínicas es un proceso laborioso que puede tardar años en completarse. Por eso, la labor de este trabajo es la de desarrollar una metodología para hacer un seguimiento de los pacientes, comparar las actividades registradas con las de las vías actuales y así diseñar vías clínicas definitivas, todo ello bajo la supervisión de un equipo médico.

Por otro lado, la minería de procesos consiste en extraer conocimiento valioso a partir de los registros de actividades que se producen en los sistemas de información, mediante la identificación de patrones, relaciones y estructuras que se encuentran en los datos [1-4]. A lo largo del trabajo se utilizarán herramientas de esta disciplina científica.

Cabe destacar que estos tipos de análisis han ganado importancia en los últimos años y se han extendido a diversas áreas y sectores, como es el caso del campo de la salud. En este contexto, el sistema sanitario genera una gran cantidad de datos y eventos que pueden ser analizados. Mediante el uso de técnicas de minería de procesos, se pueden identificar patrones y comportamientos, siendo de particular interés para este trabajo modelos matemáticos. Estos podrán ser utilizados no solo como herramienta de tratamiento de datos, sino también en la toma de decisiones en los hospitales. De tal manera que este proyecto tiene como fin facilitar el diseño de nuevas vías clínicas al personal médico. Así pues, se propone utilizar métodos de análisis matemático, basados en modelos de redes de Petri [5,13,14], para el modelado de procesos dentro del hospital que ayudarán a la toma de decisiones en el sistema de salud.

1.2 Objetivos del TFG

Los objetivos que se abordan en este trabajo son:

- **(O1) Creación de un sistema de seguimiento de pacientes a partir de datos clínicos.** Se identificarán los datos clínicos relevantes y cómo se pueden extraer y procesar para la minería de procesos. A partir de ahí, se aplicarán los algoritmos que mejor procesen estos datos y con los que se consigan los resultados que mejor representen las vías clínicas en forma de redes de Petri. Para ello se estudiarán distintos algoritmos de descubrimiento de procesos. Se buscará que esta metodología sea sistemática, lo que permitirá identificar y diseñar vías clínicas más eficientes y efectivas a partir de datos clínicos.
- **(O2) Optimización de vías clínicas para pacientes con patologías de artroplastia de cadera y rodilla.** La revisión de los procedimientos y la inclusión de nuevas tecnologías en el tratamiento de estas patologías ha hecho que se planteen nuevos protocolos de actuación para que el paciente permanezca ingresado el menor tiempo posible. Los especialistas de cadera y rodilla han proporcionado borradores de estos nuevos procedimientos, por lo tanto, la labor a desempeñar en este trabajo será la de comprobar qué actividades se realizan sobre pacientes de estas patologías en el hospital, diseñar y proponer una vía clínica basándose en este seguimiento de pacientes y comprobar si esta propuesta coincide con el borrador. De esta manera, el personal sanitario podrá sacar conclusiones y revisar o diseñar nuevas vías clínicas.
- **(O3) Inserción del método de seguimiento de pacientes.** Los algoritmos de minería de procesos que se usarán a lo largo del trabajo son los mismos que se

han implementado en el servidor que se está desarrollando en el proyecto al que está asociado este TFG (*“Evaluación y optimización de vías clínicas mediante métodos formales”*). Dicho servidor consta de una herramienta con la que se trata de ayudar al personal médico en la toma de decisiones y gestión de la administración en el hospital. Se podrá por tanto implementar en la herramienta mencionada, la manera que se desarrolle durante este trabajo de representar las vías clínicas y hacer un seguimiento de los pacientes.

1.3 Técnicas utilizadas para su resolución

A lo largo de este proyecto, concretamente para cubrir los objetivos (O1) y (O2), se ha llevado a cabo el procedimiento de extracción y tratamiento de datos para la minería de procesos mostrado en la *Figura 2*.

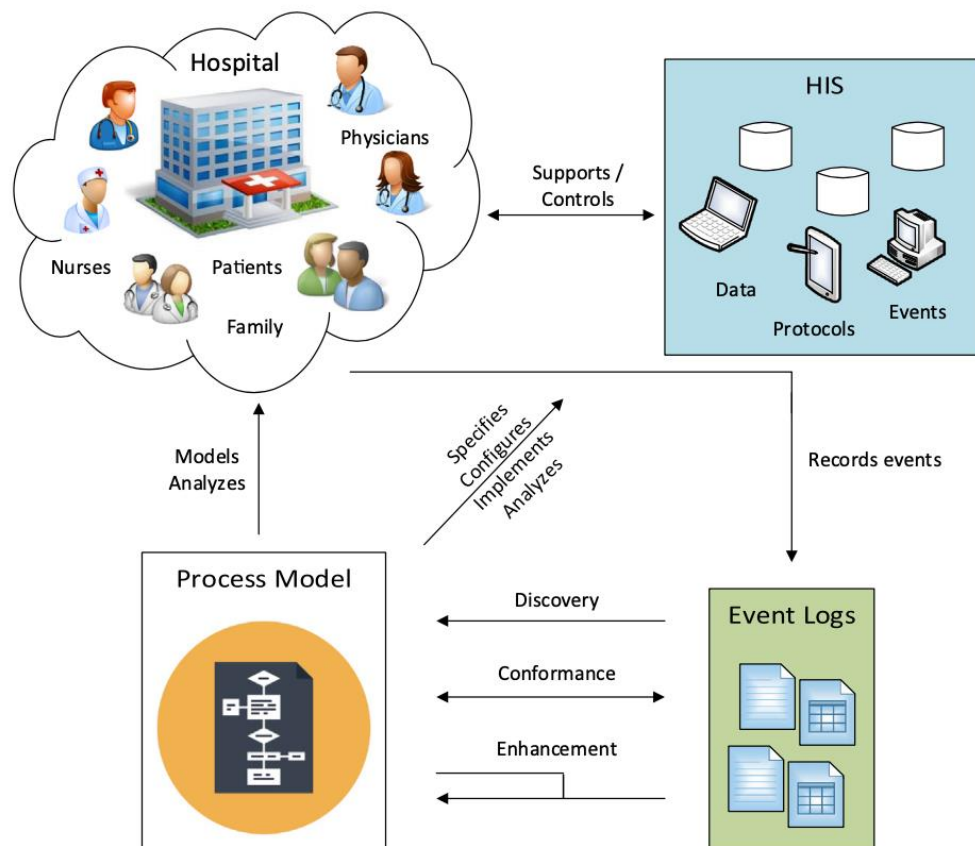


Figura 2. Minería de procesos en el sistema de gestión de la sanidad [1]

Los datos clínicos de pacientes necesarios para el desarrollo del trabajo son controlados y almacenados en los sistemas de información del hospital (*Healthcare Information System (HIS)*) y en este caso han sido proporcionados por el jefe del servicio de Cirugía Ortopédica y Traumatología del Hospital Clínico Universitario y los Médicos Internos Residentes (MIR) de su servicio, que han sido los que han llevado a cabo la recogida de datos. Simultáneamente, ha habido un trabajo de investigación y formación en minería de procesos para aplicar correctamente las técnicas de análisis sobre estos datos.

Una vez extraída esta información (*event logs* o registro de eventos), se le ha proporcionado el formato necesario para comenzar a diseñar la vía clínica, que en este contexto, será considerada un modelo del proceso. Los algoritmos de descubrimiento de procesos se aplicarán usando el lenguaje de programación *Python*, por medio de la

librería *PM4PY* (<https://pm4py.fit.fraunhofer.de/>), en un entorno de desarrollo de programación como es *Visual Studio Code* (<https://code.visualstudio.com/>).

Tras las iteraciones necesarias, el modelo final es propuesto al personal médico, presentando también las comparaciones y diferencias de este nuevo modelo con la antigua vía clínica. Ellos serán los encargados de valorar las comparaciones obtenidas en este trabajo, evaluar la viabilidad de las propuestas de vías clínicas e incorporar cambios a sus actuales protocolos si así lo considerasen.

1.4 Alcance

Este proyecto se centra en la propuesta de vías clínicas para dos patologías en concreto: artroplastia total de rodilla y artroplastia de cadera primaria, en el Hospital Clínico Universitario Lozano Blesa. No obstante, el objetivo (O1) **Creación de un sistema de seguimiento de pacientes a partir de datos clínicos** ha sido, en la práctica, llevado a cabo no solo con registros de pacientes actuales y antiguos del Hospital Clínico Universitario, sino también con datos del *Academic Medic Hospital (AMC) de Ámsterdam* de pacientes con síntomas de Sepsis. De esta patología no se diseñarán vías clínicas pero los registros de dichos pacientes sí sirven como punto de partida para aplicar la minería de procesos. Por lo tanto, la metodología diseñada ha sido contrastada con el personal médico del Hospital Clínico y puede ser igual de válida para la obtención de vías clínicas en otros hospitales u otras patologías. Además la manera de representar estos procesos será útil y se podrá implementar en la herramienta informática que se está desarrollando dentro del proyecto global.

En los siguientes capítulos de la memoria, se abordará teóricamente el campo de estudio de la minería de procesos. Seguidamente se mostrará cómo ha sido el proceso de aplicación de técnicas y algoritmos de esta disciplina para desarrollar la metodología de seguimiento de pacientes y diseño de vías clínicas. A continuación se presentará el servidor que se está desarrollando en el proyecto global, se valorará cómo ha sido la colaboración con el hospital y se expondrán los resultados obtenidos de la minería de procesos y conclusiones. Los anexos del trabajo serán utilizados para exponer las vías clínicas actuales, presentar el marco teórico sobre el análisis clásico de datos y mostrar los resultados de manera exacta y completa.

2 Campo de estudio

El propósito de este capítulo es exponer la base teórica sobre la minería de procesos, análisis clásico de datos y modelado de procesos para comprender adecuadamente el desarrollo del trabajo, que aplica estos conocimientos de forma práctica.

2.1 Minería de procesos

La minería de procesos (*Process Mining*) [2] es una disciplina científica que surge como el punto de unión entre los modelos de procesos (*Business Process Management (BPM)*) y el registro de datos de eventos (*Work-flow Management (WfM)*). Como veremos a lo largo de esta sección, esta materia ocupa términos como la minería de datos (*Data Mining*), la inteligencia empresarial (*Business Intelligence*) o el aprendizaje automático (*Machine learning*) pero es importante no confundirlo con el ya extendido término del *Big Data*. No obstante, la minería de procesos sí que se puede aplicar sobre *Big Data*.

Por un lado, el *Big Data* hace referencia a conjuntos de datos de gran volumen que son difíciles de gestionar y analizar utilizando enfoques tradicionales. Su enfoque principal será almacenar, procesar y analizar esta gran cantidad de datos para extraer conocimiento valioso y tomar decisiones informadas.

Por otro lado, la *minería de procesos*, es una disciplina específica dentro del análisis de datos que se centra en el análisis de los registros de actividades (eventos) generados por los sistemas de información para descubrir patrones, identificar problemas y mejorar los procesos de negocio. En este caso, el enfoque principal será analizar la secuencia temporal de las actividades y obtener información sobre cómo se desarrollan los procesos.

Este campo en el análisis de datos se origina en los años '90 [2] y se comienza a usar en el ámbito de la gestión de procesos de negocios. A partir de ese punto, comienzan a surgir nuevas técnicas y enfoques para extraer conocimiento de los registros de eventos. Desde entonces, la minería de procesos ha evolucionado rápidamente, con el desarrollo de algoritmos y herramientas más sofisticadas para el análisis de estos registros y la visualización de los procesos. Ha encontrado aplicaciones en diversos sectores, como la industria, los servicios, la administración pública y la salud, como es el caso de este proyecto; y se ha convertido en una herramienta clave para mejorar la eficiencia y la calidad de los distintos procesos en estos sectores.

La minería de procesos [3], tiene por tanto como objetivo principal descubrir, monitorizar y mejorar procesos reales, extrayendo información de los registros de eventos (*event logs*) que se pueden obtener de los existentes sistemas de información.

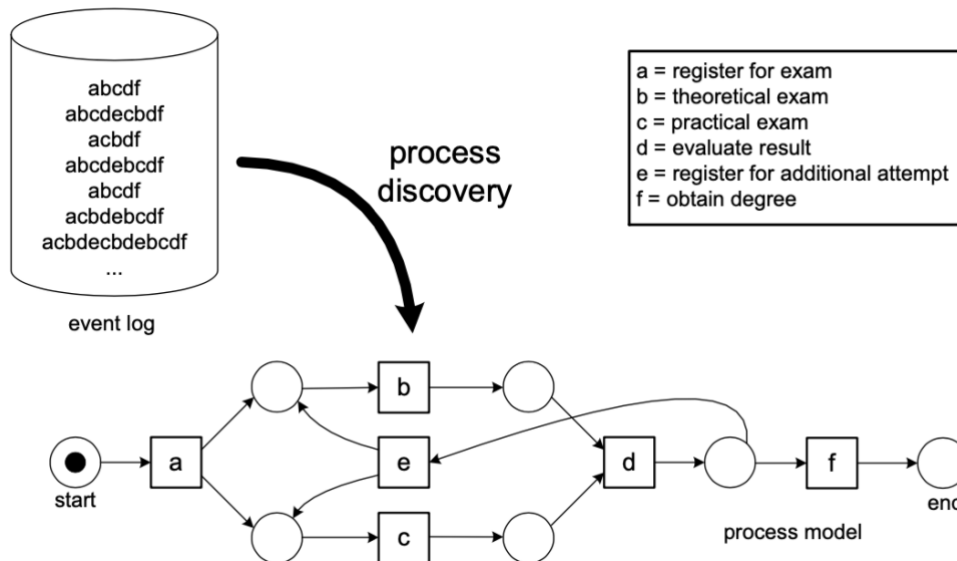


Figura 3: Minería de procesos. *Process Discovery* [3]

La minería requiere de un registro de eventos, dentro del cual, cada evento se referirá a una actividad del proceso que se vaya a modelar. Estas actividades estarán ordenadas secuencialmente y formarán lo que se denomina traza, que es el conjunto de estas actividades ordenadas en el tiempo. Así pues, el registro de eventos o *event log* estará compuesto por una serie de trazas que recogerán los posibles comportamientos del proceso, ya que consideramos que pueden ocurrir distintos casos de secuencias de actividades dentro del mismo modelo de proceso. Finalmente, la salida o *output* de este tratamiento de datos, será generalmente una Red de Petri, en el apartado 2.3 se explicarán nociones básicas sobre este modelo específico de eventos discretos.

En la *Figura 3* se ilustra este procedimiento con un ejemplo sobre el proceso de realizar un examen. En la esquina superior izquierda aparece el *event log* (registro de eventos), con letras ordenadas de distintas maneras, cada uno de los conjuntos de letras será las trazas o secuencias de inicio de actividades. En este caso, cada letra representa un evento o inicio de actividad, cuyos significados se recogen en la tabla de la esquina superior derecha. A través de la minería de procesos, se extrae del registro de datos una secuencia de pasos que modelan todo el proceso de examinación, siendo este un modelo formal, en particular una red de Petri. A través de esta red de Petri, se aprecia como el primer paso del proceso será *a*, registro para el examen; a continuación se realizarán los exámenes teóricos y prácticos, *b* y *c*, sin importar el orden entre ellos, ya que son actividades concurrentes; seguidamente se evaluará el resultado de los exámenes (*d*); si el resultado de los exámenes no es suficiente, se podrá elegir registrarse de nuevo para un intento adicional de estos exámenes (*e*) o de lo contrario, si el resultado sí es satisfactorio, ocurrirá el evento *f*, obtención del título y finalizará el proceso. De esta manera, a través de la minería de procesos, se ha obtenido un modelo del proceso de examinación (red de Petri), extrayendo información del registro de eventos que estaban almacenados en un sistema de información. Este desarrollo de minería de procesos se conoce como descubrimiento de procesos (*process discovery*); no obstante, existe un amplio abanico de técnicas de minería como pueden ser la reparación iterativa de modelos, análisis de cuellos de botella o el *conformance checking*, técnica que también se usará en este trabajo.

El *conformance checking* (verificación de la conformidad) consiste en comparar el modelo del proceso con los datos reales registrados en el *event log*. El objetivo es evaluar en qué medida los datos observados se ajustan al modelo esperado y detectar posibles desviaciones o discrepancias, para modificar y actualizar el modelo si se

considera necesario. Estas desviaciones han de ser interpretadas atentamente, ya que pueden hacer que se considere el modelo descubierto como erróneo, es decir, que no representa correctamente el proceso. No obstante, en otras ocasiones pueden demostrar que el modelo está bien construido pero lo que falla son las actividades que se llevan a cabo en la realidad, lo que significa que no se están siguiendo las indicaciones que marca el modelo del proceso.

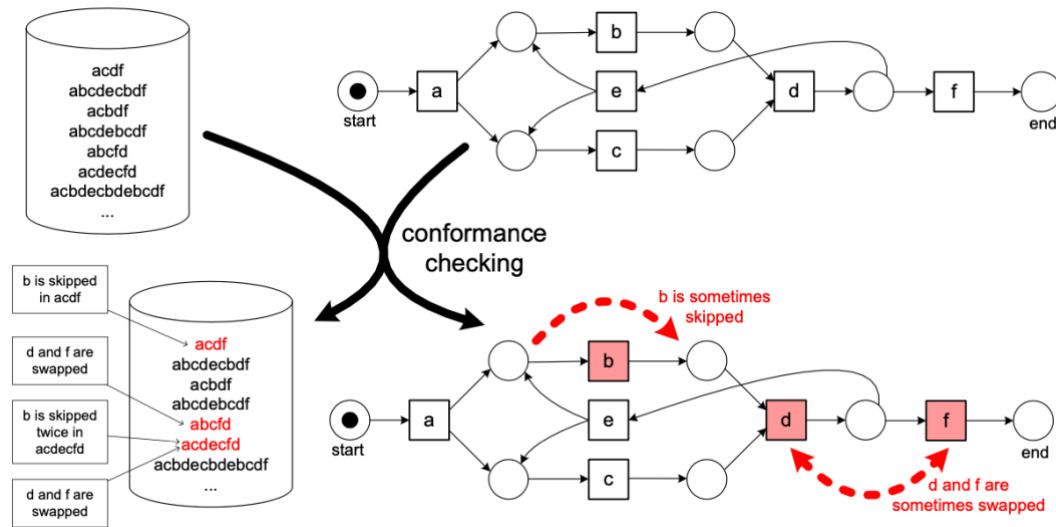


Figura 4: Mejora de modelo a partir de *conformance checking* [3]

Cabe destacar que estas técnicas de minería de procesos pueden tener problemas con registros de datos largos y complejos tanto en el tiempo de respuesta como en la interpretación de los modelos. En estos casos, se suele dividir los registros de eventos en partes más pequeñas y así hacer un estudio del proceso total de manera fragmentada.

En este trabajo se aplicarán algoritmos de minería de procesos para descubrir vías clínicas, que serán los modelos del proceso, extrayendo la información del historial clínico de cada paciente [4]. También se llevarán a cabo análisis de *conformance checking* para comprobar si los modelos descubiertos se asemejan a las vías proporcionadas por el personal médico y si estas modelan de manera adecuada los pasos que siguen los pacientes en el hospital.

2.2 Modelado de procesos con redes de Petri

Una red de Petri es un formalismo matemático que es utilizado en el ámbito de la ingeniería como herramienta para representar y modelar procesos concurrentes, tales como procesos de fabricación, sistemas de producción, logística, entre otros [15,5]. Estos procesos están compuestos de actividades o tareas que pueden interactuar entre sí, compartir recursos y competir por acceso a los mismos.

Formalmente, una red de Petri [15,5] es una 3-tupla, $N = (P, T, F)$ donde:

- P es un conjunto finito de lugares;
- T es un conjunto finito de transiciones tales que $P \cap T = \emptyset$;
- F es un conjunto de arcos tales que $F \subseteq (P \times T) \cup (T \times P)$.

Los lugares son representados gráficamente como círculos. Cada uno de ellos representa una variable de estado del sistema, siendo un estado local. Las transiciones se representan gráficamente como cuadrados e indican las acciones o eventos que pueden ocurrir en el sistema, en nuestro caso siendo inicio o finalización de unas

actividades. Su aparición cambiará el estado (local) de la red de Petri. Los arcos por su parte, conectan los lugares y transiciones, representando las relaciones de consumo y producción de marcas entre ellos.

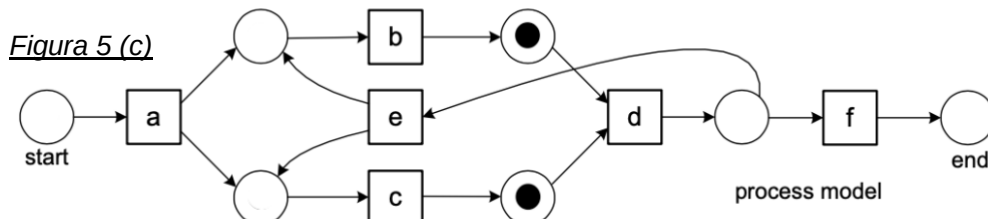
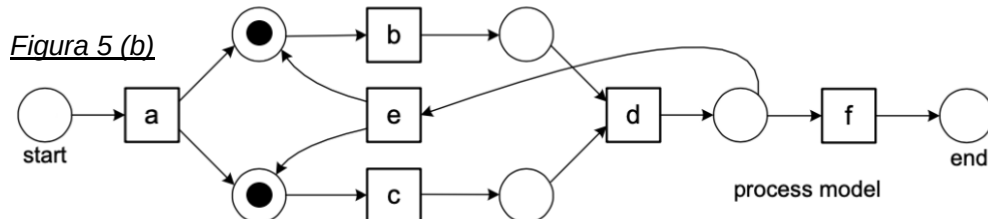
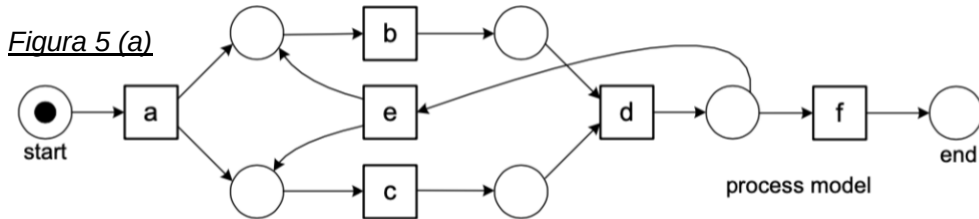
Cada lugar de la red de Petri tiene asociado un número entero no negativo, siendo el marcado del lugar. De esta manera, el marcado de la red es el conjunto de marcados de todos los lugares. Se puede expresar como un vector de dimensión igual al número de lugares.

Se define un sistema de red de Petri al par $\langle N, m \rangle$ donde:

- $N = (P, T, F)$, es una red de Petri;
- $m : P \rightarrow \mathbb{N} \cup 0$ es una función de marcado y m_0 es el marcado inicial de la red.

Las marcas de un lugar están asociadas a una o más condiciones que deben cumplirse para que la transición a la que está conectada ese lugar sea habilitada. Cuando se cumplen dichas condiciones, la transición se dispara y puede consumir/producir marcas de/en los lugares conectados a ella. El peso de los arcos indicará las relaciones de consumo o producción de marcas de los lugares y transiciones que conectan y se marcará con un número natural sobre dicho arco, siendo el peso igual a 1 si no figura ningún número sobre él. Para que una transición se dispare, el lugar o lugares de entrada a esa transición han de tener como mínimo el número de marcas que indica el arco que los conecta a dicha transición. Si eso ocurre, se dice que la transición está sensibilizada y se puede disparar; cuando lo haga, producirá tantas marcas en sus lugares de salida como indiquen los pesos de los arcos que los conectan.

A continuación, se utilizará el ejemplo de la red de Petri mostrada en la *Figura 5* para ilustrar la interpretación de este formalismo y cómo evoluciona.



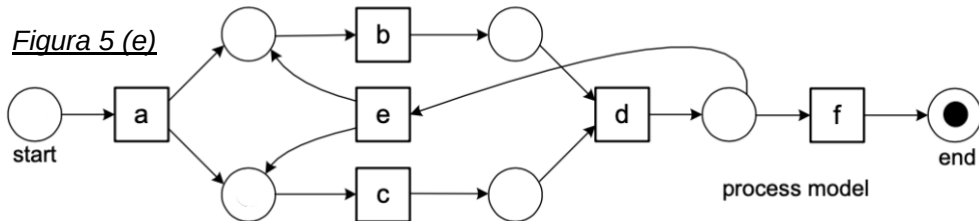
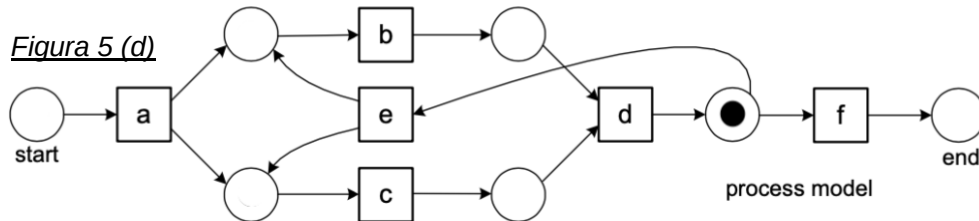


Figura 5: Evolución de estados de una red de Petri (a,b,c,d,e) [3]

En esta red de Petri, todas las transiciones consumirán tan solo una marca de cada lugar de entrada al dispararse y producirán una marca en los lugares que les precede, ya que los pesos de todos los arcos es unitario (red de Petri ordinaria). De esta manera, en el estado inicial hay una marca en el lugar *start* que sensibiliza la transición *a*. Si dicha transición se dispara, se consumirá la marca del lugar de inicio (*start*) y se producirán sendas marcas en los lugares de salida de *a*. En el nuevo marcado (o estado) que se obtiene, las transiciones *b* y *c* están sensibilizadas y se pueden disparar. El disparo de las dos transiciones lleva a la red de Petri a un nuevo marcado (*Figura 5(c)*) en el cual la transición *d* está sensibilizada. Una vez se dispara *d*, en el lugar de salida se produce una marca. La red se encuentra ahora en una situación de elección ya que las dos transiciones *e* y *f* están sensibilizadas. Según la condición que se cumpla en el sistema se disparará la transición *e* o la *f*. Si se dispara *f*, se produce una marca en el último lugar del modelo, que representa el fin del proceso.

La red de Petri es el formalismo que se va a utilizar en este trabajo para representar el seguimiento de actividades sobre los pacientes y las vías clínicas que se diseñen. Una vía clínica no deja de ser un protocolo conformado por una serie de pasos ordenados sobre los que debe llevarse cierto control para que se cumplan todos ellos. Por lo tanto, las redes de Petri son una herramienta ideal para modelar estos procesos. Además, la red de Petri que se obtiene se puede ampliar añadiendo los recursos que son necesarios en cada actividad así como duraciones de tiempo de las actividades, pudiéndose utilizar para un sistema de ayuda a las decisiones en el sistema de salud [16].

2.3 Modelado de procesos con árboles de procesos

Al igual que la red de Petri, un árbol de procesos (*process tree*) es un formalismo utilizado para representar y modelar un sistema o proceso. Su representación se basa en una estructura jerárquica que muestra las distintas ramas y subramas de decisión que se pueden seguir dentro del proceso (*Figura 6*).

Matemáticamente, se definen los árboles de procesos como:

Sea $AU\{\tau\}$ el universo de actividades donde τ es la actividad silenciosa y $\oplus = \{\rightarrow, \times, \wedge, \cup\}$ el conjunto de operadores del árbol de procesos [6].

- Sea $T \in AU\{\tau\}$ un árbol de procesos;
- Dados $n \geq 1$ árboles de proceso $T_1, T_2, \dots, T_n$ y un operador perteneciente a $\{\rightarrow, \times, \wedge, \cup\}$, $T = (T_1, T_2, \dots, T_n)$ será un árbol de procesos;
- Dados dos árboles de procesos T_1, T_2 , $T = \cup (T_1, T_2)$ será un árbol de procesos formado por 2 subárboles.

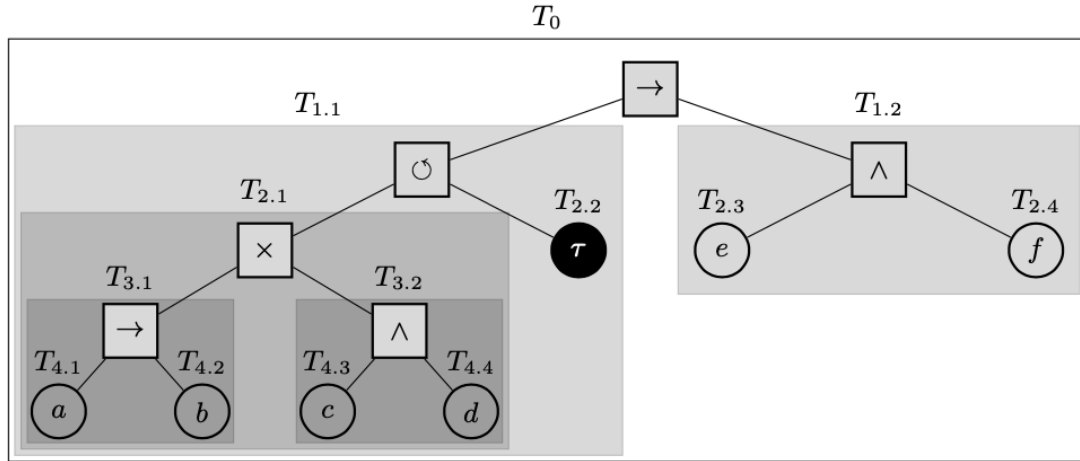


Figura 6: Árbol de procesos [6]

Este árbol estará formado por un set de nodos “hoja”; estos son los nodos inferiores de cada rama, a los que se les asocian letras. Por definición, estos nodos mantendrán asociadas a ellos actividades del proceso o la llamada actividad silenciosa, τ , que representa. Los nodos interiores y el nodo raíz (nodo superior), en cambio, contendrán los operadores del árbol. Estos operadores expresan relaciones entre subárboles y actividades:

- Secuencia ($\rightarrow$) : se ejecutarán los subárboles de izquierda a derecha.
- Elección excluyente ($\times$): solo uno de los subárboles será ejecutado.
- Paralelismo ($\wedge$): todos los subárboles serán ejecutados en paralelo.
- Bucle ($\cup$): este operador se divide en 2 subárboles, el primero de ellos se debe de ejecutar una vez (cuerpo de bucle). Si el otro subárbol se ejecuta, el primero de ellos deberá ser ejecutado. Este proceso se repetirá hasta que el segundo conjunto no se lleve a cabo.

Se procede a explicar el funcionamiento del árbol de procesos de la Figura 6. El primer operador de secuencia $\rightarrow$ marca que $T_{1.1}$ debe ejecutarse antes que el derecho $T_{1.2}$. El operador bucle $\cup$, que está formado por definición por dos subárboles, contiene un cuerpo de bucle que es el primer subárbol y una parte de repetición, el segundo subárbol. El cuerpo del bucle debe ejecutarse al menos una vez. Después, la parte de repetición puede ejecutarse opcionalmente. En caso de que se ejecute la parte rehecha, el cuerpo del bucle debe ejecutarse de nuevo a continuación. El operador de elección $\times$, es decir, un *OR-exclusivo*, indica que debe ejecutarse exactamente uno de los dos subárboles. El operador paralelo $\wedge$ indica una ejecución paralela de los subárboles. Por ejemplo, para el árbol $T_{1.2}$, las actividades e y f pueden ejecutarse en cualquier orden.

Estos árboles de procesos se utilizan como primer paso de modelado del sistema en algunos algoritmos, como es el caso del algoritmo Inductivo; para después hacer una transformación y representar el proceso en forma de red de Petri. Se profundizará sobre esta cuestión en el siguiente apartado 2.4. *Algoritmos de minería de procesos*.

2.4 Algoritmos de minería de procesos

La extracción de información a partir de los registros de eventos y la creación de los modelos de los procesos en forma de redes de Petri se lleva a cabo mediante una clase de algoritmos de descubrimiento de procesos. Los algoritmos utilizados durante este proyecto han sido el algoritmo Alpha, el algoritmo Heurístico y el algoritmo Inductivo. Se explican en esta sección la manera en la que cada uno crea los modelos de procesos y posibles diferencias entre ellos. Los ejemplos explicativos de cada algoritmo se realizarán tomando letras de la *a* hasta la *f* como si fueran eventos. Las trazas o secuencias de eventos se reflejan en la siguiente tabla, donde aparecen también las frecuencias de aparición de cada traza.

Secuencias de eventos (trazas)	Frecuencias de aparición
<a,b,c,d,e,g>	6
<a,b,c,d,e,g>	38
<a,c,d,b,f,g>	2
<a,b,d,c,e,g>	12
<a,d,c,b,f,g>	4

Tabla 1. Trazas y frecuencia de aparición para ejemplo explicativo (<https://www.futurelearn.com/info/courses/process-mining/0/steps/15639>)

2.4.1 Algoritmo Alpha

El algoritmo Alpha es uno de los primeros algoritmos de descubrimiento de procesos en poder analizar y modelar procesos con eventos concurrentes. Tiene un funcionamiento simple pero robusto y ha sentado las bases para la creación de otros algoritmos más elaborados.

Este algoritmo se encarga de leer el registro de eventos y buscar patrones concretos. Estos patrones son las relaciones y orden que hay entre los eventos de cada proceso.

- > Relación de seguir directamente (*directly follows relation*): $a > b$, estará seguida directamente por b .
- $\circ$ Relación de causalidad (*causality relation*): $a \circ b$, a está seguido por b pero nunca b estará seguida por a .
- || Relación de concurrencia (*parallel*): $a || b$, en ocasiones a está seguido de b y en otras b está seguido de a .
- # Relación de exclusión: $a \# b$ no hay relación directa entre a y b .

Una vez clasificadas las relaciones, el algoritmo crea una matriz de relaciones (*footprint matrix*) con estas notaciones (Figura 7) y a partir de ahí construirá la red de Petri correspondiente (Figura 8).

	a	b	c	d	e	f	g
a	#	→	→	→	#	#	#
b	←	#			#	→	#
c	←		#		→	#	#
d	←			#	→	→	#
e	#	#	←	←	#	#	→
f	#	←	#	←	#	#	→
g	#	#	#	#	←	←	#

Figura 7: Algoritmo Alpha. Matriz de relaciones [7]

Con la matriz de relaciones, el algoritmo analizará las relaciones entre actividades del siguiente modo. La actividad *a* no precede a ninguna otra, por eso es la transición de salida del lugar inicial; *b*, *c* y *d* preceden a la actividad *a* y a su vez tienen relación de concurrencia entre ellas, por eso se representan en paralelo a continuación de *a*; *b* y *d* preceden a *f* y *c* y *d* preceden a *e*, así que el lugar de salida de *b* se conectará a *f*, el de *c* a *e* y el de la actividad *d* a ambas, *e* y *f*; por último, *e* y *f* anteceden a *g*, por lo que ambas transiciones tendrán el mismo lugar de salida, que será el lugar de entrada de la última transición o actividad, *g*.

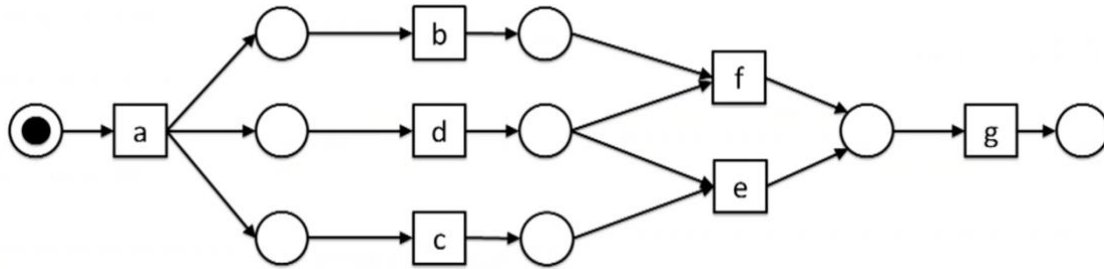


Figura 8: Algoritmo Alpha. Red de Petri [7]

El algoritmo Alpha tiene como limitaciones el ruido dentro del registro de eventos, los bucles cortos y puede que a partir de una misma matriz cree redes de Petri con estructuras diferentes.

2.4.2 Algoritmo Heurístico

El algoritmo Heurístico [8] puede considerarse una mejora del algoritmo Alpha, ya que tiene en cuenta la frecuencia con la que ocurren las actividades dentro del proceso. Además detecta y es capaz de omitir en el modelado los pequeños bucles y las actividades menos frecuentes. Con este algoritmo, se utilizarán las frecuencias de aparición de las trazas en el registro que aparecen al comienzo de la sección.

El algoritmo parte de la matriz de relaciones con la que se trabajaba en el anterior ejemplo, se crea una matriz de frecuencias de relaciones directas ($\succ$), como se aprecia en la Figura 9.a. A partir de esta matriz, se crea la matriz de dependencias (Figura 9.b) utilizando la ecuación: $a \rightarrow b = \frac{|a \succ b| - |b \succ a|}{|a \succ b| + |b \succ a| + 1}$. Con ella se establece una relación de dependencia y frecuencias entre eventos e imponiendo valores umbrales para omitir actividades o relaciones de dependencia poco frecuentes, el algoritmo construirá la red de Petri del proceso (Figura 10).

$ \Rightarrow $	a	b	c	d	e	f	g
a		.98	.67	.80			
b	-.98		.82	.67		.86	
c	-.67	-.82		.90	.92		
d	-.80	-.67	-.90		.95	.97	
e			-.92	-.95			.95
f		-.86		-.97			.98
g					-.95	-.98	

$ \> $	a	b	c	d	e	f	g
a		56	2	4			
b			44	12		6	
c		4		46	12		
d		2	4		18	38	
e							18
f							44
g							

Figura 9: Matriz de relaciones directas (a), matriz de dependencias (b) [7]

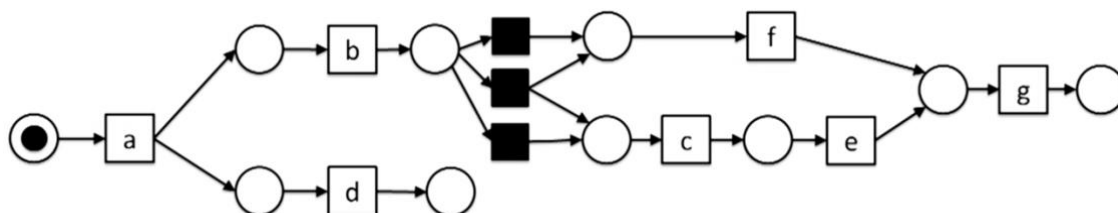


Figura 10: Algoritmo Heurístico. Red de Petri [7]

Con este algoritmo, las propiedades de repetición y generalización de la red creada no serán ideales y si el registro con el que se trabaja tiene un tamaño considerable, esta red no será simple de analizar. Adicionalmente, incluye en su estructura unas transiciones negras, que no aportan información sobre ningún evento del proceso. Dichas transiciones no están asociadas a ninguna actividad y por tanto no necesitan condición adicional para ser disparadas.

2.4.3 Algoritmo Inductivo

El algoritmo Inductivo puede considerarse de nuevo una mejora del algoritmo Alpha. En esta ocasión, se encarga de separar internamente el registro de eventos en los patrones que más se repitan. En este caso de estudio, los patrones que más se repiten son las actividades de inicio y final, que serán *a* y *g*; actividades concurrentes, como son *b*, *c* y *d*; o actividades excluyentes, como *e* y *f*. Con este análisis, el algoritmo Inductivo crea un árbol de proceso (Figura 11) y a partir de este árbol se construye una red de Petri (Figura 12), haciendo uso de transiciones que no requieren de ninguna condición para que se habiliten.

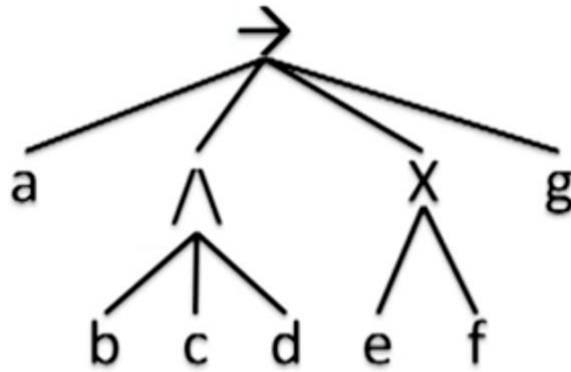


Figura 11: Algoritmo Inductivo. Árbol de proceso [7]

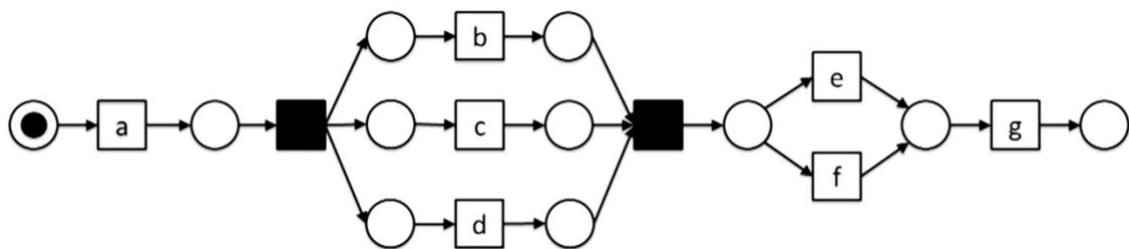


Figura 12: Algoritmo Inductivo. Red de Petri [7]

2.5 Análisis clásico de datos

El análisis de datos con modelos de regresión y gráficos estadísticos se llevará a cabo en este trabajo como otra disciplina más de la ciencia de datos. Se utilizará para estudiar la relación entre distintas variables observadas en la recogida de datos de los pacientes. El marco teórico sobre el análisis clásico de datos se puede consultar en el *Anexo [2]*

3 Minería de procesos para el diseño de vías clínicas

A lo largo de este capítulo se expondrá la aplicación de la minería de procesos a distintos casos de estudio para desarrollar una metodología de seguimiento de las actividades realizadas sobre pacientes y extracción de vías clínicas, a partir de los registros del *Academic Medic Hospital* en primera instancia y del Hospital Clínico Universitario. También se presentará el servidor web que se está desarrollando en el proyecto al que está asociado este trabajo y se realizará un análisis de riesgos del trabajo.

En los dos siguientes apartados se expondrán los trabajos de investigación realizados sobre datos del hospital holandés y antiguos datos del Hospital Clínico Universitario. En ellos, se han aplicado los algoritmos de descubrimiento de procesos ya expuestos y técnicas de *conformance checking*.

3.1 Trayectorias de pacientes que padecen de Sepsis

En esta sección se comenzará a aplicar la minería de procesos sobre datos clínicos. Se analizará una muestra de 1050 pacientes con síntomas de Sepsis, una afección potencialmente mortal. El objetivo no será diseñar vías clínicas de esta patología pero sí descubrir el modelo de proceso de la trayectoria de estos pacientes dentro del hospital, lo cual resultará útil para el futuro análisis de pacientes del HCU.

El registro de eventos utilizado en este punto procede del *Academic Medic Hospital (AMC)* [11] de Ámsterdam, Países Bajos. Consta de un conjunto de datos de personas que presentaban síntomas de sufrir Sepsis e incluían actividades en salas de emergencia y tratamientos aplicados al paciente, así como pruebas en laboratorio y la admisión y alta del paciente.

El estudio de estos datos surge del artículo *Analyzing the Trajectories of Patients with Sepsis using Process Mining* [12], el cual, en su análisis, modela el proceso con algoritmos inductivos. Dichos algoritmos obtienen una red de Petri intuitiva y más fácil de visualizar. Es por este motivo por el que el primer modelado del presente trabajo se hace con el algoritmo Inductivo.

La red de Petri resultante tras aplicar dicho algoritmo es la siguiente:

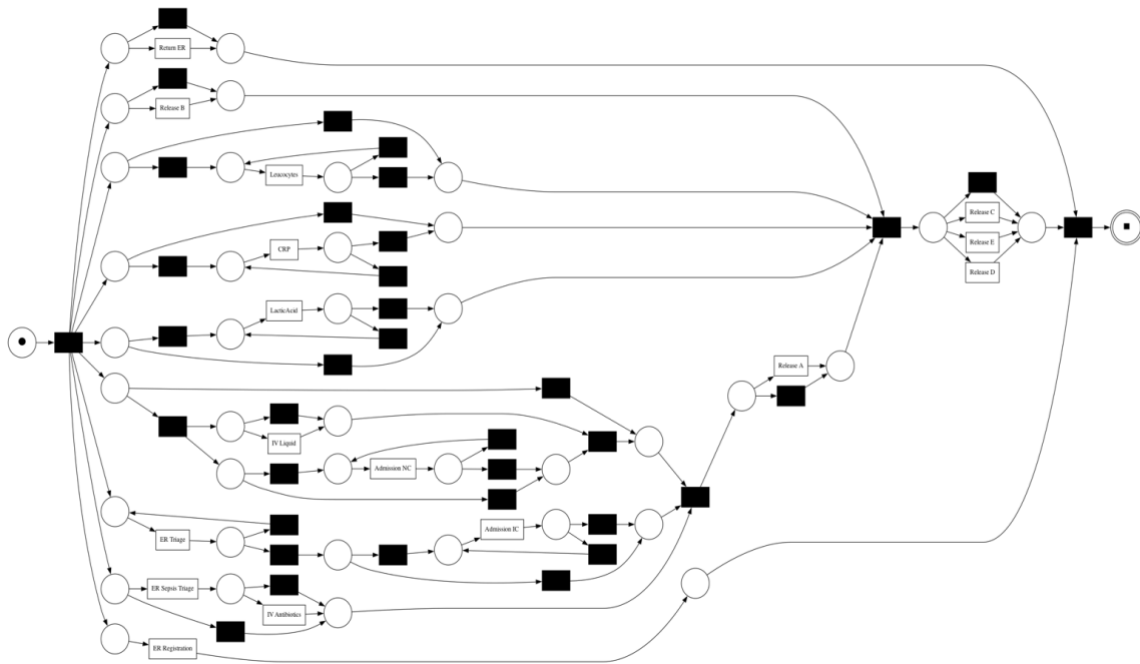


Figura 13: Red de Petri (Sepsis). Algoritmo Inductivo

Se puede comprobar que al principio de la red hay una transición que no necesita una entrada adicional para ser disparada y marca una serie de lugares que modelan los distintos conjuntos de pasos a seguir. La manera de representar estos pasos significa que son actividades concurrentes, es decir, pueden realizarse al mismo tiempo, sin importar su orden. Además, hay pasos como *Return ER* o *Release B*, entre otros, que no en todos los casos se cumplen, como denota la transición negra que les acompaña.

Con este algoritmo se obtiene un modelo que cumple con todas las trazas del registro de datos, por lo que si se realiza un estudio de *conformance checking*, la conformidad será del 100%. La manera de representar la red de Petri resulta visual a la hora de hacer un seguimiento de los pasos seguidos, incluso para aquellos con poca experiencia con este formalismo. No obstante, para trabajos posteriores se debería

depurar la representación ya que las transiciones que no representan ningún evento dificultan la comprensión.

El siguiente algoritmo aplicado ha sido el Heurístico y se ha definido un umbral de omisión de actividades y relaciones de dependencia del 50%. Como se comprueba en la *Figura 14*, la manera de modelar el proceso es totalmente distinta al anterior algoritmo y en este caso no es una red de Petri fácilmente comprensible.

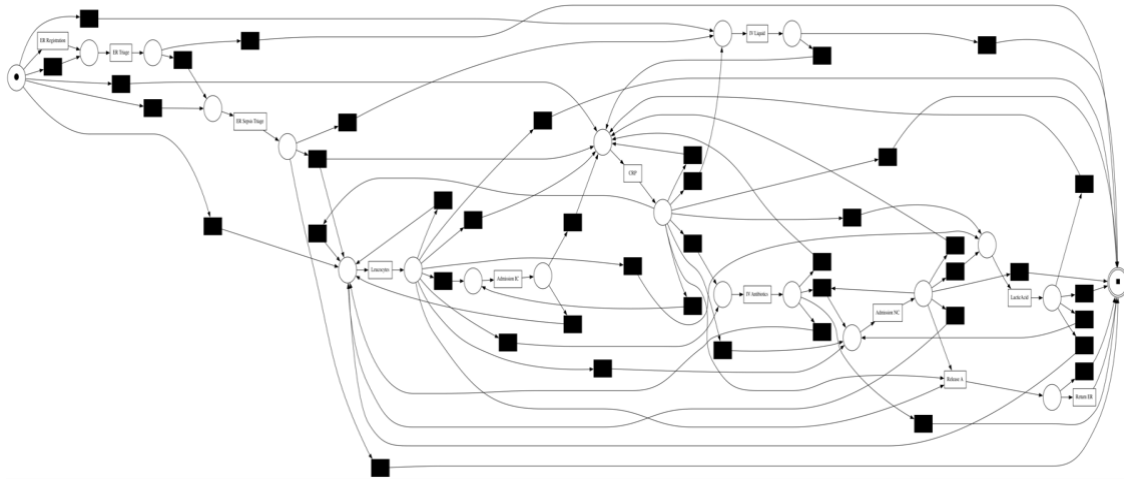


Figura 14: Red de Petri (Sepsis). Algoritmo Heurístico por defecto

Tomando como base este algoritmo, se han estudiado distintas variantes modificando los umbrales de omisión de actividades poco frecuentes y el único resultado a resaltar ha sido el obtenido estableciendo unos valores umbrales del 99% (*Figura 15*), el resto creaban una red de Petri más confusa todavía.

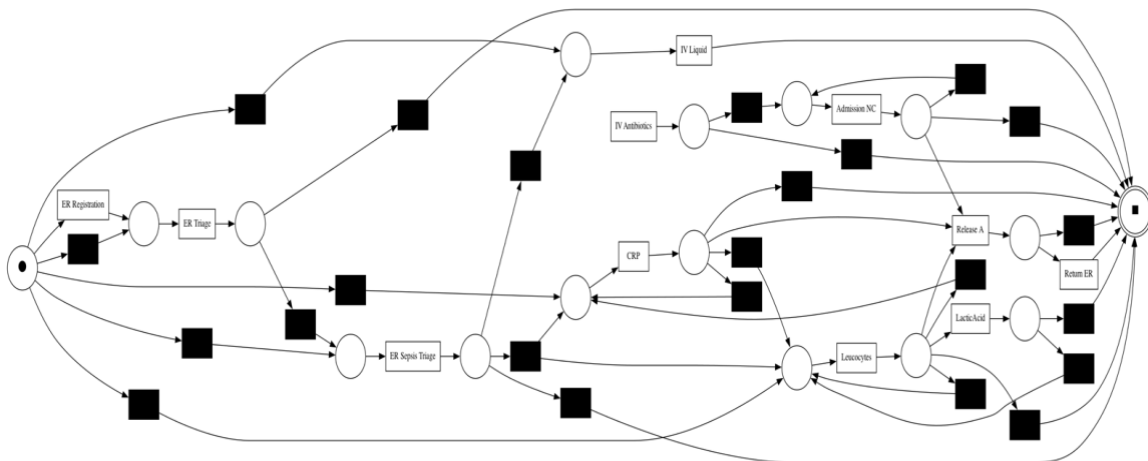


Figura 15: Red de Petri (Sepsis). Algoritmo Heurístico modificado

Este es el modelo más simplificado que se ha obtenido con el algoritmo Heurístico. De hecho al aplicar técnicas de *conformance checking* se obtiene una aptitud del modelo del 0.9368, mayor que el calculado por defecto. Esta aptitud del modelo hace referencia a lo bien que reproduce la red de Petri todo los casos del registro. Debido al extenso tamaño de este registro de datos, el algoritmo Heurístico crea un modelo que no se comprende fácilmente.

El tercer algoritmo aplicado ha sido el Alpha, cuyos resultados no han sido los esperados:

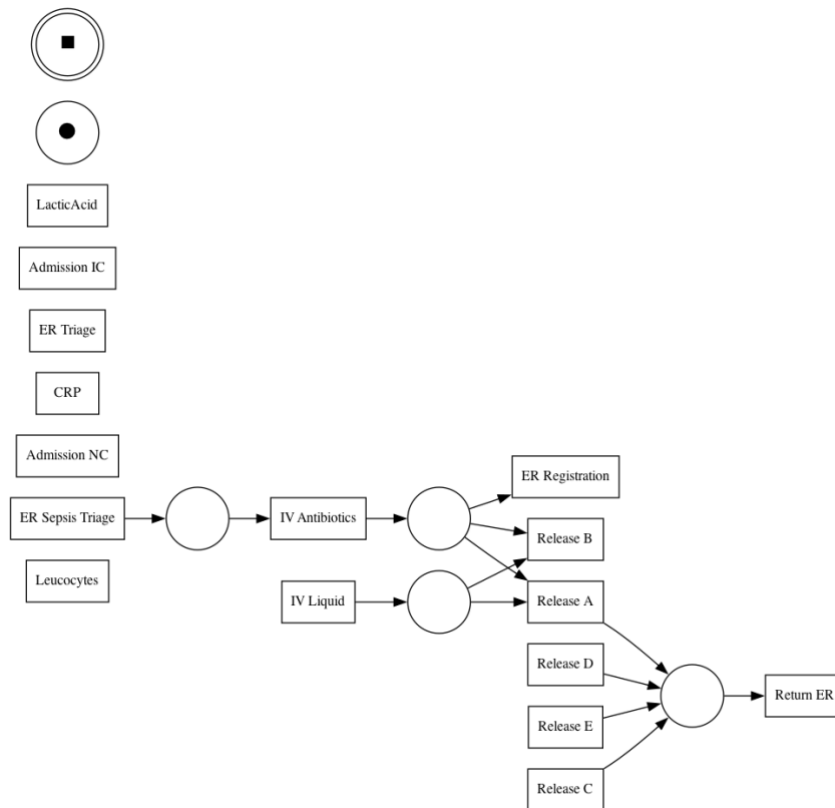


Figura 16: Red de Petri (Sepsis). Algoritmo Alpha

El algoritmo Alpha genera una red de Petri con un lenguaje mayor que el del mismo registro de datos, mientras que los otros algoritmos intentan reducirlo para aproximarse solo al propio registro. De esta manera, el algoritmo Alpha será capaz de reproducir, a priori, más casos de los que presenta el registro, si no son muy diferentes a los anteriores.

La manera que tiene el algoritmo Alpha de modelar el proceso es a través de transiciones ligadas a eventos (las situadas en la parte izquierda de la *Figura 16*) que se pueden disparar en cualquier momento. De esta forma, las actividades que aparezcan varias veces a lo largo del proceso, podrán ser disparadas sin necesidad de que estén precedidas por alguna actividad en concreto. Esta representación no es la que se busca para diseñar las vías clínicas ya que no puede hacerse un seguimiento de las actividades que se realizan. Por eso, los siguientes estudios se centrarán en análisis con los algoritmos Inductivo y Heurístico.

3.2 Estudio de pacientes con prótesis total de cadera

Los registros de pacientes utilizados en este apartado son antiguos datos del año 2019 y proceden del servicio de *Cirugía Ortopédica y Traumatología* del Hospital Clínico Universitario de Zaragoza. Recogen pasos del proceso de las intervenciones quirúrgicas para pacientes que necesitan prótesis total de cadera (PTC) y a partir de ellos, se va a intentar extraer un modelo que recoja y ordene la serie de actividades llevadas a cabo sobre el paciente.

Previo a la aplicación de algoritmos, se ha analizado el registro de los pacientes y es necesario advertir que estas actividades están registradas tan solo el día en el que se llevaron a cabo, no la hora exacta. Por lo tanto, el modelo que se obtenga ordenará las actividades por días. Este hecho no es un problema, ya que las vías clínicas actuales clasifican las actividades a realizar también por días.

En el apartado anterior, el modelo del proceso más intuitivo y claro se obtuvo con el algoritmo Inductivo, por eso se ha aplicado en primer lugar este algoritmo al registro de datos. Tras extraer un modelo preliminar y depurarlo durante varias iteraciones para así evitar bucles y transiciones que no aportan información, esta ha sido la red de Petri obtenida:

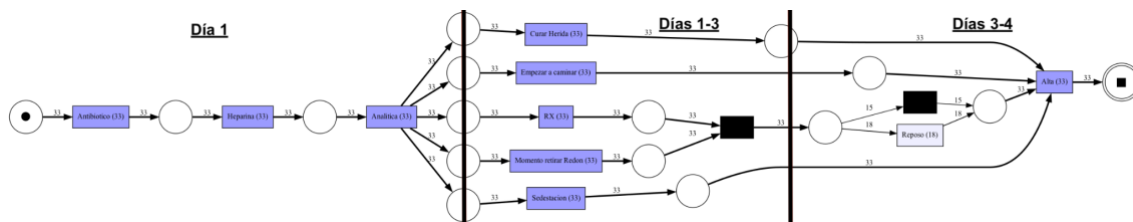


Figura 17: Red de Petri (datos PTC). Algoritmo Inductivo

A diferencia de casos anteriores, este modelo está programado para recoger las frecuencias con las que ocurren las distintas actividades asociadas a las transiciones de esta red, a modo de información adicional para el personal médico. Además, se han separado los días en los que se realiza cada actividad.

Este algoritmo proporciona un modelo que tiene en cuenta todos los casos y por eso muestra las actividades de los días 1-3 como concurrentes, que no siempre se realizan en el mismo orden. Para obtener un modelo más generalista, evitando los casos poco frecuentes, se ha utilizado el algoritmo Heurístico.

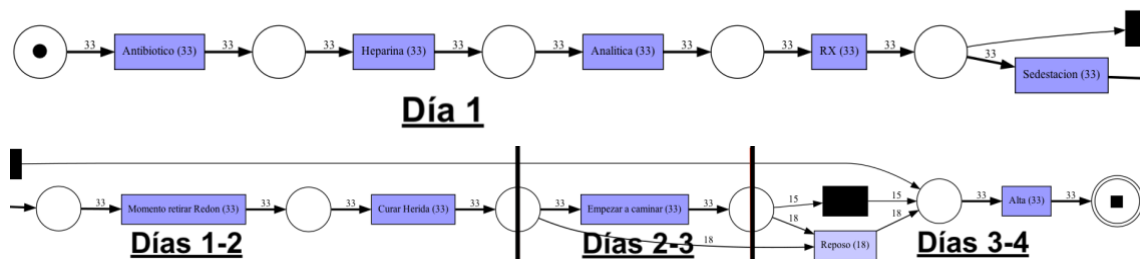


Figura 18: Red de Petri (datos PTC). Algoritmo Heurístico

El tamaño del registro no es tan grande como en el anterior apartado de pacientes con Sepsis, por lo que el algoritmo Heurístico no tiene problemas para procesar los datos y muestra una red de Petri sencilla, con las actividades ordenadas secuencialmente.

Tras el análisis y minería de procesos con estos datos, se ha definido la metodología para realizar el seguimiento de los pacientes a partir de datos existentes. Dicha metodología se recoge en los siguientes pasos:

1. Aplicar algoritmo Inductivo.
2. Análisis del modelo obtenido y detección de posibles mejoras.

3. Depurar el registro de datos para eliminar bucles y transiciones innecesarias.
4. Reiterar el proceso hasta elaborar una red de Petri adecuada.
5. Inclusión de frecuencia de actividades y división por días.
6. Posible caso general aplicando el algoritmo heurístico.

3.3 Diseño de vías clínicas

Una vez se ha diseñado el procedimiento y un sistema de seguimiento de pacientes, se han solicitado registros de pacientes al Hospital Clínico Universitario para aplicar esta metodología. Los registros provienen del Servicio de Cirugía Ortopédica y Traumatología y corresponden a 50 pacientes que han sido intervenidos para la implantación de prótesis de cadera (PTC) y otros 49 pacientes con prótesis de rodilla (PTR).

Como se pudo comprobar en las secciones anteriores de este capítulo, tanto el algoritmo Heurístico como el algoritmo Alpha, no modelan con tanta precisión como el algoritmo Inductivo procesos con elevado número de registro de eventos. Durante el estudio de estos registros se ha comprobado este hecho y por eso en esta sección, los resultados más significativos han sido obtenidos aplicando el algoritmo Inductivo.

Se comenzará estudiando los casos de prótesis de cadera aplicando la extracción del modelo del proceso por técnicas de minería de procesos. Este es el seguimiento de pacientes obtenido (*Figura 19*). Para una mejor visualización, se puede consultar la *Figura 19* a mayor tamaño en el *Anexo 4*.



Figura 19: Seguimiento de pacientes con prótesis total de cadera(PTC)

Esta red de Petri ha sido obtenida con el algoritmo Inductivo. Las actividades han sido clasificadas por días del proceso que sigue el paciente. Analizando esta red detenidamente, se puede comprobar que la mayoría de actividades han sido realizadas en 50 ocasiones, es decir, para todos los pacientes. Las actividades que no han sido realizadas tal número de veces son "Manchado de apósito", "Rescate analgésico 1 DPO", "Dolor segundo día" y "Valoración deambulacion tercer día". Las razones por las que estos eventos tienen una menor frecuencia de aparición pueden ser variadas, puede ser que la recogida de datos no haya sido precisa o que simplemente son eventos poco frecuentes pero importantes de registrar. En cualquier caso, el personal médico es el encargado de valorar estos aspectos y sacar conclusiones de los resultados aquí obtenidos.

Los siguientes pacientes estudiados han sido los casos de prótesis de rodilla, aplicando las mismas técnicas de descubrimiento del proceso. El seguimiento de pacientes con esta patología es el siguiente (*Figura 20*). Para una mejor visualización, se puede consultar la *Figura 20* a mayor tamaño en el *Anexo 4*.



Figura 20: Seguimiento de pacientes con prótesis total de rodilla (PTR)

Al igual que en el estudio del proceso de prótesis de cadera, la mayoría de actividades se efectúan tantas veces como pacientes hay en el registro, en esta ocasión, 49 veces. Los eventos que ocurren con menor frecuencia son "Solicitud de pruebas cruzadas", "Rescate analgésico", "Dolor segundo día" y "Valoración deambulacion tercer día".

Previo a la recogida de datos y la minería de procesos, se pidió al personal médico de este departamento que redactara una propuesta de vía clínica para ambos procesos de intervención quirúrgica. De esta manera se podrían comparar los resultados obtenidos en este estudio con lo que los sanitarios consideran que ocurre o debería de ocurrir en el hospital. Al cotejar el seguimiento de pacientes obtenido de la minería de procesos con las vías proporcionadas desde el hospital, se consiguen resultados satisfactorios. Estos son que la secuencia de actividades descrita por ambas vías coincide con el seguimiento de pacientes, a excepción de los casos poco frecuentes y de 6 actividades en cada patología. Este hecho se ha comunicado a los sanitarios y el motivo de ello es que son actividades que no se suelen registrar en el momento en el que se realizan y por ello, no pueden añadirse como evento dentro del modelado del proceso.

El hecho de que haya actividades que no se registren dentro del sistema de información del hospital es un aspecto que se debate en el capítulo 6. *Problemas encontrados*, pero es importante resaltar que limita la capacidad de modelado del proceso.

Las vías clínicas proporcionadas por los sanitarios, son idénticas, a excepción de 5 actividades adicionales para prótesis de rodilla. Esto significa que salvo la intervención, los pasos que tiene que seguir el paciente desde su preingreso son prácticamente iguales si sufre una de estas dos patologías, exceptuando los 5 pasos adicionales. Este suceso puede llevar a pensar que las vías clínicas obtenidas en este trabajo, debieran ser prácticamente iguales.

De hecho, si se comparan las dos redes de Petri obtenidas como modelo de ambos procesos, guardan cierta similitud, aunque se haya obtenido un modelo más simple en el caso de la prótesis de rodilla. Las redes no son completamente iguales porque el registro de pacientes de cada patología es distinto y es importante destacar que la variabilidad de los datos clínicos es significativa, aún en conjuntos de extensión no muy elevada. Además, cabe recordar que la vía clínica no es más que una serie de recomendaciones y protocolos que se deberían de seguir según una patología específica, es decir, el personal médico es quien toma las decisiones sobre el paciente, ayudándose de las vías clínicas. Cada paciente presenta una casuística y características únicas, por eso estos procesos pueden discernir en ciertas ocasiones.

3.4 Despliegue de metodologías en servidor web

El proyecto al que está asociado este trabajo de fin de grado está desarrollando un servidor. Este servidor se trata de una herramienta de gestión de la información basado en la arquitectura de *API Rest* y está compuesto por una estructura de tres capas. Las tres capas son:

- ***Backend***: capa encargada de realizar todos los cálculos necesarios para el funcionamiento del servidor y su visualización. Está desarrollado en una tecnología *Python* con un marco de trabajo (*framework*) que se llama *Flask* (<https://palletsprojects.com/p/flask/>), muy útil para el desarrollo de aplicaciones web. Se ha decidido este *framework* debido a su simplicidad y la fácil integración con el resto de capas.
- ***Frontend***: capa que presenta la interfaz del usuario, es decir, se visualiza la aplicación y el usuario interactúa con el servidor web. El *frontend* se comunica con el *backend* para solicitar los datos necesarios para mostrarlos en la interfaz de usuario y para enviar datos ingresados por el usuario de vuelta al *backend* para su procesamiento. Está desarrollado con un *framework* llamado *ReactJS* (<https://react.dev>). Se ha optado por esta solución ya que permite la

modularización de componentes visuales y el rápido tiempo de respuesta en cliente.

- **Base de datos:** Se utiliza una base de datos relacional para almacenar los datos, gestionarlos y actualizarlos. El *backend* se comunica a través de un *ORM* de *Python* con la base de datos para almacenar, recuperar datos y operar con ellos. Una base de datos relacional es un tipo de base de datos que organiza y estructura los datos en tablas y los interrelaciona a través de claves o números identificativos. Se ha decidido el uso de *PostgreSQL* debido a su licenciamiento gratuito y la sencillez de la misma.

La estructura e interfaz del servidor se encuentra en el *Anexo [3] Estructura del servidor web*

La labor de este trabajo en dicho proyecto es la de obtener los algoritmos de descubrimiento de procesos que mejor modelen las actividades que se realizan dentro del tratamiento de un paciente. El algoritmo que se ha implementado en el servidor es el Inductivo, por la simplicidad de las redes de Petri que obtiene y porque es el algoritmo que mejor se comporta en el caso de que se trate con registros de pacientes con elevado número de actividades, tal y como se aprecia en la *Figura 25* del apartado 6. *Resultados*.

4 Tratamiento de datos y colaboración con el hospital

Durante este capítulo se valorará cómo ha sido el trabajo con el personal médico, la recogida de los datos de pacientes, su procesado posterior y la explicación de los resultados obtenidos a los sanitarios.

4.1 Recogida de datos

Cabe destacar antes de la explicación sobre cómo ha sido el proceso de recogida de datos, que la recopilación de los mismos se ha llevado a cabo respetando la privacidad de los pacientes. **Todos los datos que se han utilizado en este proyecto han sido anonimizados, dado que el único identificador que consta de cada paciente es el Número de Historial Clínico (NHC).**

El proceso de recogida de todas las actividades realizadas a los pacientes ha sido efectuado por los Médicos Internos Residentes (MIR) del Hospital Clínico Universitario. El procedimiento ha sido manualmente, paciente por paciente. Es decir, se ha revisado el *Historial Clínico* de cada paciente y se han registrado las actividades una tras otra sobre una hoja de cálculo (*Excel*). Esta es una práctica que lleva de 10 a 15 minutos por paciente, ya que no todas las actividades se registran en el *Historial Clínico Electrónico*, sino en papel. Si todo el seguimiento estuviese informatizado, la recopilación de información habría sido más rápida y menos costosa para los sanitarios.

4.2 Procesado de datos

La minería de procesos y el análisis de datos mediante algoritmos requiere que los registros de datos presenten un formato específico (*Dataframe*). El *Dataframe* es una estructura de datos en forma de tabla en la que cada columna representa una variable y cada fila contiene los valores de esas variables. Es utilizado para almacenar los datos de manera estructurada y acceder a ellos fácilmente.

En este caso el formato necesario es que cada actividad que sigue el paciente esté identificada con el *Número de Historial Clínico (NHC)* del propio paciente; que los

formatos de la fecha en que se realizan las actividades deben de ser iguales; o que los valores numéricos utilizen comas en vez de puntos como signo de separación de la parte decimal. Estas especificaciones no siempre se han cumplido a la hora de plasmar los registros de pacientes sobre una hoja de cálculo. Es por eso que previo a la aplicación de algoritmos, se ha realizado un filtrado y procesamiento de todos estos datos, utilizando la librería *Pandas*, una librería del lenguaje de programación *Python* que facilita la manipulación y el análisis de datos.

En la siguiente figura se muestra cómo es el registro de un paciente operado de prótesis de rodilla, con un *Dataframe* apto para trabajar con él.

NHC	Patología	Cirujano	Tipo de Prótesis	Actividad	Timestamp	Hora - Formato (Ejemplo: 12:30)	Recursos - Formato (Ejemplo: AntibioticoX,Cama,Habitacion)
410784	Gonartrosis derecha	Cesar Garcia Gutierrez	Persona	Rx preiq	2021-11-08 00:00:00		Sala de rayos, personal, material informatico
410784				Consentimiento informado	2021-08-23 00:00:00	11:17:00	Personal, material
410784				Valoracion preanestesia	2023-05-04 00:00:00		Personal, material
410784				Comprobacion anestesia ingreso	2023-05-07 00:00:00		Personal
410784				Comprobacion Consentimiento	2023-05-07 00:00:00		Personal
410784				Pauta de medicacion y Antibiotico induccion	2023-05-07 00:00:00		Personal, medicacion
410784				Pauta HBPM preiq	2023-05-07 00:00:00	20:00:00	Personal
410784				Solicitud pruebas cruzadas	2023-05-07 00:00:00		Personal, material para extraccion
410784				IQ	2023-05-08 00:00:00	08:30:00	Personal, material protesico y quirofano, quirofano
410784				Comprobacion ayunas IQ	2023-05-08 00:00:00		personal
410784				AB IQ	2023-05-08 00:00:00		Personal, medicacion
410784				Protocolo IQ	2023-05-08 00:00:00		Personal
410784				Solicitud AS	2023-05-08 00:00:00		Personal
410784				Solicitud Rx	2023-05-08 00:00:00		Personal
410784				Pauta AB postiq	2023-05-08 00:00:00		Personal
410784				Rehabilitación	2023-05-09 00:00:00	10:45:00	Personal
410784				Retirada redon primer dia	2023-05-09 00:00:00	12:22:00	Personal, material
410784				Sedestacion primer dia	2023-05-09 00:00:00		Personal
410784				Revision AS primer dia	2023-05-09 00:00:00		Personal
410784				Rescate analgesico	2023-05-09 00:00:00		Personal, Medicacion
410784				Manchado aposito	2023-05-09 00:00:00		Personal, material
410784				Tratamiento AB	2023-05-09 00:00:00		Personal, material, medicacion
410784				HBPM post iq	2023-05-09 00:00:00		Personal, Material, medicacion
410784				Rx control	2023-05-09 00:00:00		Sala de rayos, personal, material informatico
410784				Bipedestacion y deambulacion segundo dia	2023-05-10 00:00:00		Personal,
410784				Valoracion deambulacion tercer dia	2023-05-11 00:00:00		Personal
410784				Entrega de ALTA	2023-05-11 00:00:00		Personal, material

Figura 21: Registro de paciente con *Dataframe* adecuado

Una vez que los datos presentan el formato correcto, se lleva a cabo la minería de procesos. Todo este proceso de minería se ha desempeñado usando el lenguaje de programación *Python*, por medio de la librería *PM4PY*, en un entorno de desarrollo de programación como es *Visual Studio Code*. Esta librería está especializada en el análisis de procesos. Proporciona un conjunto de herramientas y algoritmos para realizar actividades de descubrimiento de procesos, *conformance checking* y mejora de procesos utilizando técnicas de minería de procesos.

Adicionalmente, para el análisis de datos también se han utilizado las librerías *Numpy* (<https://numpy.org/>) y *Seaborn* (<https://seaborn.pydata.org/#>). La primera de ellas está especializada en el procesamiento numérico, muy útil en concreto para este proyecto, para el tratamiento de vectores (*arrays*); y la segunda, proporciona una interfaz de alto nivel para crear gráficos estadísticos atractivos e informativos.

Por otro lado, todo el código de programación ha sido almacenado en la plataforma en línea de *GitHub* (<https://github.com/>). *GitHub* se utiliza para gestionar repositorios de código fuente como los usados para estos procesos de minería y análisis de datos. Proporciona además un entorno colaborativo y un sistema de control de versiones.

En la *Figura 22* se explica el flujo de datos que ocurre cuando se trabaja con *GitHub*. Del repositorio en local con el que se programa, el código se almacenará en el repositorio remoto de la plataforma online y allí se gestionará el sistema de almacenado

y versiones. Desde ese repositorio se podrán volver a descargar los archivos para seguir trabajando en local.

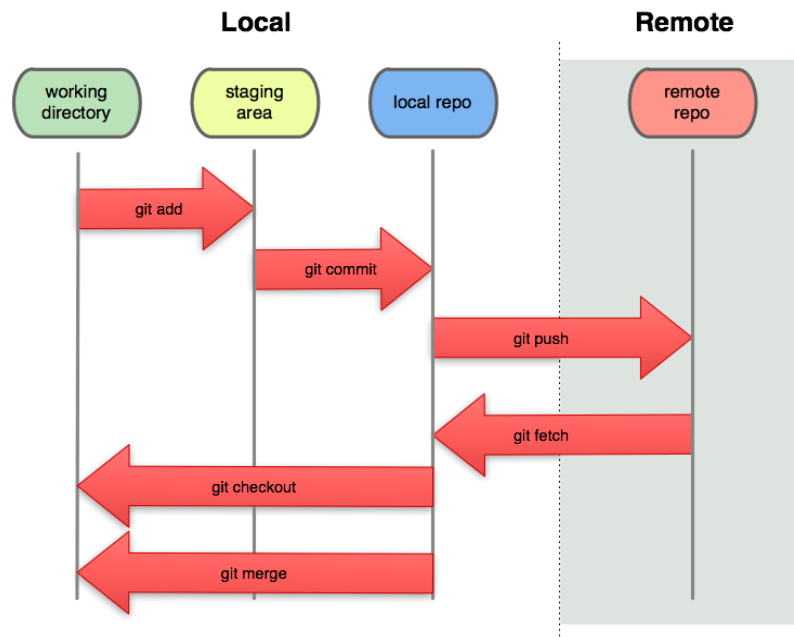


Figura 22: Flujo de datos en *GitHub*
(<https://bluuweb.github.io/tutorial-github/01-fundamentos/#¿que-es-git>)

Se ha creado un repositorio público de esta plataforma en el que se encuentran todos los archivos de minería de procesos y análisis clásico de datos utilizados a lo largo del proyecto:

https://github.com/danielsanzz/Mineria_Procesos_TFG

4.3 Colaboración con el personal sanitario

A lo largo del desarrollo del proyecto se han efectuado diversas reuniones con el personal médico del *Servicio de Cirugía Ortopédica y Traumatología del Hospital Clínico Universitario Lozano Blesa* para sentar las bases de este trabajo final. Los asuntos abordados durante dichas reuniones fueron:

- Presentación del problema sobre las vías clínicas.
- Propuesta de soluciones y objetivos del TFG.
- Concreción de las patologías a estudiar.
- Explicación del registro de actividades a recoger.
- Explicación y validación de resultados.

Una vez obtenidos los resultados de la minería de procesos, se han presentado al personal médico las comparaciones entre el seguimiento de pacientes intervenidos en cada patología y las vías clínicas proporcionadas sobre cada proceso. Se ha explicado la dinámica que tienen las redes de Petri para modelar procesos y se han ido analizando los días de la estancia de un paciente en el hospital, las actividades que marca la vía clínica que se deberían de realizar y las actividades que de verdad se realizan o registran para el total de casos estudiados.

Estos análisis comparando cada actividad del proceso han resultado interesantes de analizar para el equipo de traumatología por los siguientes motivos:

- Desconocimiento de que hubiese actividades que no se registrasen en el historial médico electrónico. Esto puede dar lugar a fallos como que se realice una actividad más veces de las debidas (p.ej. Pauta de AB) y en definitiva acarree problemas para el paciente.
- Los datos de frecuencias de disparo de las transiciones reflejan las veces que se lleva a cabo un evento. Esta información es útil en pasos que están valorando en eliminar de la vía clínica (p.ej. Solicitud de pruebas cruzadas o Rescate analgésico).
- Cambio de orden cronológico de algunas actividades (p.ej. Checklist o Ácido Tranexámico).
- Eliminación de actividades dentro de la vía clínica (p.ej. Bloqueo nervioso anestésico).

5 Análisis de riesgos del proyecto

El siguiente análisis tiene como objetivo identificar y evaluar los riesgos potenciales asociados con este trabajo de fin de grado. La parte técnica del trabajo se centra en la utilización de metodologías diseñadas con minería de procesos para realizar un seguimiento de pacientes del Hospital Clínico Universitario en forma redes de Petri a partir de registros de sus datos clínicos. A través de este análisis, se busca anticipar los posibles desafíos y contratiempos que podrían surgir a lo largo del desarrollo del proyecto. Al identificar estos riesgos, se podrá diseñar estrategias efectivas de gestión de riesgos que permitan minimizar su impacto y garantizar el éxito del proyecto.

1. **Privacidad y seguridad de los datos.** Uno de los riesgos más importantes en proyectos que involucran datos médicos es la privacidad y seguridad de la información. Existe el riesgo de que los datos de los pacientes sean accedidos o divulgados de manera no autorizada. Para mitigar este riesgo, se han anonimizado los datos de los pacientes y se ha trabajado con el Número de Historial Clínico (NHC) como único identificador del paciente, para asegurar la confidencialidad de los registros.
2. **Calidad y confiabilidad de los datos.** La calidad de los datos de entrada puede afectar la precisión y validez de los resultados obtenidos. Existe el riesgo de que los registros del hospital contengan errores, inconsistencias o datos faltantes. Como medidas preventivas ante este riesgo, se ha realizado una limpieza y validación exhaustiva de los datos proporcionados por el hospital antes de aplicar la minería de procesos.
3. **Selección y diseño de algoritmos.** La elección y diseño de los algoritmos de minería de procesos pueden influir en la calidad y relevancia de las redes de Petri extraídas. Existe el riesgo de utilizar algoritmos inadecuados o aplicarlos incorrectamente, lo que podría llevar a conclusiones erróneas. Para gestionar de manera efectiva este riesgo, se han estudiado los algoritmos disponibles para distintos casos de estudio y se ha evaluado su idoneidad para el contexto médico, validando los resultados obtenidos con expertos médicos.
4. **Interpretación y análisis de los resultados.** Una vez obtenidas las redes de Petri, es necesario interpretar y analizar adecuadamente los resultados para extraer conclusiones significativas. Existe el riesgo de malinterpretar los patrones identificados o sacar conclusiones erróneas sobre los procesos médicos. Con el fin de minimizar este riesgo, se debe contar con expertos en el área médica que puedan validar y contextualizar los resultados obtenidos.

5. **Precisión del modelado y cobertura de casos.** Existe el riesgo de que el modelado por Redes de Petri de los procesos no capture con precisión todas las actividades y situaciones relacionadas con los pacientes. La complejidad y variabilidad de los casos clínicos pueden dificultar la representación completa de todos los escenarios posibles. Además, puede haber actividades atípicas o excepcionales que no se ajusten a los patrones típicos capturados por la minería de procesos. Esta limitación puede afectar la precisión de los resultados obtenidos y llevar a conclusiones incompletas.
6. **Integración con el entorno hospitalario.** El trabajo involucra la extracción de datos del registro del hospital, lo cual requiere integración con los sistemas existentes. Existe el riesgo de enfrentar desafíos técnicos y de compatibilidad al integrar los algoritmos de minería de procesos con los sistemas de información del hospital. Es por eso por lo que dichos algoritmos se están implementando en el servidor de gestión de la información hospitalaria durante su fase de desarrollo.

6 Problemas encontrados

Durante el transcurso de este trabajo, han surgido problemas relacionados con la coordinación y gestión de la información. A continuación, se enumeran los más relevantes:

- **Actividades no registradas.** No todas las actividades que se realizan sobre un paciente se registran de manera electrónica. Este problema limita el alcance de la minería de datos, puesto que solo se pueden modelar los procesos con las actividades que están registradas. El análisis médico también se verá afectado, puesto que a la hora de comparar la vía clínica actual con el seguimiento de los pacientes, no se podrá comprobar si las actividades que no aparecen lo hacen por el motivo de que no se llevan a cabo o simplemente porque no están registradas.
- **Retrasos en la obtención de datos de pacientes.** Para recolectar la información de los historiales clínicos de cada paciente se valoró en primera instancia volcar los datos registrados en el sistema informático sobre una hoja de cálculo. No obstante, esta opción se descartó ya que la información obtenida hubiera sido incompleta, por lo que se optó por recopilar manualmente, de registros informáticos y otros apuntados en papel. Adicionalmente, esta extracción de datos no es un proceso inmediato. Por eso, el periodo de espera para recibir estos registros se demoró más de lo deseado y la minería de procesos se detuvo durante alguna semana.
- **Escasez de información sobre los pacientes.** El análisis clásico de datos llevado a cabo no se centra en las actividades que sigue el paciente, sino en parámetros que identifican al paciente (edad, sexo, lugar de procedencia) y otras variables de interés médico (hemoglobina antes y después de la operación, patologías previas y otros). Durante la recogida de información, todos estos parámetros no fueron recogidos por lo que el análisis ha tenido que realizarse con un registro antiguo de pacientes, del año 2019. Si se hubiese efectuado con los registros actuales, hubiera sido un análisis más completo, puesto que el tamaño muestral de los datos actuales es mayor al de los antiguos.
- **Datos registrados de forma no deseada.** Para la minería de procesos y análisis de datos, es necesario que los registros tengan un formato determinado. Los datos proporcionados no siempre han tenido el formato deseado y es por eso que se ha tenido que llevar a cabo un procesamiento de estos registros previo a la minería de procesos y análisis de datos. El formato no deseado de

los registros puede ser debido a que han sido varios los sanitarios que han recogido los datos y no se haya llegado a un consenso previo de cómo extraer y plasmar la información en las hojas de cálculo.

- **Resistencia al cambio.** Dejar atrás el antiguo sistema para empezar a usar el nuevo, implica un cambio significativo, el cual podría haberse comenzado a dar ya. Desafortunadamente, no es este el caso, ya que a la hora de instalar en ordenadores del hospital el servidor que se está desarrollando se encuentran dificultades como ordenadores obsoletos o sin conexión a internet. Tanto este trabajo como el proyecto general han despertado el interés del equipo de médicos que han colaborado pero para implantar de manera efectiva todas las propuestas, se necesitaría el apoyo y la intención de áreas más grandes del hospital.

7 Resultados

7.1 Algoritmos óptimos de minería de procesos

En los apartados 3.1 *Trayectorias de pacientes que padecen de Sepsis* y 3.2. *Estudio de pacientes con prótesis total de cadera* se han estudiado distintos algoritmos de descubrimiento de procesos para modelar la trayectoria de los pacientes dentro del hospital.

Tras analizar los modelos que se obtienen de cada caso de estudio con estos algoritmos, se ha seleccionado el **algoritmo Inductivo** como el que mejor modela el seguimiento de pacientes y vías clínicas. La red de Petri que se obtiene al aplicarlo es una red sencilla y fácil de seguir y entender para aquellos que no estén familiarizados con este tipo de notación, como puede ser el personal sanitario.



Figura 23: Registro de pacientes (PTR). Algoritmo Inductivo

Un ejemplo de representación de la trayectoria del paciente con este algoritmo es la *Figura 23*, donde están representados los pasos que se han registrado sobre los pacientes operados de prótesis de rodilla durante el día previo a la intervención. El seguimiento de los pacientes completo para los casos de prótesis de rodilla y de cadera se encuentra en el *Anexo 4*.

Los otros algoritmos estudiados, algoritmo Alpha y algoritmo Heurístico, representan estos procesos de una forma menos ordenada y más difícil de comprender que en el caso del algoritmo Inductivo.

7.2 Comparación de vías clínicas con registros de pacientes

Tanto en las patologías de prótesis de cadera como en las de prótesis de rodilla se han modelado en forma de redes de Petri las vías clínicas propuestas por los sanitarios. De esta manera se podrán comparar con el registro de actividades realizadas a los pacientes obtenidos en el apartado 3.3 *Diseño de vías clínicas*. Esta comparación será útil para elaborar nuevas vías clínicas ajustándose a la evidencia de los datos hasta ahora recogidos. Un ejemplo de estas comparaciones es el mismo caso de estudio que se plasmaba en el resultado anterior. Se compara ahora con el protocolo que marca la

vía clínica y se resaltan en rojo aquellas actividades que no han sido registradas pero sí constan en la vía clínica; en naranja las que están registradas pero en otro día y en verde aquellas que se registran pero no están contempladas en la vía clínica. El seguimiento de los pacientes se representa con el fondo de las actividades coloreado, mientras que red de Petri de la vía clínica proporcionada por los sanitarios tiene fondo blanco.

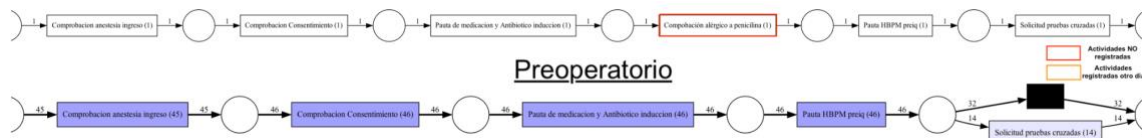


Figura 24: Comparación de vía clínica con registro de pacientes (PTR)

Los resultados completos se pueden encontrar en el *Anexo 4* y han sido presentados al equipo de médicos que ha colaborado con el trabajo. Se han analizado todas las comparaciones y los médicos han extraído interesantes conclusiones como las mostradas en el apartado 4.3 *Colaboración con el personal sanitario*. Se demuestra así la utilidad del seguimiento de pacientes y comparación con las vías clínicas actuales.

7.3 Implementación de metodologías en servidor web

Este trabajo colabora con el proyecto al que está asociado, en la labor de diseño de metodologías para el seguimiento de pacientes y extracción de vías clínicas, la cual ha sido mostrada en los dos apartados anteriores. Se muestra a continuación la implementación de dicha metodología en el servidor web que se está desarrollando (*Figura 25*). Según la patología que sufra el paciente, la aplicación indicará la vía clínica que se debe seguir sobre ese paciente y al pulsar sobre “Descubrir red de Petri”, se representará dicha vía en forma de red de Petri.

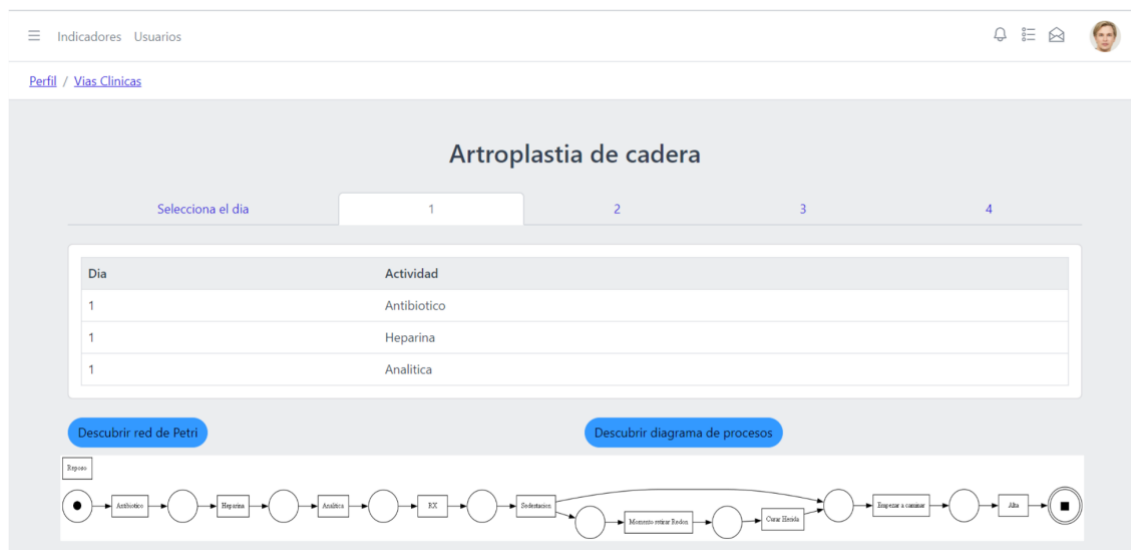


Figura 25. Servidor. Seguimiento del paciente en el detalle de las vías

7.4 Análisis clásico de datos

El análisis de datos realizado sobre antiguos registros de pacientes ha consistido en un modelo de regresión lineal y otro de regresión logística para predecir valores de las respectivas variables respuesta en cada caso. Se ha hecho un estudio de la posible relación entre distintas variables y parámetros de pacientes, para comprobar el efecto que pueden tener sobre otros, parámetros que a priori no parecen tener relación. En la parte final del análisis se han obtenido gráficos de conteo y de cajas para extraer más información sobre los pacientes y el proceso hospitalario (procedencia del paciente, motivos de intervención, días de operación, ingreso y alta...).

El estudio al completo de estos datos se encuentra en el *Anexo 5*.

8 Conclusiones

Una vez expuesto el desarrollo del trabajo de fin de grado se analiza el cumplimiento de los objetivos propuestos al comienzo del mismo y algunos de los resultados obtenidos.

1. Creación de un sistema de seguimiento de pacientes a partir de datos clínicos.

Por medio de los algoritmos de minería de procesos, concretamente el algoritmo Inductivo, se ha desarrollado un sistema de seguimiento y representación de las actividades que se realizan a los pacientes. Tan solo con los registros que se proporcionen de estas actividades, se podrá plasmar en forma de red de Petri la trayectoria que siguen los pacientes dentro del hospital. Me gustaría destacar el interés mostrado por el equipo de médicos del área de *Cirugía Ortopédica y Traumatología* del HCU en esta manera de representar los procesos clínicos; llegando incluso a mencionar la utilidad de la red de Petri para mostrar el seguimiento de actividades del paciente. Este formalismo es muy válido para dicha labor y por medio del marcado de la red, se representarían sus distintos estados, que en este caso se interpretarían como las actividades que se han llevado a cabo y las que quedan por realizar, lo que ayudaría por ejemplo a que no se repitan actividades ya efectuadas .

2. Optimización de vías clínicas para pacientes con patologías de artroplastia de cadera y de rodilla.

La labor de este trabajo desde este enfoque es la de realizar el seguimiento de los pacientes con dichas patologías, compararlos con sus respectivas vías clínicas actuales y mostrarlo al personal sanitario. Las decisiones de rediseño de las vías son por supuesto competencia de los especialistas médicos, pero a la hora de mostrarles los resultados, han quedado sorprendidos con el hecho de que haya actividades que no se reflejan en las vías clínicas actuales y que haya otras que sí aparecen pero no se registran, por lo que deben de empezar a controlarse. Esto demuestra que aunque haya interés por renovar o implantar nuevas vías clínicas para la homogeneización de los procesos, hay supuestos de actividades que no se habían valorado previamente y que la extracción de la trayectoria de los pacientes a partir de datos clínicos resulta útil para el diseño de vías clínicas.

3. Estandarización y digitalización de la gestión hospitalaria.

La representación del seguimiento de pacientes ha sido implementada con éxito en el servidor web que se desarrolla a nivel del proyecto general. Es una metodología que se está poniendo a prueba con pacientes que siguen las vías clínicas de artroplastia de cadera y de rodilla porque son los datos que se nos han proporcionado. No obstante, al incluir el algoritmo Inductivo, permite modelar procesos sin importar el tamaño del registro ni la patología, si los datos se presentan en el formato adecuado. Por este motivo, si el servidor sigue desarrollándose en otras áreas del hospital, podrá ser igual de válido a la hora de representar el seguimiento de pacientes con patologías propias de dichas áreas.

9 Bibliografía

- [1] Rojas, E., Munoz-Gama, J., Sepúlveda, M., & Capurro, D. (2016). Process mining in Healthcare: A literature review. *Journal of Biomedical Informatics*, 61, 224–236. doi:10.1016/j.jbi.2016.04.007.
- [2] van der Aalst, W. M. P. (2011). *Process Mining : Discovery, Conformance And Enhancement Of Business Processes*. 1st ed. Heidelberg: Springer-Verlag, pp.125-211.
- [3] van der Aalst, W. M. P. (2014). Process mining in the large: A tutorial. *Business Intelligence*, 33–76. doi:10.1007/978-3-319-05461-2_2.
- [4] Munoz-Gama, J., Martin, N., Fernandez-Llatas, C., Johnson, O. A., Sepúlveda, M., Helm, E., ... Zerbato, F. (2022). Process mining for healthcare: Characteristics and challenges. *Journal of Biomedical Informatics*, 127, 103994. doi:10.1016/j.jbi.2022.103994.
- [5] Colom, J.-M., & Desel, J. (2013). *Application and theory of Petri Nets and concurrency 34th International Conference, petri nets 2013, Milan, Italy, June 24-28, 2013: Proceedings*. Heidelberg: Springer.
- [6] Schuster, D., van Zelst, S. J., & van der Aalst, W. M. (2020). Incremental discovery of hierarchical process models. *Research Challenges in Information Science*, 417–433. doi:10.1007/978-3-030-50316-1_25
- [7] University of Technology, T. U. E. (2013). Retrieved from <https://www.futurelearn.com/>
- [8] Weijters, A. J. M. M. , van der Aalst, W. M. P. & de Medeiros A. K. A.. Process mining with the heuristics miner-algorithm. Technical report, Technische Universiteit Eindhoven, 2006.
- [9] *Probabilidad y Estadística*(2023b). Retrieved from <https://www.probabilidadyestadistica.net/>
- [10] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2022). In *An introduction to statistical learning: With applications in R* (pp. 59–189). Boston: Springer.
- [11] Mans, R. S., Schonenberg, M. H., Song, M., van der Aalst, W. M., & Bakker, P. J. (2008). Application of process mining in healthcare – a case study in a Dutch Hospital. *Biomedical Engineering Systems and Technologies*, 425–438. doi:10.1007/978-3-540-92219-3_32.
- [12] F. Mannhardt, D. Blinde, *Analyzing the trajectories of patients with sepsis using process mining, RADAR+ EMISA@ CAiSE 1859 (2017) 72–80.*
- [13] Mahulea, C., Mahulea, L., Garcia-Soriano, J.M, & Colom J.M. (2018). Modular Petri Net Modeling of Healthcare Systems. *Flexible Services and Manufacturing Journal*, 30(1-2): 329-357, June 2018.
- [14] Bernardi, S., Mahulea, C. & Albareda, J (2019). *Toward a decision support system for the clinical pathways assessment. Discrete Event Dynamic Systems: Theory and Applications*, 29(1):91-125, March 2019.
- [15] Silva, M. (2002). *Las Redes de Petri : en la Automática y la Informática*. 10.26754/uz.84-7288-045-1 (https://www.researchgate.net/publication/325164472_Las_Redde_de_Petri_en_la_Automatica_y_la_Informatica).
- [16] Clavel, D., Mahulea, C., Albareda J. & Silva, M. (2020). *A Decision Support System for Elective Surgery Scheduling under Uncertain Durations. Applied Sciences, Special Issue on "Medical Informatics and Data Analysis"*, 10(6):1937.

Anexos

Anexo 1. Vías clínicas existentes

Las vías clínicas son instrumentos que se usan para coordinar, planificar y marcar la secuencia de procedimientos a seguir sobre el paciente según una patología en concreto. Pueden tener un carácter general y ser útiles también para un grupo de enfermedades relacionadas.

Abarcan el período de tiempo que va desde el ingreso del paciente hasta que este recibe el alta. Todo ello es necesario para conseguir la máxima eficiencia en el proceso asistencial de los pacientes, ya que permite la estandarización de cuidados para un grupo de pacientes con una misma patología y un curso clínico similares.

Pese a su utilidad, resulta llamativa la escasa implantación de esta metodología en los hospitales, como es el caso del Hospital Clínico Universitario. Por eso se pretende mejorar el seguimiento y representación de pacientes que sufren estas patologías y ayudar en el diseño de las vías clínicas.

Se presentan en este anexo cómo son las vías clínicas actuales proporcionadas por médicos MIR que han colaborado en el proyecto.

ARTROPLASTIA TOTAL DE RODILLA

Preingreso:

- Radiografía, telemétricas
- Causa de la intervención
- Consentimientos de la cirugía
- Valoración preanestésica en consultas

PrelQ:

- Comprobar valoración por anestesia/ petición anestesia intrahospitalaria
- Comprobación de CI
- Pauta de medicación + Antibiótico en inducción
- Alérgico a penicilina si / no
- Pauta de HBPM 12h preiq
- Solicitud de PC según criterios actuales

IQ:

- Comprobaciones ayunas
- Comprobación suspensión de HBPM 12h preiq
- Administración de AB
- Check list
- Duración de IQ
- Protocolo IQ
- Solicitud de analítica
- Solicitud de Rx
- Tratamiento AB post IQ
- Ácido tranexámico Sí/No
- Bloqueo nervioso anestésico
- Infiltración de cápsula posterior

1DPO:

- Retirada de Redón (si redón -300ml)
- Sedestación en sillón
- Revisión de AS postiq
- Transfusión Si/no
- Manchado de apósito quirúrgico
- Tratamiento AB post iq administrado
- HBPM administrada
- Comprobación de la flexoextensión
- Retirada de vendaje compresivo

2DPO:

- Dolor petición de rescates analgésicos
- Bipedestación y deambulación

3DPO:

- Valoración de deambulación
- Entrega de documentación y explicación de precauciones al alta
- Alta

ARTROPLASTIA DE CADERA PRIMARIA EVOLUTION/ SWEYMULLER

Preingreso:

- Radiografía, telemétricas
- Causa de la intervención
- Consentimientos de la cirugía
- Valoración preanestésica en consultas

PrelQ:

- Comprobar valoración por anestesia/ petición anestesia intrahospitalaria
- Comprobación de CI
- Pauta de medicación + Antibiótico en inducción
- Alérgico a penicilina si / no
- Pauta de HBPM 12h preiq
- Solicitud de PC según criterios actuales

IQ:

- Comprobaciones ayunas
- Comprobación suspensión de HBPM 12h preiq
- Administración de AB
- Check list
- Duración de IQ
- Protocolo IQ
- Solicitud de analítica
- Solicitud de Rx
- Tratamiento AB post IQ

1DPO:

- Retirada de Redón (si redón -300ml)
- Sedestación en sillón
- Revisión de AS postiq
- Transfusión Si/no
- Manchado de apósito quirúrgico
- Tratamiento AB post iq administrado
- HBPM administrada

2DPO:

- Dolor petición de rescates analgésicos
- Bipedestación y deambulación

3DPO:

- Valoración de deambulación
- Entrega de documentación y explicación de precauciones al alta
- Alta

Anexo 2. Análisis clásico de datos

El análisis clásico de datos es un enfoque estadístico utilizado para analizar y comprender la relación entre variables y realizar predicciones o inferencias a partir de los datos observados. El objetivo principal es encontrar patrones, tendencias y relaciones subyacentes entre los datos, lo que permite tomar decisiones informadas y respaldadas por evidencia.

Dentro de este análisis clásico, se ha optado por hacer estudios de regresión lineal, polinomial y logística. Se describirán a continuación conceptos básicos de cada uno de ellos.

Previo a la presentación de estos estudios, resulta importante definir los siguientes parámetros estadísticos [9]:

- Covarianza: es una medida de dispersión que indica el grado de variación conjunta de dos variables. Sirve para analizar la dependencia entre dos variables. Si la covarianza es positiva, cuando una variable aumente, la otra también lo hará; si es negativa, el aumento de una provocará la disminución de la otra y viceversa; y si es igual a 0, no hay relación entre variables.
- Coeficiente de correlación de Pearson (r): o simplemente de correlación, es una medida estadística que cuantifica la relación entre dos variables. Es un valor que varía de -1 a 1 , siendo $r = -1$ una relación lineal perfecta negativa entre ambas variables; $r = 1$ una relación lineal perfecta positiva; $r = 0$ indicará que no hay relación lineal entre ellas. Los escenarios intermedios señalarán cuánto de próxima es esa relación a ser negativamente perfecta ($-1 < r < 0$) o positivamente perfecta ($0 < r < 1$)

Regresión lineal

La regresión lineal es una técnica estadística utilizada para modelar y analizar la relación entre una variable dependiente continua y una o más variables independientes [10]. Estos dos tipos de variables son:

- Variable dependiente: es la variable que se intenta predecir o explicar a través del modelo de regresión. También se conoce como variable respuesta.
- Variables independientes: son las variables que se utilizan para predecir o explicar la variable dependiente. También se conocen como variables explicativas.

Los modelos de regresión lineal asumen una relación lineal entre las variables independientes y la variable dependiente, es decir, los cambios en las variables explicativas se asocian a cambios proporcionales en la variable dependiente. Esta relación se puede expresar matemáticamente como una recta $y = \beta_0 + \beta_1 x$, con y como variable respuesta y x como variable independiente. El coeficiente de intercepción β_0 representa el valor de y cuando las variables independientes sean nulas; mientras que el coeficiente β_1 indica el cambio promedio en la variable respuesta asociado con un cambio de una unidad en la variable dependiente, es decir, es la pendiente de esa recta.

En el caso de que se quisiera relacionar la variable dependiente con más de una variable independiente, supóngase hasta p variables, la expresión que explique la variable será del tipo $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$.

El motivo por el que se utiliza la regresión lineal es para construir modelos en los que dados un conjunto de variables independientes y sus correspondientes variables respuesta, sean capaces de trazar la recta que mejor se ajuste a este conjunto de

parejas de datos y así poder predecir futuras variables dependientes. Esto se conseguirá utilizando el método de los mínimos cuadrados.

Un parámetro importante a la hora de tratar con modelos de predicción mediante técnicas de regresión, es el coeficiente de determinación R^2 . Indica qué cantidad de variación en y puede explicarse por la dependencia de x , utilizando el modelo de regresión con el que se trabaje. Este coeficiente toma valores de 0 a 1, de manera que un mayor R^2 indica un mejor ajuste y significa que el modelo puede explicar mejor la variación de la salida con diferentes entradas. En el caso de que $R^2=1$, se tendrá el ajuste perfecto, ya que los valores de las respuestas predichas y reales se ajustan completamente entre sí

En la siguiente figura se muestra un ejemplo de aplicación de este modelo predictivo (Figura 26). La variable dependiente será en este caso el número de ventas (*sales*) y la variable explicativa será TV, refiriéndose al presupuesto invertido en anuncios de televisión para mejorar esas ventas. Como se acaba de exponer, se le dan valores de la y con su correspondiente x y será el modelo el que trace la recta que mejor se ajuste y explique la relación que hay entre sendas variables.

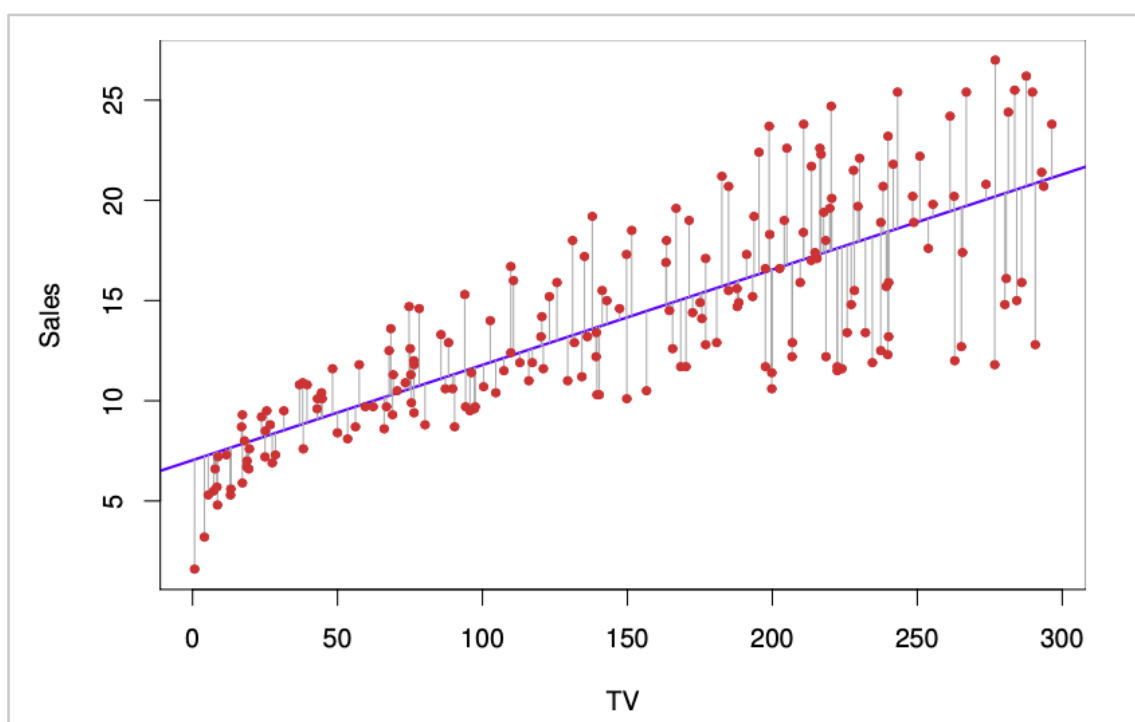


Figura 26: Regresión lineal [10]

Cabe mencionar en esta sección la regresión polinómica. Este tipo de regresión no es más que una extensión del modelo lineal, al que se le añaden variables predictoras extra, obtenidas elevando las variables independientes a alguna potencia. De tal manera que la expresión que explique la variable respuesta será del tipo $y = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_p x_i^p$, siendo i el identificador de la variable predictora y p el grado de la potencia más alta. Se presenta en la siguiente figura un ejemplo de explicación de la variable y mediante regresión polinómica.

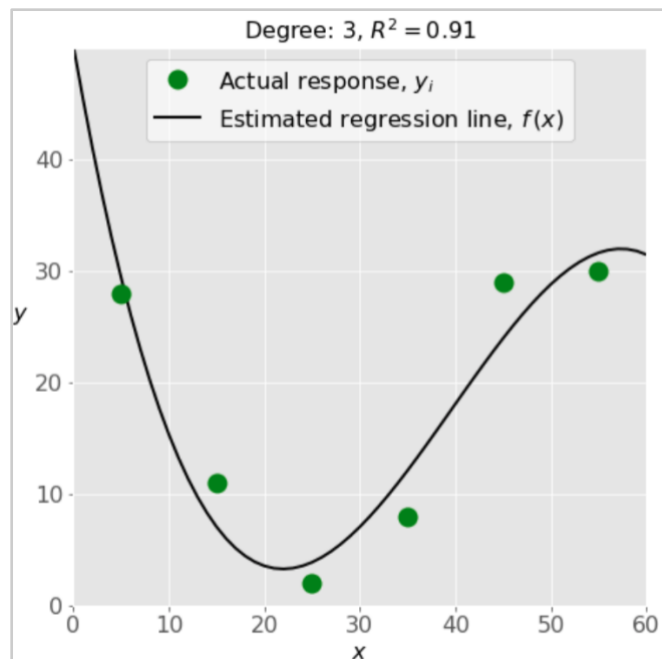


Figura 27: Regresión polinómica (<https://medium.com/@curryrowan/simplified-logistic-regression-classification-with-categorical-variables-in-python-1ce50c4b137>)

Analizando la covarianza y coeficiente de correlación de Pearson entre distintas variables continuas del historial clínico de los pacientes, se buscarán aquellas variables que tengan más relación entre sí y se intentarán obtener modelos explicativos para esas parejas de variables.

Regresión logística

La regresión logística es otro tipo de análisis de regresión, usado para modelar y analizar la relación entre una variable dependiente categórica (binaria o multinomial) y una o más variables independientes. Para una explicación más simple de esta regresión, se usarán variables dependientes tan solo binarias.

A diferencia de la regresión lineal, la regresión logística no asume una relación lineal entre las variables independientes y la variable dependiente. En cambio, utiliza una función logística para modelar la probabilidad de que la variable dependiente pertenezca a una categoría específica [10].

Dicha probabilidad, $p(X) = Pr(Y = 1|X)$, con Y como variable dependiente y X como independiente, no puede ser predicha con un modelo lineal, ya que la variable dependiente solo toma valores de 0 y 1. Si se intentase predecir de manera lineal, la aproximación sería tan imprecisa como la mostrada en la Figura 28.a. Es por eso que se usará la *función logística*:

$$p(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} \quad (\text{Fuente [10]})$$

Con esta función, se calculará la probabilidad de la variable dependiente “Default” en este caso, de manera más precisa (Figura 28.b)

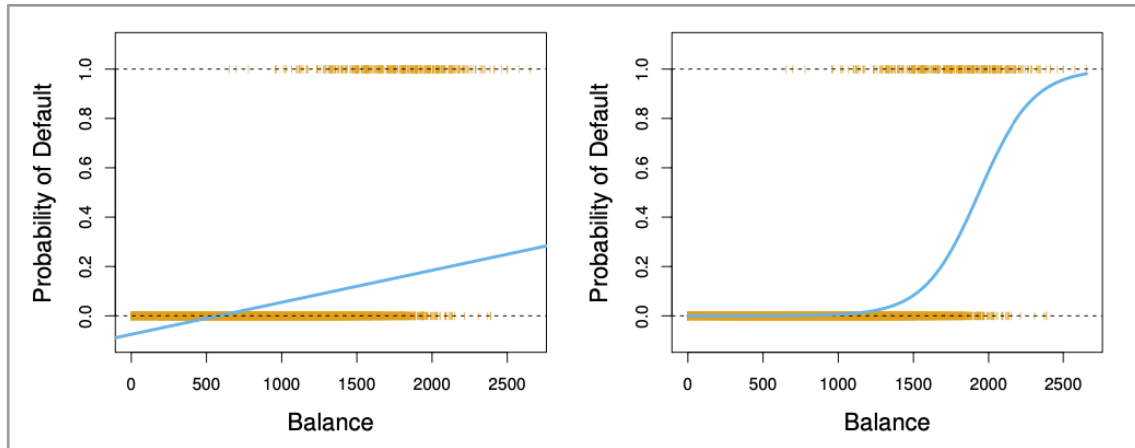


Figura 28: Regresión logística. Aproximación lineal (a) y aproximación logística (b) [10]

Al contrario que en el caso de la regresión lineal, la precisión de las predicciones de modelos de regresión logística es más complicada de cuantificar. Por eso, una herramienta que se suele utilizar en este tipo de regresión es la matriz de confusión (Figura 29). En ella, se plasmarán los valores predichos de la variable dependiente y, tomando el 1 como valor positivo y el 0 como negativo, se clasificarán en valores positivos verdaderos (*TP*) o falsos (*FP*) y valores negativos verdaderos (*TN*) o falsos (*FN*). De esta manera se podrán cuantificar los valores que han sido predichos correctamente, para las dos categorías que presenta la variable dependiente binaria. Este mismo análisis se podrá realizar para variables categóricas multinomiales, añadiendo al 0 y 1, tantos valores como categorías tenga la variable categórica.

		PREDICTIVE VALUES	
		POSITIVE (1)	NEGATIVE (0)
ACTUAL VALUES	POSITIVE (1)	TP	FN
	NEGATIVE (0)	FP	TN

Figura 29: Regresión logística. Matriz de confusión
<https://medium.com/@curryrowan/simplified-logistic-regression-classification-with-categorical-variables-in-python-1ce50c4b137>

Gráficos estadísticos

Adicionalmente, se utilizarán gráficos de conteo (*countplots*) y gráficos de cajas (*boxplots*) como herramientas de visualización del análisis estadístico entre variables continuas y categóricas.

Los gráficos de conteo se emplean para contar y visualizar la frecuencia de la categoría elegida del conjunto de datos; mientras que los gráficos de caja muestran de manera resumida la distribución de un conjunto de datos y proporcionan información sobre su centralidad, dispersión, asimetría y presencia de valores atípicos.

Esta información de los gráficos de caja incluye (*Figura 30*):

- Mediana(*median*): se muestra como una línea dentro de la caja del gráfico y marca el valor medio del conjunto de datos. Divide la distribución en dos partes iguales, con el 50% de los datos por encima y el 50% por debajo. Se conoce también como el segundo cuartil.
- Cuartiles (*quartiles*): en este tipo de gráficos se dividen los datos en cuartiles, donde el primer cuartil (Q1) representa el 25% inferior de los datos, el segundo cuartil (Q2) es la mediana y el tercer cuartil (Q3) representa el 75% inferior de los datos, es decir, que solo un 25% de los datos es mayor a esa separación.
- Rango intercuartílico (RIC, en inglés *IQR*): es la distancia entre el tercer cuartil y el primer cuartil ($Q3 - Q1$). La caja del *boxplot* se extiende desde el primer cuartil hasta el tercer cuartil, y el RIC proporciona una medida de la dispersión de los datos.
- Bigotes (*whisker*): son líneas que se extienden desde la caja y representan la variabilidad de los datos fuera del RIC. Los bigotes generalmente se extienden hasta el valor más alejado dentro de un rango específico.
- Valores atípicos: los gráficos de caja también identifican los valores atípicos, que son puntos de datos que se encuentran más allá de los bigotes. Estos valores pueden indicar observaciones inusuales o errores en los datos y son importantes para tener en cuenta al interpretar la distribución de los datos.

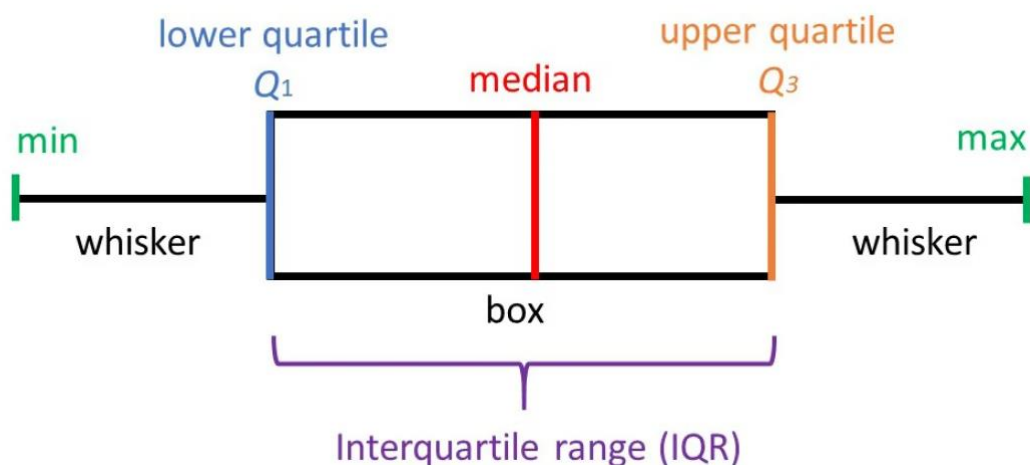


Figura 30: Ejemplo de gráfico de caja
(<https://www.simplypsychology.org/boxplots.html>)

Anexo 3. Interfaz del servidor web

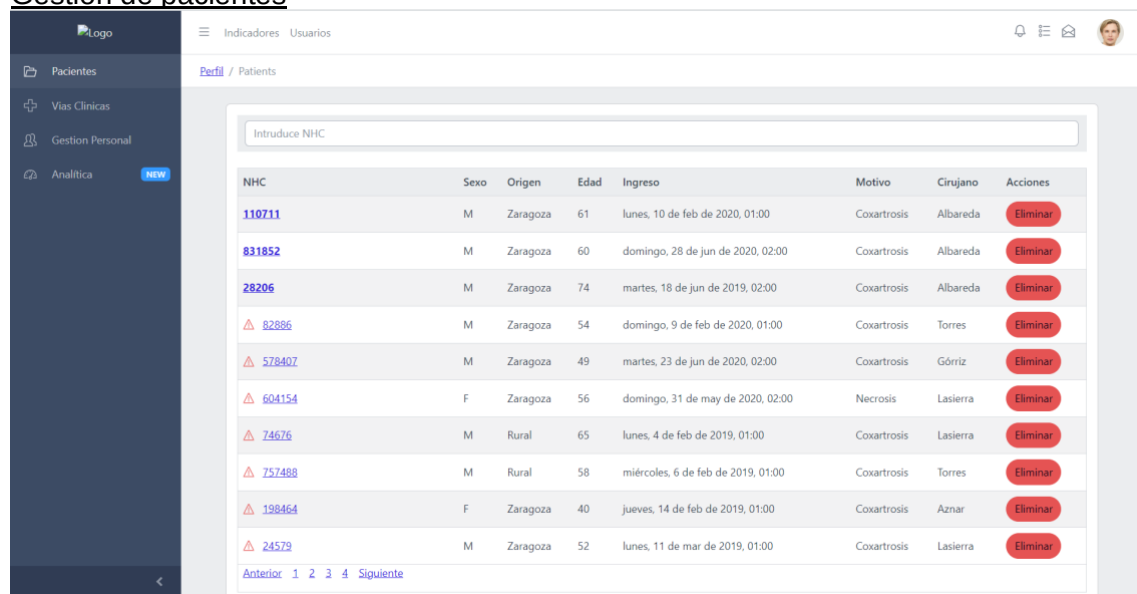
El servidor web que se está desarrollando en el proyecto al que está asociado este trabajo se compone de las siguientes funcionalidades:

Página de login



Figura 31. Servidor. Página de login

Gestión de pacientes



NHC	Sexo	Origen	Edad	Ingreso	Motivo	Cirujano	Acciones
110711	M	Zaragoza	61	lunes, 10 de feb de 2020, 01:00	Coxartrosis	Albareda	Eliminar
831852	M	Zaragoza	60	domingo, 28 de jun de 2020, 02:00	Coxartrosis	Albareda	Eliminar
28206	M	Zaragoza	74	martes, 18 de jun de 2019, 02:00	Coxartrosis	Albareda	Eliminar
82886	M	Zaragoza	54	domingo, 9 de feb de 2020, 01:00	Coxartrosis	Torres	Eliminar
578407	M	Zaragoza	49	martes, 23 de jun de 2020, 02:00	Coxartrosis	Górriz	Eliminar
604154	F	Zaragoza	56	domingo, 31 de may de 2020, 02:00	Necrosis	Lasierra	Eliminar
74676	M	Rural	65	lunes, 4 de feb de 2019, 01:00	Coxartrosis	Lasierra	Eliminar
757488	M	Rural	58	miércoles, 6 de feb de 2019, 01:00	Coxartrosis	Torres	Eliminar
198464	F	Zaragoza	40	jueves, 14 de feb de 2019, 01:00	Coxartrosis	Aznar	Eliminar
24579	M	Zaragoza	52	lunes, 11 de mar de 2019, 01:00	Coxartrosis	Lasierra	Eliminar

Figura 32. Servidor. Gestión de pacientes

Detalle de los pacientes

Indicadores Usuarios

Perfil / Patients

Paciente	Via Clínica	Finalizado
110711	Artroplastia de cadera Modificar Via	Via finalizada ✓

Día	Actividad	Acciones
1	Antibiotico ✓	Ver comentarios Eliminar
1	Heparina ✓	Ver comentarios Eliminar
1	Analítica ✓	Ver comentarios Eliminar
1	RX ✓	Ver comentarios Eliminar
1	Sedestacion ✓	Ver comentarios Eliminar
2	Empezar a caminar ✓	Ver comentarios Eliminar
2	Momento retirar Redon ✓	Ver comentarios Eliminar
2	Curar Herida ✓	Ver comentarios Eliminar
3	Reposo ✓	Ver comentarios Eliminar
.	Alta ✓	Ver comentarios Eliminar

Figura 33. Servidor. Detalle de pacientes

Vías clínicas

Logo

Indicadores Usuarios

Perfil / Vías Clínicas

Pacientes

Vías Clínicas

Gestión Personal

Analítica NEW

Patología

Artroplastia de cadera

Artroplastia total de rodilla

Figura 34. Servidor. Vías clínicas

Detalle de las vías

Indicadores Usuarios

Perfil / Vías Clínicas

Artroplastia de cadera

Selecciona el día

1 2 3 4

Día	Actividad
1	Antibiotico
1	Heparina
1	Analitica

Descubrir red de Petri

Descubrir diagrama de procesos

Figura 35. Servidor. Detalle de las vías

Gestión del personal

Indicadores Usuarios

Perfil / Dashboard

ID	Nombre	Apellidos	Tipo	Edad	Especialidad	Fecha de registro	Area	Acciones
12345	Antonio	Torres	ADJUNTO	39	Cirugía ortopédica y traumatología	7/3/2023, 14:33:35	Sector 3	Añadir contraseña Eliminar
12346	Luis	Góriz	ADJUNTO	48	Cirugía ortopédica y traumatología	7/3/2023, 14:33:35	Sector 3	Añadir contraseña Eliminar
12348	Miguel	Ruiz Frontera	ADJUNTO	31	Cirugía ortopédica y traumatología	7/3/2023, 14:33:35	Sector 3	Añadir contraseña Eliminar
12349	Jose Manuel	Lasierra	ADJUNTO	61	Cirugía ortopédica y traumatología	7/3/2023, 14:33:35	Sector 3	Añadir contraseña Eliminar
12350	Jose Maria	Aznar	ADJUNTO	66	Cirugía ortopédica y traumatología	7/3/2023, 14:33:35	Sector 3	Añadir contraseña Eliminar
12351	Francisco Javier	Nicolau	ADJUNTO	49	Cirugía ortopédica y traumatología	7/3/2023, 14:33:35	Sector 3	Añadir contraseña Eliminar
12347	Jorge	Albareda	ADJUNTO	64	Cirugía ortopédica y traumatología	7/3/2023, 14:33:35	Sector 3	Añadir contraseña Eliminar

Crear doctor

Figura 36. Servidor. Gestión del personal

También tendrá una pestaña donde se mostrarán análisis estadísticos de los datos de los pacientes como los realizados en este trabajo en la sección 2.5 *Análisis clásico de datos*, pero se encuentra actualmente en desarrollo debido a la escasez de datos para la inferencia de conclusiones.

Anexo 4. Resultados completos para cada caso de estudio

En este anexo se muestran las redes de Petri completas que modelan el seguimiento de los pacientes para las patologías de artroplastia de rodilla (ATR) y artroplastia de cadera (ATC). Seguidamente, se presentan las comparaciones de dichos seguimientos con las vías clínicas actuales, tal y como se han presentado al equipo de médicos que han colaborado en el trabajo.

Seguimiento de pacientes (ATR)

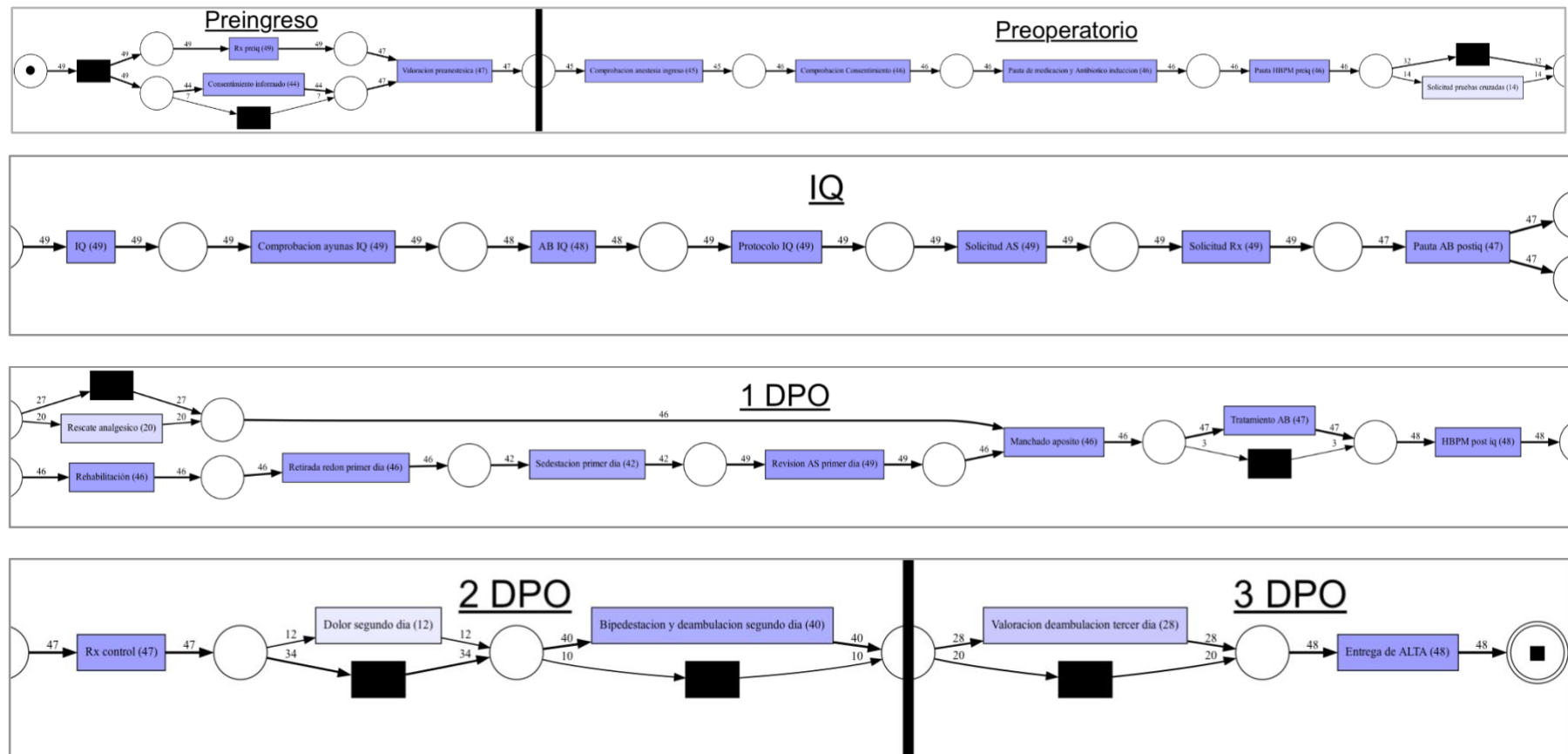
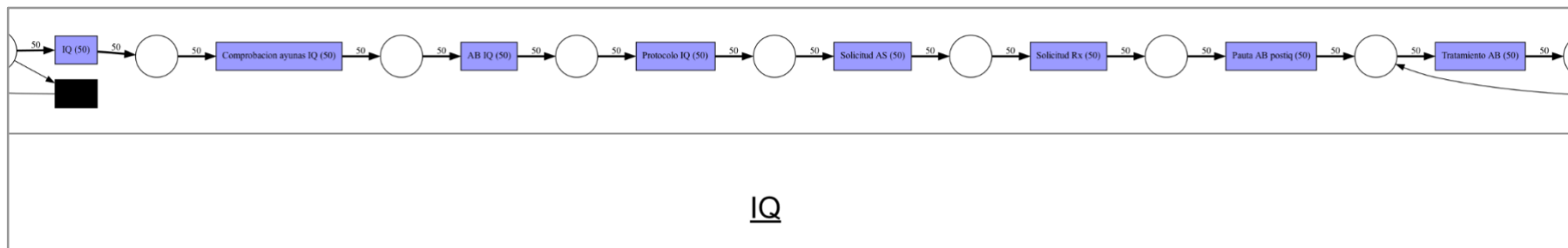
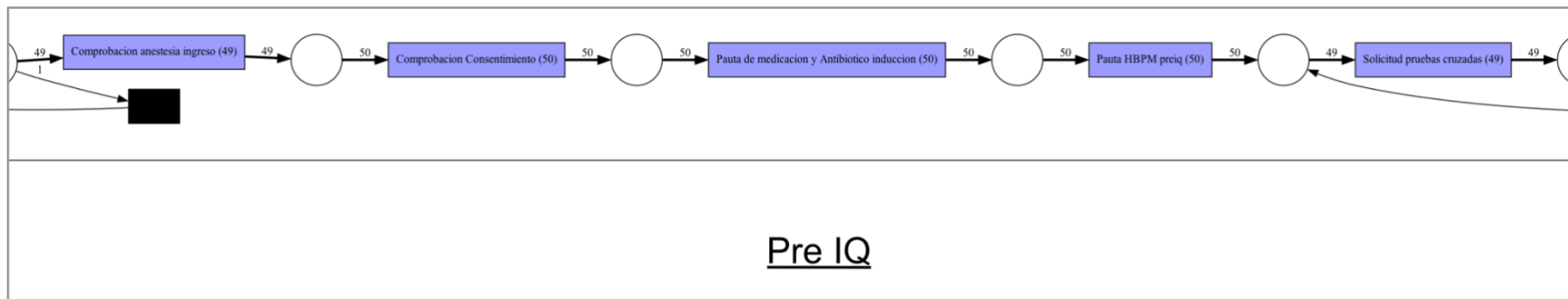
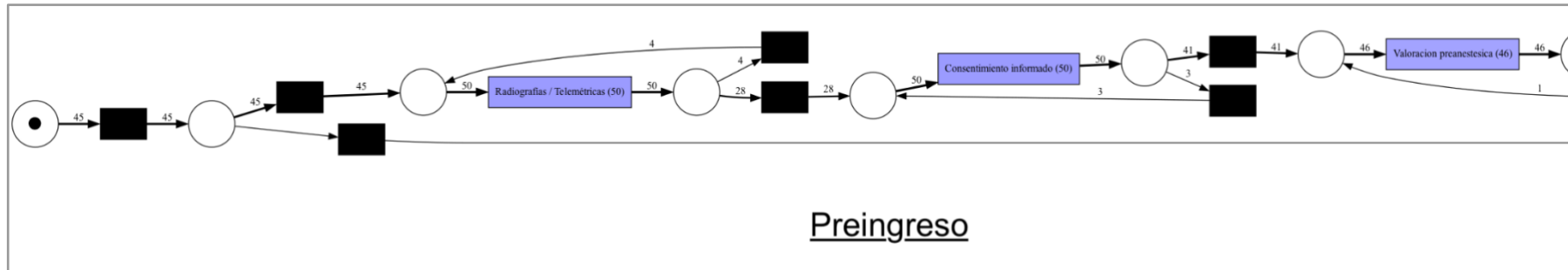


Figura 37: Red de Petri para el seguimiento de pacientes (ATR)

Seguimiento de pacientes (ATC)



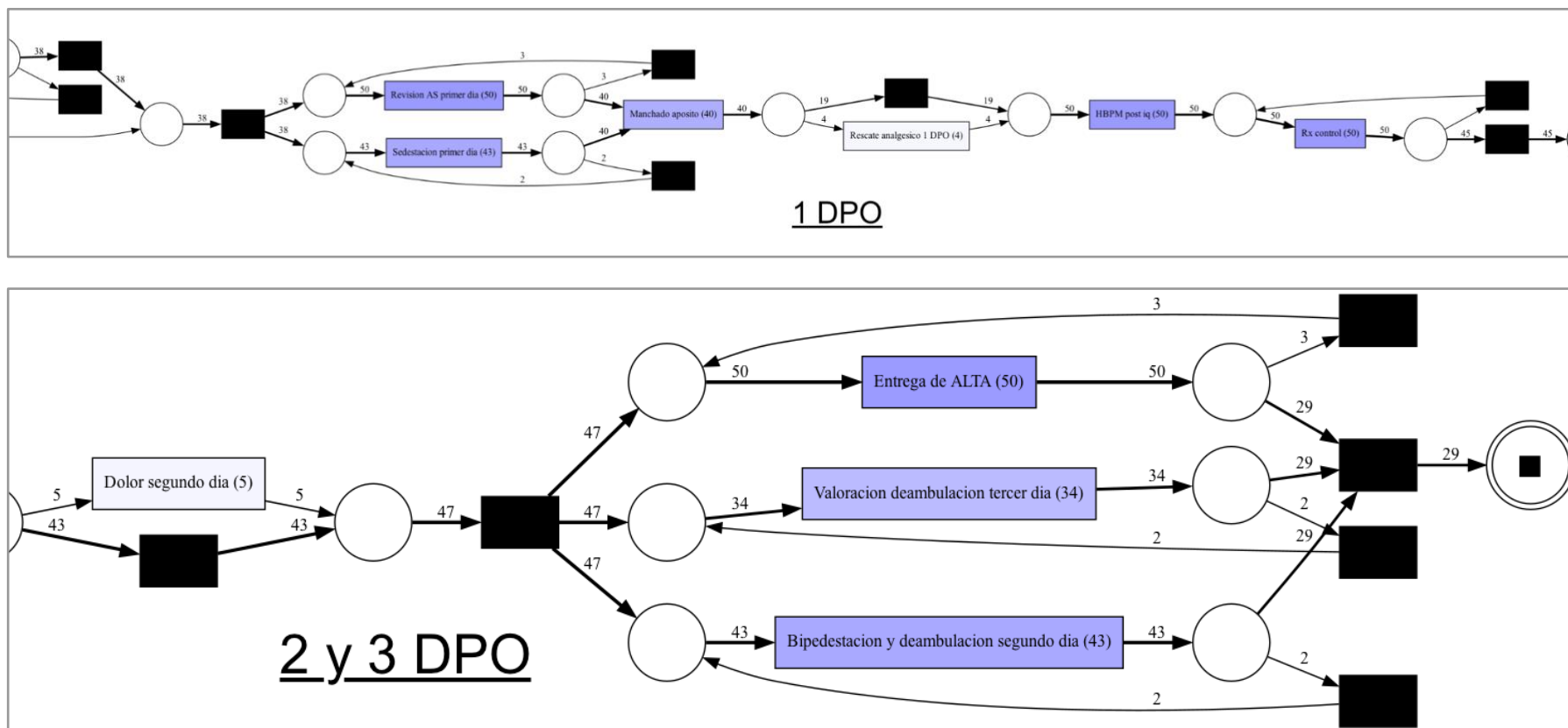
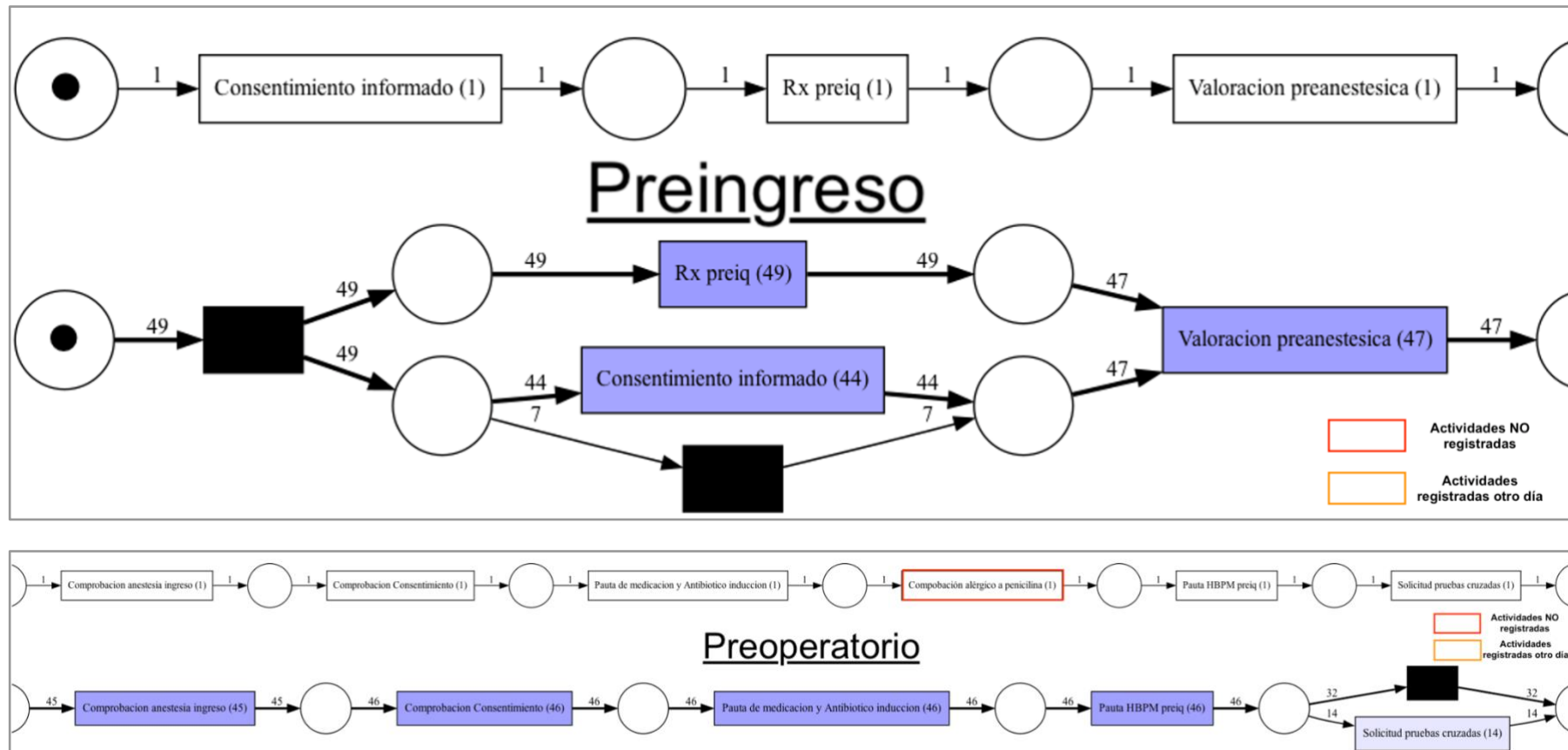
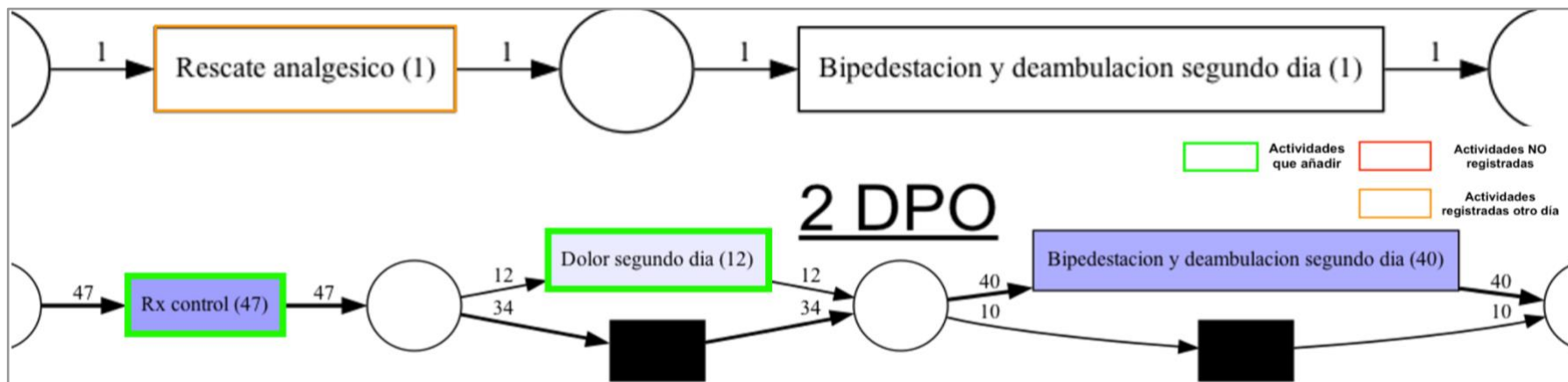
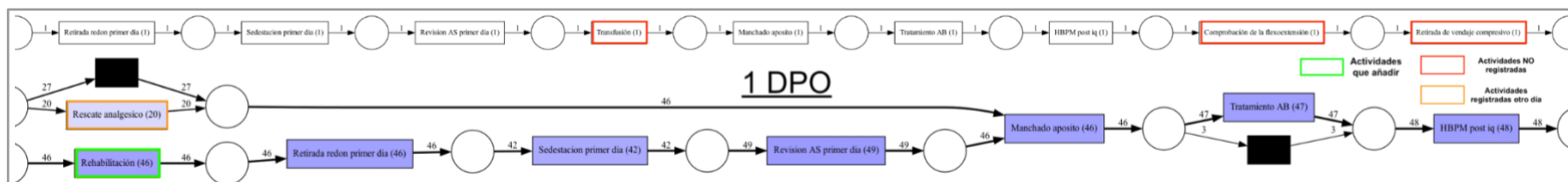
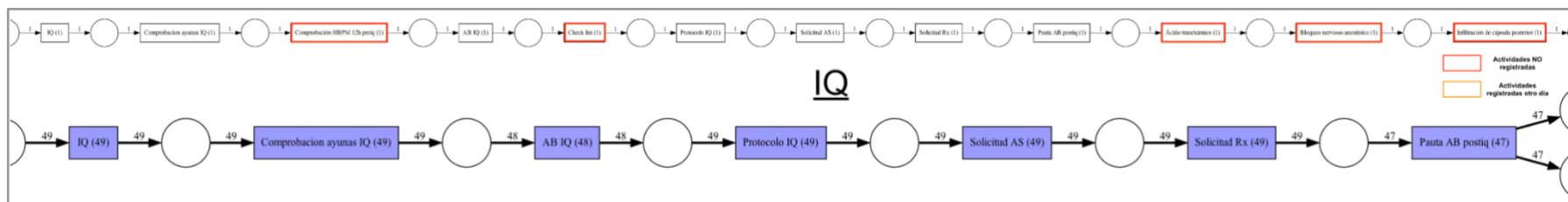


Figura 38: Red de Petri para el seguimiento de pacientes (ATC)

Se compara ahora con el protocolo que marca la vía clínica y se resaltan en rojo aquellas actividades que no han sido registradas pero sí constan en la vía clínica; en naranja las que están registradas pero en otro día y en verde aquellas que se registran pero no están contempladas en la vía clínica. El seguimiento de los pacientes se representa con el fondo de las actividades coloreado, mientras que red de Petri de la vía clínica proporcionada por los sanitarios tiene fondo blanco.

Comparación de seguimiento de pacientes con vías clínicas (ATR)





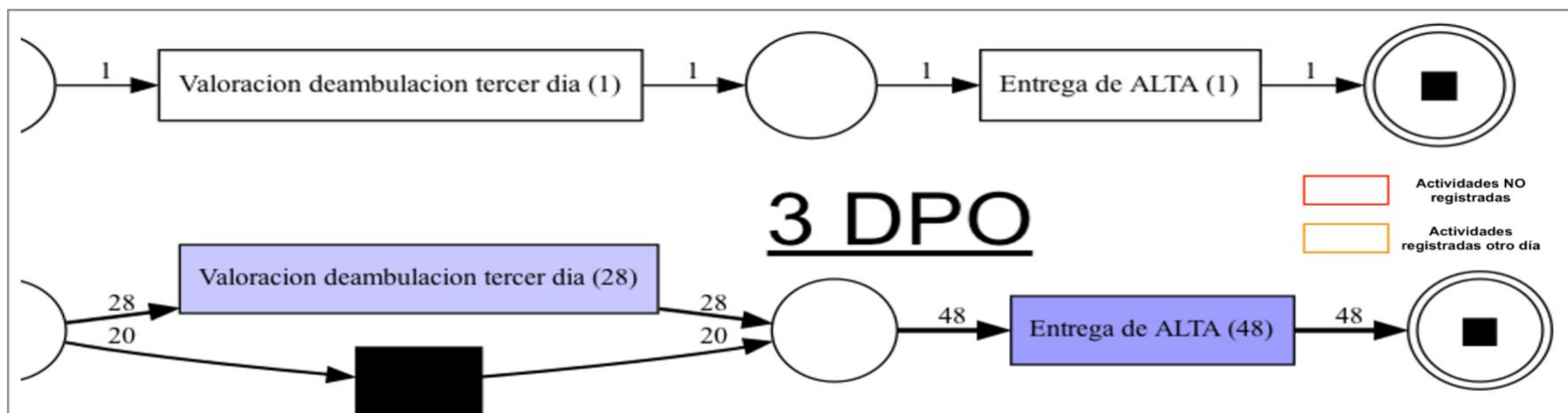
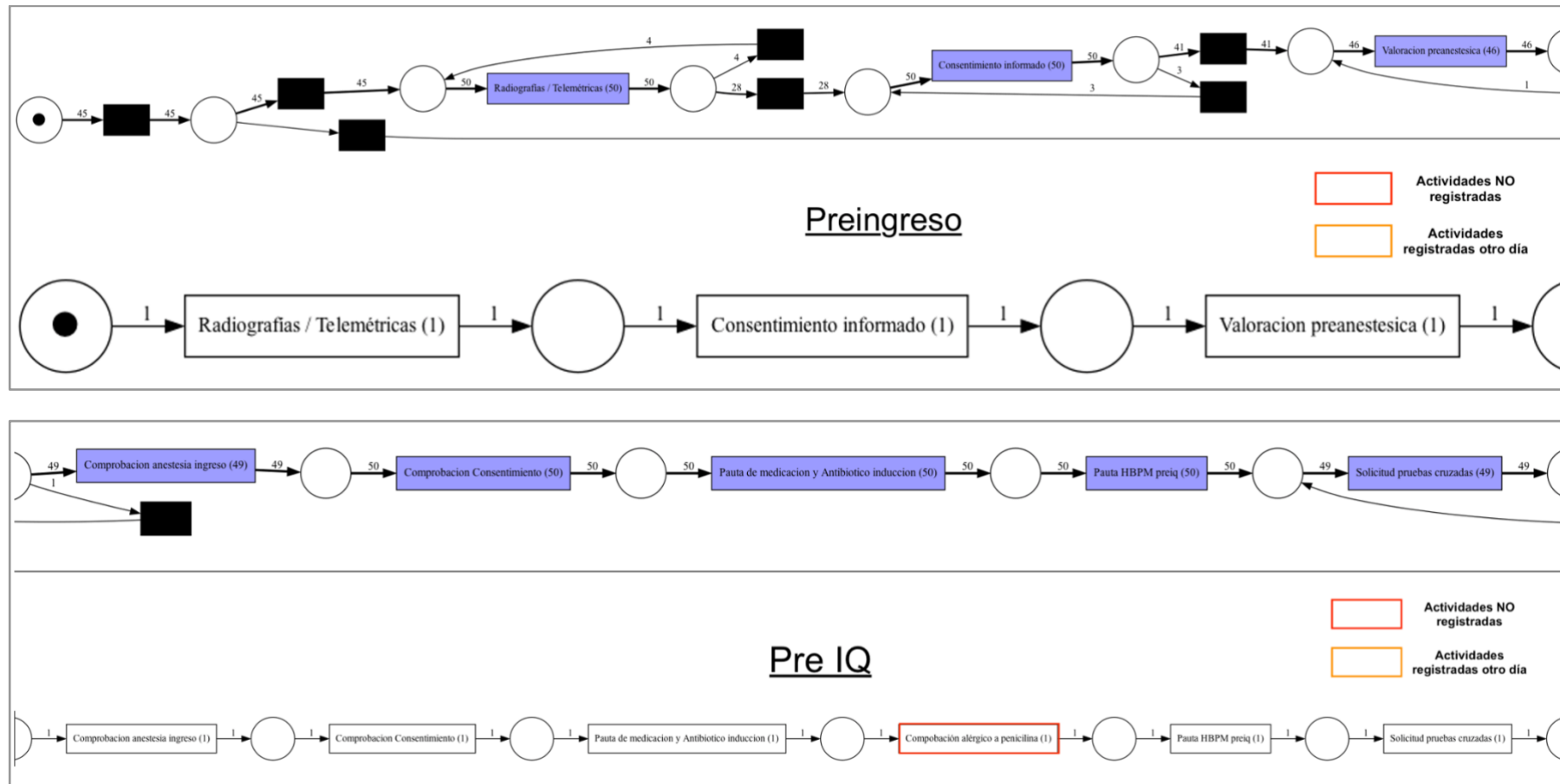
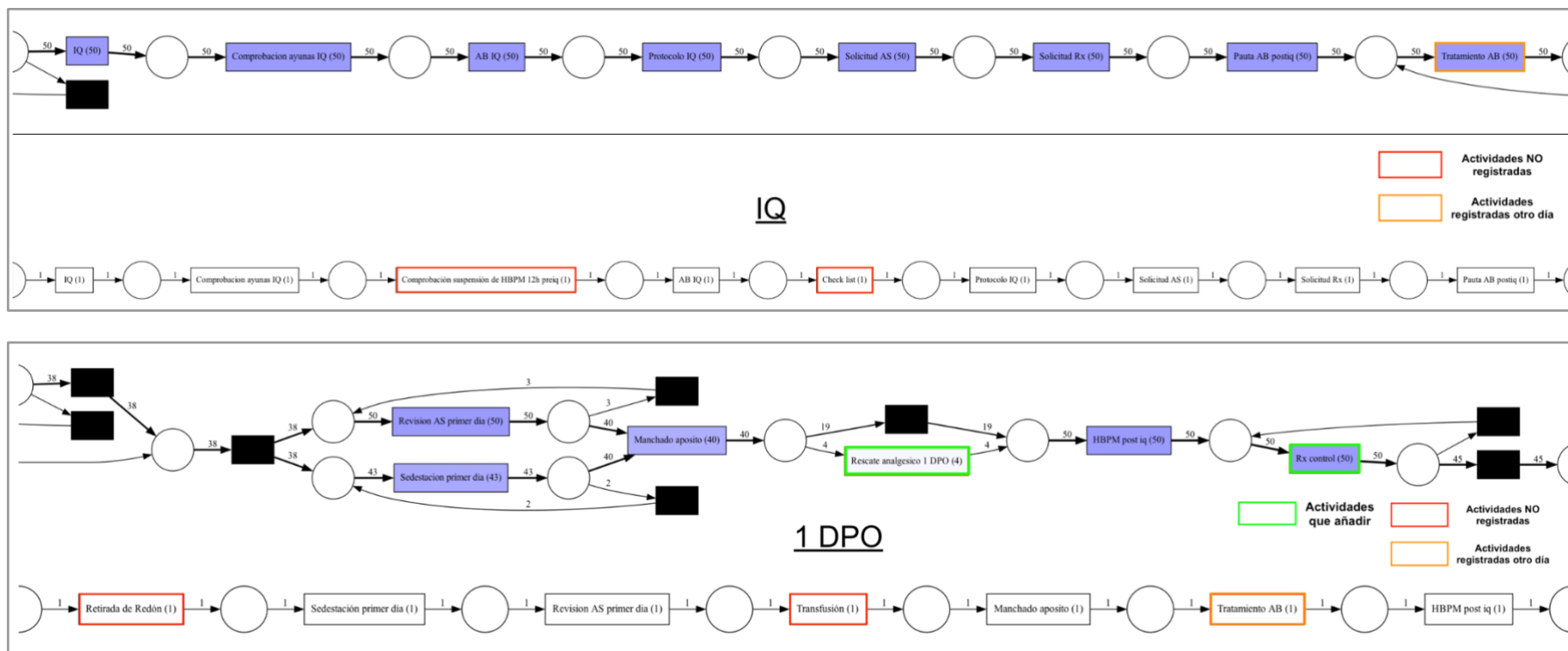


Figura 39: Comparación entre las redes de Petri del seguimiento de pacientes y la vía clínica(ATR)

Comparación de seguimiento de pacientes con vías clínicas (ATC)





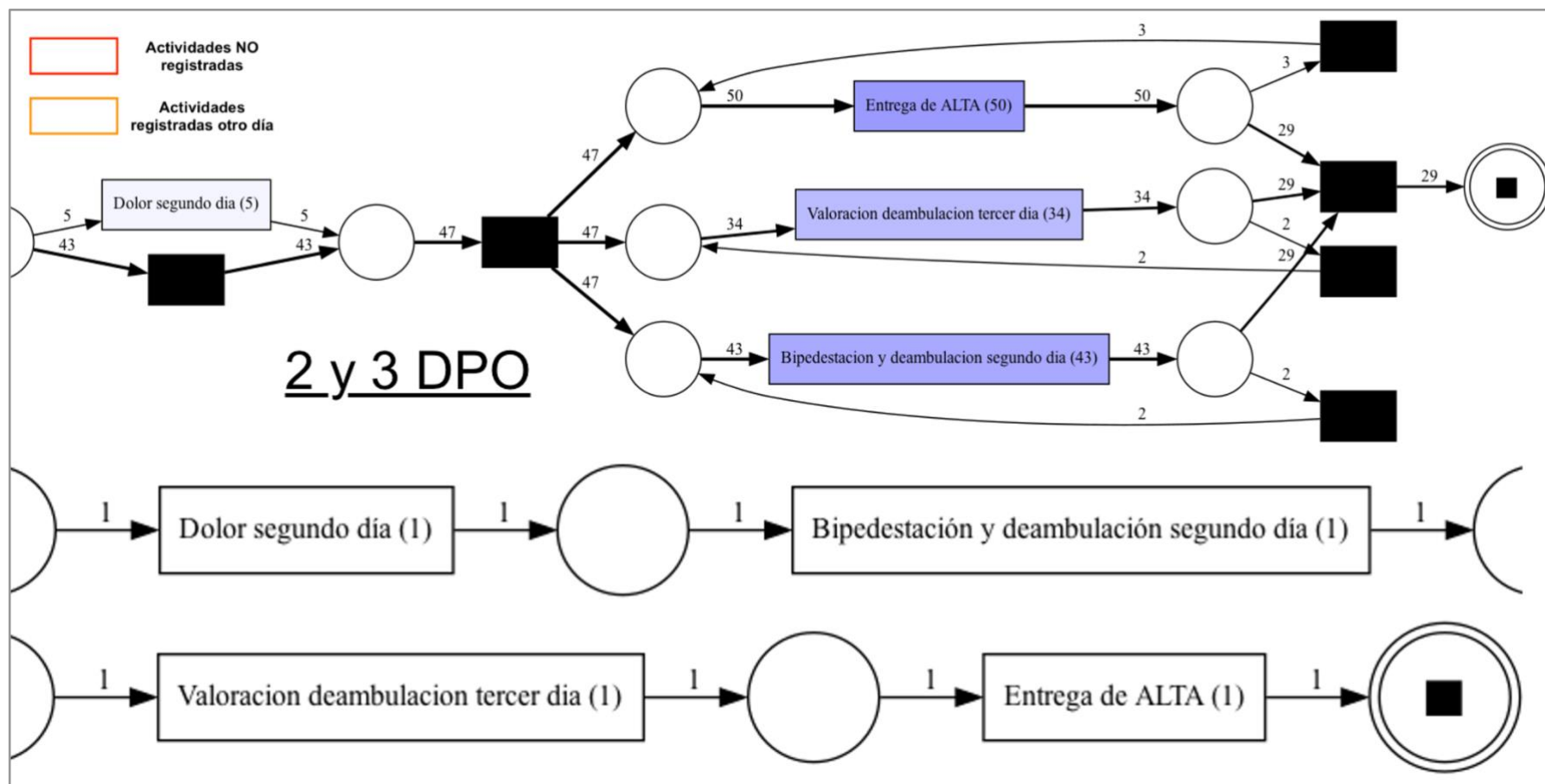


Figura 40: Comparación entre las redes de Petri del seguimiento de pacientes y la vía clínica(ATC)

Anexo 5. Resultados del análisis clásico de datos

En este anexo se plasman todos los resultados obtenidos en el análisis clásico de datos. Contiene modelos de regresión lineal y logística, estudio de la relación de parámetros de pacientes y gráficos estadísticos para mayor información sobre el proceso interno del hospital y sus pacientes.

Durante el estudio sobre la relación entre las variables continuas del registro de pacientes se han extraído los siguientes resultados:

Pareja de variables (x-y)	Covarianza	Coefficiente de Pearson
Días ingresado-HTO_POST	-0,2433	-0,0834
Edad-HB_POST	-1.6415	-0.1261
Días ingresado-HB_POST	-0,0549	-0,0550
Edad-HTO_POST	-8,5087	-0.2239
Edad-días ingresado	0,2523	0,0268
HTO_PRE-HTO_POST	6,3631	0,7023
HBE_PRE-HBE_POST	0,6965	0,7437

Tabla 2: Medidas de relación entre variables continuas

Con este estudio se comprueba que tan solo las variables de HTO_PRE-HTO_POST y HBE_PRE-HBE_POST tienen coeficientes de Pearson $>0,7$ (elevados), es decir, son las únicas que se puede considerar que estén lo suficientemente relacionadas. Es un resultado lógico, ya que son variables de parámetros de la sangre previos y posteriores a una operación. Lo interesante de este análisis de datos es que ninguna de las variables estudiadas (días ingresado, edad) tienen coeficientes de Pearson con los parámetros HB (hemoglobina) y HTO (hematocrito) como para afirmar que pueden tener algún efecto o relación con ellos.

Se muestran a continuación la representación gráfica de los modelos de regresión lineal, de los cuales se consiguen coeficientes de determinación de 0.49 y 0.55, respectivamente, para demostrar cómo explican los modelos de regresión lineal las parejas de variables con indicadores de correlación más altos:

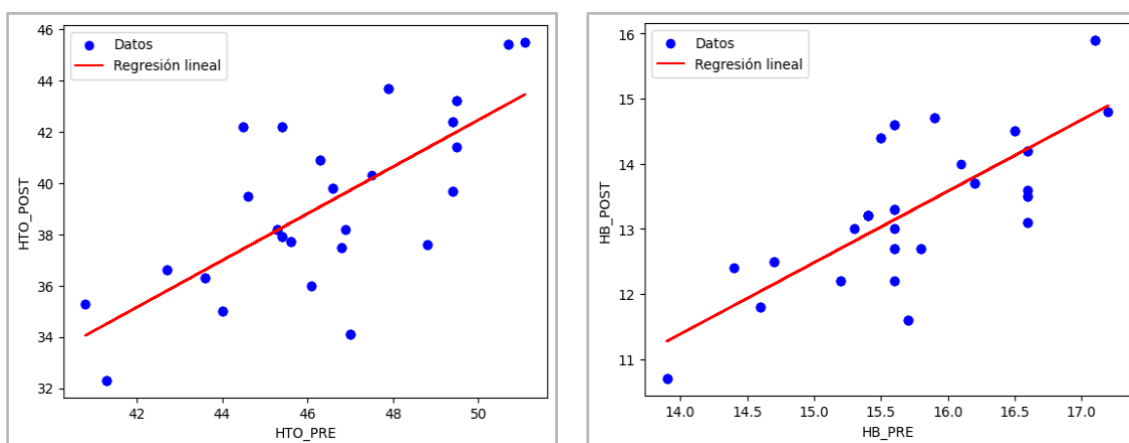


Figura 41: Modelos de regresión de Hematocrito (a) y Hemoglobina (b)

También se ha realizado un modelo de regresión logística para intentar predecir si un paciente proviene del entorno rural o de Zaragoza, en función de sus valores de hemoglobina previos y posteriores a la intervención quirúrgica. En este caso se ha

utilizado un set de 23 conjuntos de datos de pacientes para entrenar el modelo y 10 registros para comprobar su funcionamiento e intentar predecir de dónde provienen esos 10 pacientes. En la siguiente *Figura 42* se muestra la matriz de confusión del modelo. En ella se aprecia cómo el modelo ha predicho bien tan solo un caso de negativo verdadero y 4 casos positivos verdaderos, mientras que ha predicho mal otros 5 valores, por lo que con los datos tomados, este modelo solo ha predicho bien la procedencia de los pacientes según sus valores de hemoglobina en el 50% de las ocasiones. Esto indica que los valores de hemoglobina del paciente no son suficientes como para predecir su procedencia.

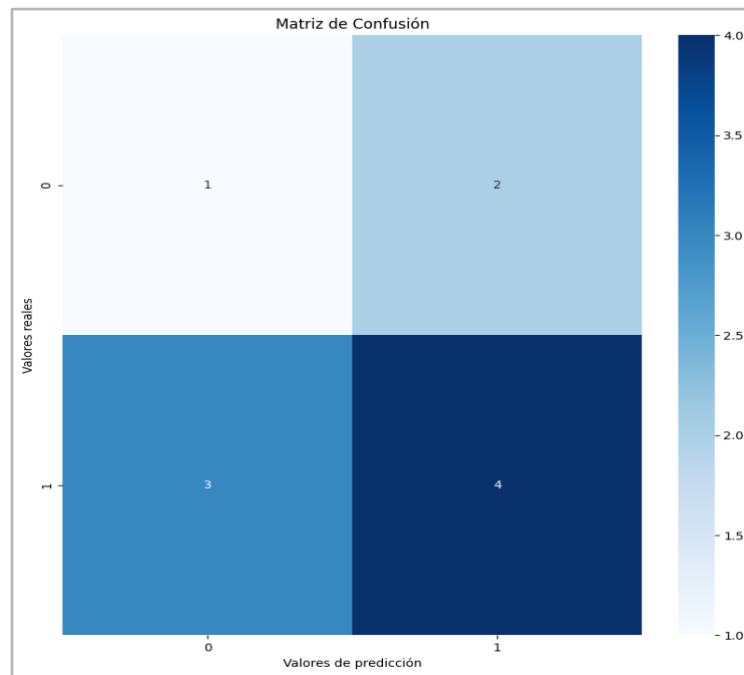


Figura 42: Matriz de confusión del modelo de regresión

Como última parte del análisis de datos se han utilizado gráficos de caja y de conteo a modo de análisis estadístico entre variables continuas y categóricas.

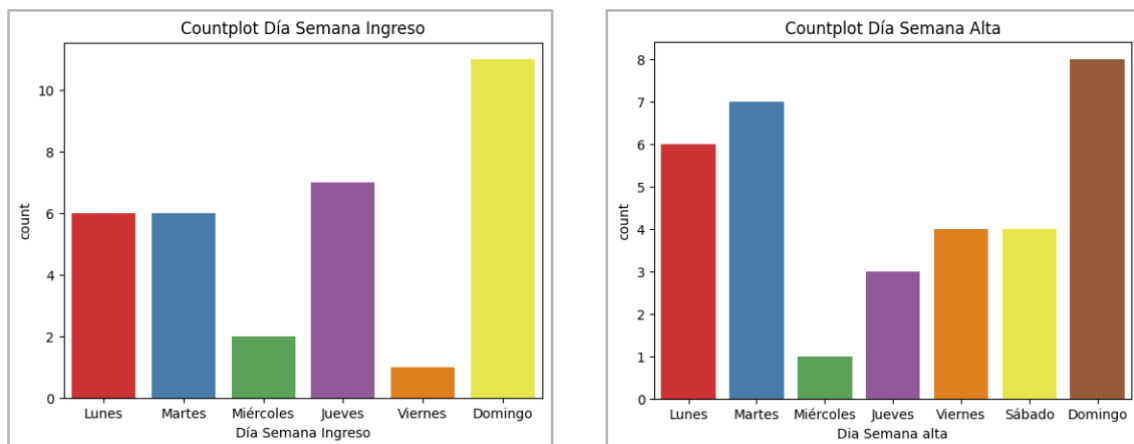


Figura 43: Gráficos de conteo de días de ingreso (a) y días de alta (b)

Se puede observar que hay una tendencia tanto de ingresar en el hospital como de recibir el alta hospitalaria, los 3 días entre los domingos y martes, siendo domingo el día que más ingresos y altas hay a la semana.

En la siguiente *Figura 44*, aparecen los días de la semana en los que se han llevado a cabo las operaciones de los pacientes del registro. Se observa que hay más operaciones los viernes, posiblemente para evitar operar en fin de semana, pero la diferencia de operaciones no es significativa con el resto de los días. Lo que sí que es reseñable es que las operaciones no se realizan en fin de semana, puesto que los datos registrados pertenecen a pacientes con cirugías programadas.

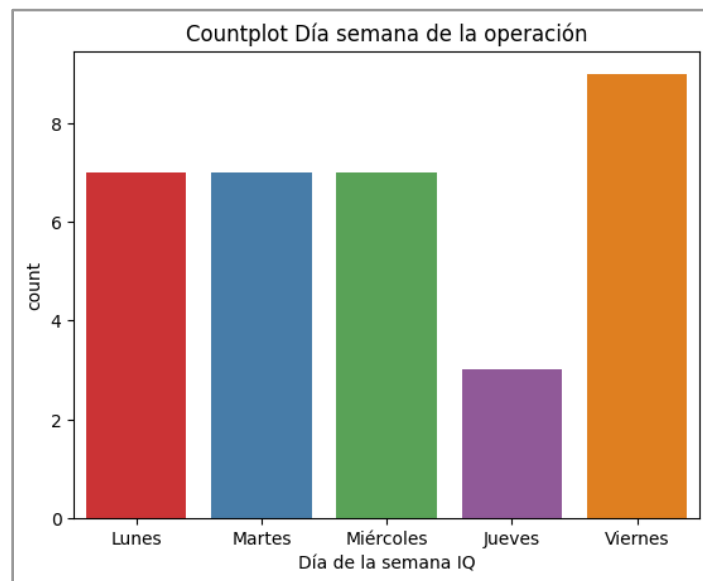


Figura 44: Gráfico de conteo de los días de la semana en los que se realizan las operaciones

En los siguientes gráficos de cajas se agrupan los pacientes según su edad y de dónde proviene cada uno, si de Zaragoza o el entorno rural (*Figura 45.a*); y según su edad y el motivo de su intervención (*Figura 45.b*).

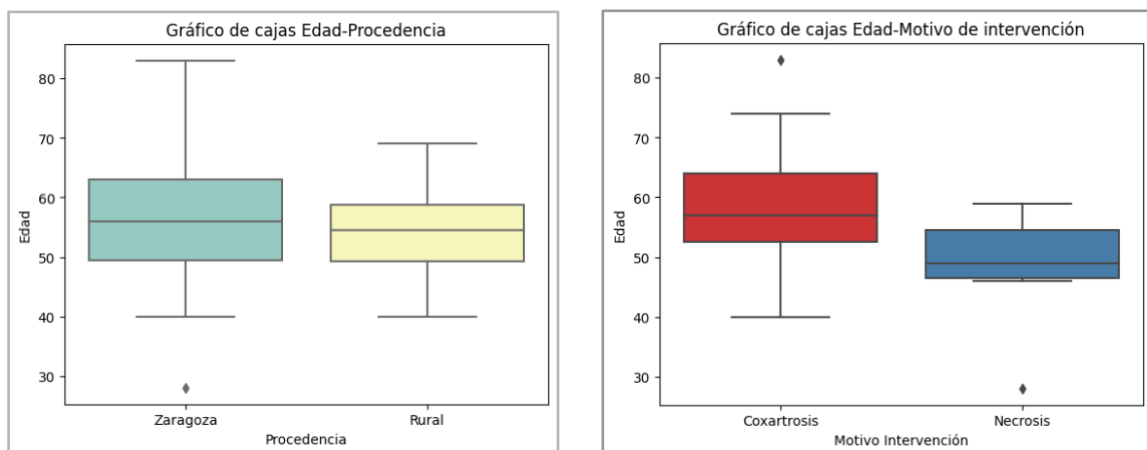


Figura 45: Gráficos de cajas. Edad-Procendencia (a) y Edad-Motivo de intervención(b)

Con respecto a la primera figura, se aprecia que la mediana de edad en ambos grupos es en torno a 55 años, pero que hay pacientes de mayor edad que provienen de Zaragoza, ya que el gráfico de Zaragoza tiene un tercer cuartil por encima de los 60

años y unos bigotes que se extienden hasta máximo de 83 años, además de un valor atípico de 28 años. El gráfico de entorno rural en cambio, tiene un rango intercuartílico más fino, lo que implica que hay menor dispersión en la edad de pacientes de fuera de Zaragoza y no presenta valores atípicos, con edad mínima de 40 años y máxima de 69 años.

Por otro lado, en la *Figura 45.b* se comprueba que hay más casos de pacientes intervenidos por coxartrosis y que la mediana de edad se sitúa en los 56 años, mientras que en casos de necrosis, la mediana son 49 años. Para este segundo motivo de intervención, los pacientes tienen entre 45 y 60 años, con un caso atípico de 28 años. En los casos de necrosis, la edad de los pacientes va desde los 40 hasta los 75 años y hay un caso atípico de un paciente con 83 años.

Este último gráfico (*Figura 46*) muestra la distribución de la edad de los pacientes agrupados en días que han permanecido ingresados. La mayoría de ellos han permanecido ingresados de 4 a 6 días, a excepción de una persona de 58 años que estuvo 8 días y dos personas de 54 y 46 que lo hicieron 7 días. Los pacientes que permanecieron 6 días tienen una mediana superior al resto, de 65 años, con un máximo de edad de 83 y un mínimo de 40 años. Los bigotes de esta caja son los más alargados, lo que demuestra mayor variabilidad en las edades de pacientes que permanecen 6 días ingresados. En cuanto a pacientes que estuvieron 4 o 5 días en el hospital, para aquellos que permanecieron 4 días, hay mayor dispersión de años que para los que estuvieron 5 días, eso sí, resalta el valor atípico de 28 años y otro de 74.

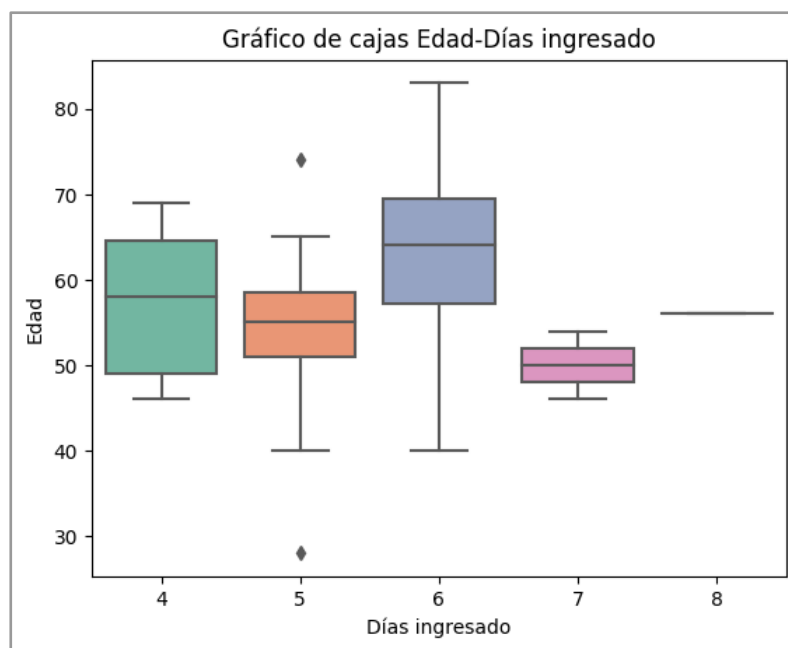


Figura 46: Gráfico de cajas. Edad-Días ingresado

Anexo 6. Gestión del proyecto

Se presenta la distribución de esfuerzos llevada a cabo durante el proyecto con una tabla donde se recogen las horas aproximadas de cada tarea y un diagrama de Gantt:

Tareas	Tiempo empleado (horas)
Trabajo de investigación sobre la minería de procesos	35
Presentación del problema y propuesta de soluciones al equipo médico	9
Reuniones de control en la universidad	18
Introducción a entornos de programación	13
Recogida y procesado de datos	10
Implementación de algoritmia y análisis de datos	46
Diseño de metodología para el seguimiento de pacientes	14
Análisis de resultados	11
Explicación y validación de resultados al equipo médico	4
Memoria	70
Total	230

TAREA	FECHA COMIENZO	FECHA FIN	DURACIÓN EN SEMANAS	Mes	Febrero				Marzo				Abril				Mayo				Junio			
				Semanas	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Trabajo de investigación sobre la minería de procesos	4	9	5																					
Presentación del problema y propuesta de soluciones al equipo médico	7	9	2																					
Reuniones de control en la universidad	7	15	8																					
Introducción a entornos de programación	6	9	3																					
Recogida y procesado de datos	10	13	3																					
Implementación de algoritmia y análisis de datos	10	16	6																					
Diseño de metodología para el seguimiento de pacientes	11	16	3																					
Análisis de resultados	13	16	3																					
Explicación y validación de resultados al equipo médico	14	16	2																					
Memoria	13	17	4																					

Diagrama de Gantt