



Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

**ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA AGRONÓMICA Y
BIOCIENCIAS**

**NEKAZARITZAKO INGENIARITZAKO ETA BIOZIENTZIETAKO GOI
MAILAKO ESKOLA TEKNIKOA**

*Reducción de las listas de espera mediante minería de procesos y algoritmos de
predicción*

presentado por

Ander Aquerreta Montoro

aurkeztua

GRADO EN CIENCIA DE DATOS
GRADUA DATUEN ZIENTZIETAN

Junio, 2022 / 2022, *Ekaina*

Resumen

En este trabajo lo que se pretende es mejorar la eficiencia de los servicios médicos mediante el uso de minería de procesos y la minería de datos. Este trabajo se enmarca en el servicio de cirugía plástica del Servicio Navarro de Salud-Osasunbidea, aunque este trabajo puede ser extrapolable para los demás servicios médicos. El propósito del presente estudio consiste en examinar el proceso que los pacientes atraviesan al recibir atención en este servicio, empleando técnicas de minería de procesos y minería de datos para generar modelos de predicción. Los resultados nos muestran que el flujo de pacientes que hay actualmente no es el más adecuado ya que hay muchos pacientes que se les deriva al servicio desde atención primaria y que, sin embargo, después de la primera consulta de este servicio, no se les vuelve a citar posteriormente y, por tanto, no requieren necesariamente de este servicio. Si pudiéramos reducir este tipo de pacientes, se podrían reducir considerablemente las listas de espera.

Palabras clave

Cirugía plástica, consulta, contacto, ciencia de datos, minería de procesos, minería de datos.

Abstract

The aim of this work is to improve the efficiency of medical services using process mining and data mining. This work focuses on the plastic surgery service of the Navarro Health Service-Osasunbidea, although this work can be extrapolated to other medical services. The purpose of this study is to examine the process that patients go through when receiving care in this service, using process mining and data mining techniques to generate predictive models. The results show us that the flow of patients that currently exists is not the most adequate, since there are many patients who are referred to the service from primary care and who, however, after the first consultation with this service, are not returned to quote later and, therefore, do not necessarily require this service. If we could reduce this type of patients, the waiting lists could be considerably reduced.

Key Words

Plastic surgery, consultation, contact, data science, process mining, data mining.

Agradecimientos

En primer lugar, me gustaría agradecer a, mi director de prácticas, Sergio Santana por haber estado conmigo ayudándome desde que empezamos con este problema mientras aún estaba en prácticas hasta el día de hoy. También agradezco a toda la gente del pabellón G del Hospital Universitario de Navarra por todo el apoyo que me brindaron durante mi estancia allí, me sentí muy integrado y he aprendido mucho gracias a todos.

Por otro lado, quiero agradecer a Andrea San Martín, jefa de cirugía plástica, por toda la ayuda y las directrices dadas. Sin su visión tan cercana de la especialidad no habría sido posible conseguir tan buenos resultados.

En tercer lugar, agradezco a todas las personas de Bardena Milenia y NaIA, por todo el seguimiento continuo durante el trabajo, por las críticas constructivas y las soluciones alternativas dadas, desde que les presenté el proyecto hasta el final.

Además, también agradezco a Mikel Galar el apoyo durante la realización del TFG, por el seguimiento continuo en la redacción, y por los consejos durante el proceso.

Por último, pero no menos importante, me gustaría agradecer a toda mi familia y amigos, gracias a ellos soy quien soy y, sin su apoyo, todo el camino hasta aquí no habría sido lo mismo. De verdad, muchas gracias.

Índice

1.	Introducción.....	6
2.	Preliminares.....	8
2.1.	Minería de procesos.....	8
2.2.	Minería de datos.....	11
3.	Descripción del problema.....	13
3.1.	Contexto.....	13
3.2.	Tipos de consulta.....	14
3.2.1.	Primera consulta procedente de atención primaria (1P).....	15
3.2.2.	Primera consulta procedente de atención especializada (1E).....	15
3.2.3.	Interconsulta (I).....	16
3.2.4.	Primera consulta procedente de urgencias (1U).....	16
3.2.5.	Urgencia médica (U).....	17
3.2.6.	Consulta sucesiva (2).....	17
3.2.7.	Consulta no presencial (4).....	18
3.2.8.	Cirugía programada (QP).....	18
3.2.9.	Cirugía Urgente (QU).....	19
3.2.10.	Ingreso hospitalario (IH).....	19
3.2.11.	Alta hospitalaria (AH).....	20
3.3.	Análisis de los tipos de consulta.....	20
4.	Minería de procesos.....	22
4.1.	Preparación de los contactos.....	22
4.2.	Definición de los procesos.....	25
4.2.1.	Criterios de inicio.....	25
4.2.2.	Criterios de fin.....	25
4.2.3.	Excepciones.....	25
4.3.	Algoritmo para establecer los casos.....	26
4.4.	Visualización de los procesos.....	26
4.4.1.	Mapas de procesos.....	26
4.4.2.	Procesos frecuentes.....	30
5.	Minería de datos.....	31
5.1.	Elección de los datos.....	31
5.1.1.	Variables de entrada.....	32
5.1.2.	Variable de salida.....	32
5.2.	Preprocesamiento de datos.....	33
5.3.	Clasificación.....	35

6. Conclusiones.....	37
7. Bibliografía.....	37

1. Introducción

Las listas de espera es uno de los principales problemas a los que se enfrenta actualmente la sanidad pública. En octubre de 2022 se llegó a registrar en el Hospital Universitario de Navarra esperas de primera consulta de hasta 234 días en alguna especialidad médica, siendo de 70 días la media de todas las especialidades. Esto supone un aumento del 40% desde octubre de 2019 [1].

La aplicación de la ciencia de datos en el ámbito de la medicina es un tema reciente, pero su potencial para mejorar la eficiencia de los servicios hospitalarios es innegable. El uso de algoritmos de predicción y análisis estadístico de las consultas puede ser de gran ayuda en la optimización de los procesos médicos y, a consecuencia de esto, en la identificación temprana de enfermedades.

La ciencia de datos se puede definir como "un concepto para unificar estadística, análisis de datos, aprendizaje automático, y sus métodos relacionados, con el objetivo de comprender y analizar la información y los fenómenos reales" [2]. Sin embargo, esta no es la única definición que existe. De esta definición surge la figura del científico de datos, que se está volviendo cada vez más relevante en la sociedad debido a la enorme cantidad de datos. Un científico de datos es un profesional que combina conocimientos estadísticos con habilidades en el manejo de bases de datos, API (Application Programming Interface) y técnicas de minería de datos. Además, debe poseer habilidades de programación y estar capacitado para analizar y encontrar patrones y tendencias significativas en grandes conjuntos de datos [3].

La ciencia de datos ha demostrado ser una herramienta esencial en numerosas áreas, como la industria, el comercio, la educación y la investigación científica, y su aplicación en la medicina está cobrando cada vez más importancia. En este sentido, ya se han realizado estudios que han aplicado modelos de machine learning para la detección temprana de enfermedades en diferentes campos médicos como la psiquiatría [4]. En España también ha salido algún estudio donde se utiliza la teoría de grafos para definir la esquizofrenia[5]. Sin embargo, son aún escasos los artículos que se han centrado en la aplicación de la ciencia de datos para analizar las rutas de los pacientes dentro de cada servicio.

En este sentido podemos mencionar un estudio realizado en 2022 sobre la aplicación de la minería de datos en la gestión del desempeño de los hospitales públicos [6]. Los resultados de este estudio sugieren que la minería de datos puede ser una herramienta efectiva en la mejora de los servicios médicos.

En este contexto, podemos definir un caso médico o proceso como el camino que ha tenido el paciente desde su primera consulta hasta que se le da el alta por una misma enfermedad o problema. Dentro de cada caso puede haber uno o más contactos con el servicio. Los contactos lo vamos a definir como cualquier tipo de visita del paciente al servicio médico. Durante el trabajo utilizaremos indistintamente tipos de consulta o contactos para referirnos a esto mismo.

Dentro del Servicio Navarro de Salud y, más en concreto, en la base de datos Bardena Core (Base Análisis Resultados de Navarra) nos encontramos que gran parte de estos casos médicos no están correctamente identificados. Hay muchos contactos, pero sin nada que los identifique en un mismo caso. En este sentido, es común encontrar situaciones en las que un paciente ha tenido múltiples citas con la misma especialidad médica, pero por problemas de salud diferentes que, en este momento, no pueden ser diferenciados.

Esta falta de información dificulta significativamente el análisis y la interpretación de los datos, lo que puede llevar a decisiones inexactas y a errores en la atención al paciente. Por lo tanto, el objetivo de esta investigación es establecer un método para identificar el inicio y el final de los casos o procesos médicos dentro de la base de datos y aplicar técnicas de machine learning para optimizar el análisis de los datos y mejorar la calidad de la atención médica. El uso de dos potentes herramientas de la ciencia de datos, como son la minería de procesos y la minería de datos, puede resultar en una reducción significativa de las listas de espera del servicio de cirugía plástica. Todo esto siempre y cuando se apliquen modificaciones respecto los “problemas” detectados (cuellos de botella, retrasos, posibilidad de meter más recursos...).

Para lograr nuestro objetivo hemos dividido en varios subobjetivos:

- Establecer los criterios de inicio y fin de los casos de cirugía plástica. Para lograr este objetivo, hemos contado con la ayuda de Andrea San Martín, jefa del servicio de cirugía plástica en el Hospital Universitario de Navarra. Gracias a sus directrices hemos podido establecer unos criterios muy acordes con la realidad.
- Llevar a cabo la minería de procesos para la identificación y análisis de los procesos de cirugía plástica. Esta técnica nos permitirá visualizar los procesos y sus interacciones mediante el uso de grafos. Esta visualización nos ayudará a identificar posibles cuellos de botella y puntos de mejora en el proceso, lo que a su vez nos permitirá mejorar la eficiencia del servicio.

- Crear de un algoritmo de predicción que permita predecir el siguiente paso en el proceso de cada paciente a partir de datos genéricos de entrada. Esto nos permitirá anticiparnos a posibles problemas en el proceso y tomar medidas para reducir el tiempo de espera del paciente.

La implementación de este enfoque permitirá la identificación y seguimiento de los casos médicos, lo que mejorará significativamente la calidad de la atención médica y permitirá una gestión más efectiva de la base de datos. Además, este enfoque puede ser aplicado en otras áreas de la medicina y de la investigación clínica para mejorar el análisis y la interpretación de grandes conjuntos de datos médicos.

2. Preliminares

En esta sección vamos a explicar las herramientas que hemos utilizado para la realización de este trabajo.

2.1. Minería de procesos

La minería de procesos es una disciplina que utiliza técnicas de análisis de datos para descubrir, monitorear y mejorar los procesos en las organizaciones a partir del análisis de los datos generados, los cuales registran los eventos o actividades realizadas dentro de los procesos. El objetivo principal de la minería de procesos es proporcionar información precisa y en tiempo real sobre los procesos de negocio, lo que permite a las organizaciones identificar oportunidades de mejora, detectar desviaciones, predecir resultados y tomar decisiones basadas en hechos concretos.

La minería de procesos proporciona un nuevo medio para mejorar los procesos en una variedad de dominios de aplicación, con especial interés en la medicina, por la cantidad de procesos que hay[7]. Hay dos factores principales que impulsan esta nueva tecnología. Por un lado, cada vez se registran más eventos, lo que proporciona información detallada sobre la historia de los procesos. A pesar de la enorme cantidad de datos de eventos, la mayoría de las organizaciones diagnostican problemas basándose en la intuición en lugar de los hechos[8].

La minería de procesos es una disciplina emergente que proporciona conjuntos completos de herramientas para brindar ideas basadas en hechos y apoyar mejoras en los procesos. Esta nueva disciplina se basa en enfoques impulsados por modelos de procesos y minería de datos. Sin embargo, la minería de procesos no se limita al descubrimiento de procesos. Al acoplar estrechamente los datos de eventos y los modelos de procesos, es posible

verificar la conformidad, detectar desviaciones, predecir retrasos, respaldar la toma de decisiones y recomendar rediseños de procesos. La minería de procesos da vida a modelos de procesos estáticos y coloca los volúmenes de datos masivos de hoy en día en un contexto de procesos[8].

La minería de procesos tiene tres partes diferenciadas (Ilustración 1. Diagrama de las tres partes de la minería de procesos):

- **Process Discovery:** Definir y analizar los procesos. Encontrar cuellos de botella y posibles mejoras.
- **Conformance checking:** Enfrentar el proceso definido anteriormente con el registro de eventos generado.
- **Enhancement:** En esta parte se realizan realizar mejoras de los modelos sobre los problemas detectados. Dentro de estos análisis estaría ver las posibles interacciones entre los recursos (profesionales sanitarios de las diferentes especialidades, en el caso de que se tenga esta información y sea relevante, ver posibles embudos o personas que tienen más trabajo que el que pueden llevar a cabo).

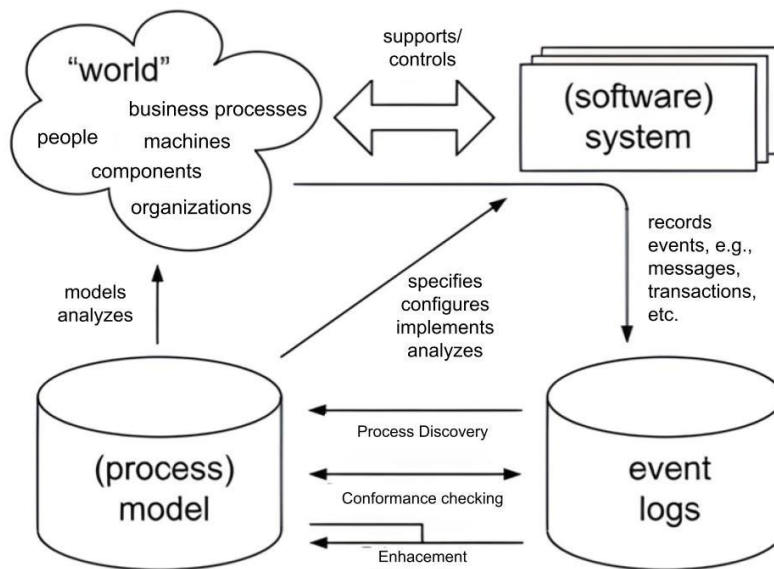


Ilustración 1. Diagrama de las tres partes de la minería de procesos.[8]

Para la minería de procesos hemos hecho uso de la librería bupaR de R. Concretamente utilizaremos lo que está definido como eventlog (registro de eventos). Para poder hacer uso de esta herramienta debemos tener definidas las siguientes cosas:

- **resource:** Puede ser un hospital, médico etc. El que realiza la actividad.
- **case_id:** el identificador del caso. Es el objeto que vamos a seguir a lo largo del proceso.
- **activity_id:** Esto es cada uno de los distintos puntos de contacto con el sistema.
- **lifecycle_id:** este es el estado del proceso. En nuestro caso hay 3 estados posibles, que son inicio, fin o en proceso.
- **Timestamp:** esta es la columna en la que están las fechas y horas de cada uno de los distintos eventos.

Con estos datos ya podemos crear nuestro objeto eventlog, con el que podemos visualizar los procesos en forma de grafos.

Un grafo es un conjunto de objetos llamados vértices o nodos unidos por enlaces llamados aristas o arcos, que permiten representar relaciones binarias entre elementos de un conjunto[9]. En otras palabras, un grafo es una representación visual de objetos y las relaciones entre ellos. Los nodos pueden representar objetos individuales, como puntos en un mapa o personas en una red social, y las aristas representan las conexiones entre ellos. En nuestro caso los nodos representan los tipos de contactos que puede tener un paciente.

Además, nuestros procesos los representan un tipo mas específico de grafos, los grafos dirigidos. Un gráfico dirigido (o simplemente dígrafo) consiste en un conjunto finito no vacío de elementos llamados vértices y un conjunto finito de pares ordenados de vértices distintos llamados arcos[10]. En un grafo dirigido, cada arco tiene una dirección y se representa por una flecha que apunta desde el vértice de origen hacia el vértice de destino. Como nuestros pacientes son dirigidos desde un tipo de contacto a otro, entonces tienen una dirección unidireccional. En la siguiente ilustración se puede observar una visualización sencilla de los procesos (Ilustración 2. Ejemplo de visualización de procesos mediante la librería bupaR.).

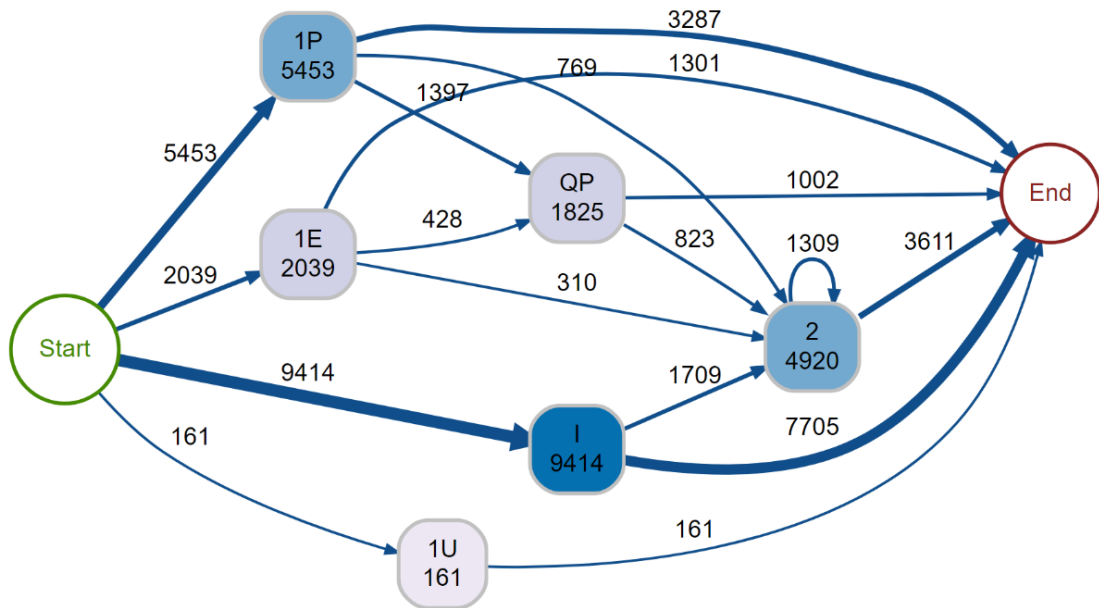


Ilustración 2. Ejemplo de visualización de procesos mediante la librería bupaR.

2.2. Minería de datos

Antes de explicar lo que es la minería de datos es necesario definir que es el KDD. La Extracción de Conocimiento en Bases de Datos (KDD, por sus siglas en inglés) es un análisis y modelado exploratorio automático de grandes repositorios de datos. KDD es el proceso organizado de identificar patrones válidos, novedosos, útiles y comprensibles de grandes conjuntos de datos complejos[11].

La Minería de Datos es el núcleo del proceso KDD (Ilustración 3. Proceso KDD[11].) y trata de resolver problemas mediante el análisis de datos ya presentes en las bases de datos[12]. Consiste en el uso de técnicas avanzadas de análisis de datos para identificar patrones, relaciones y tendencias en los datos, y luego transformar esta información en conocimiento accionable que pueda ayudar a la toma de decisiones en diferentes campos, desde la investigación científica hasta el marketing empresarial.

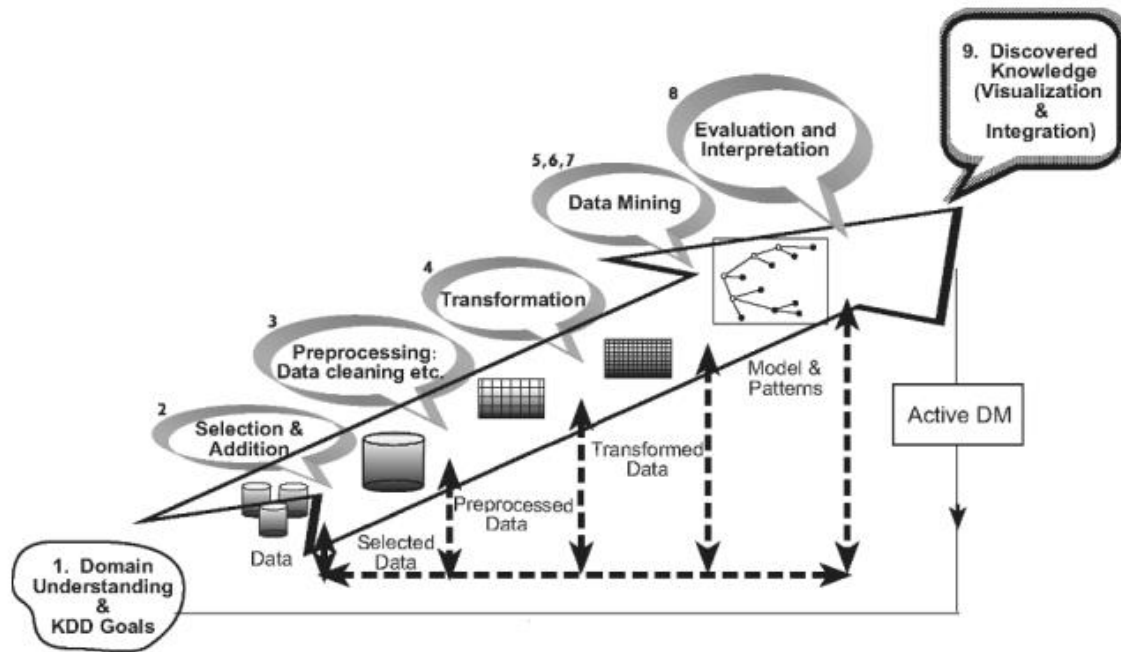


Ilustración 3. Proceso KDD[11].

La minería de datos implica el uso de diversas técnicas y herramientas de análisis de datos, incluyendo estadísticas, aprendizaje automático, visualización de datos y minería de texto. Los datos pueden provenir de una amplia variedad de fuentes, incluyendo bases de datos empresariales, redes sociales, sitios web, sensores y dispositivos de IoT.

Entre los principales objetivos de la minería de datos se encuentran la identificación de patrones ocultos, la segmentación de datos en grupos o categorías, la predicción de resultados futuros y la identificación de anomalías o excepciones en los datos. Estos conocimientos pueden aplicarse en diferentes campos, como la identificación de patrones de fraude en transacciones financieras, la segmentación de audiencias en marketing digital, la predicción temprana de enfermedades o, como en nuestro caso, para mejorar la eficiencia de los servicios médicos.

Para la parte de minería de datos vamos a hacer uso del lenguaje de programación Python. Python es una herramienta de programación popular y ampliamente utilizada en la creación de algoritmos de predicción. Su sintaxis simple y legible hace que sea fácil para los programadores entender y escribir código, lo que reduce la complejidad de la tarea de creación de algoritmos de predicción.

Además, Python cuenta con una gran cantidad de bibliotecas de aprendizaje automático, como scikit-learn, TensorFlow y Keras, que ofrecen una amplia variedad de algoritmos de aprendizaje automático para la creación de modelos de predicción precisos.

La versatilidad de Python como lenguaje de programación permite su integración con otras herramientas y lenguajes de programación, lo que facilita la creación de sistemas complejos de predicción. En resumen, Python es una herramienta poderosa y versátil para la creación de algoritmos de predicción, lo que lo convierte en una opción popular en el campo de la ciencia de datos.

En nuestro caso, al tener datos tabulares, utilizaremos la librería sklearn para los métodos de aprendizaje automático y para el preprocesamiento de datos y, para la preparación de los datos, utilizaremos las librerías pandas y numpy.

Por último, para hacerlo más visual haremos diversos gráficos con las librerías matplotlib y seaborn.

3. Descripción del problema

3.1. Contexto

Durante la realización de las prácticas curriculares en el Hospital Universitario de Navarra se propusieron tres retos para realizar durante las mismas. Dos de ellos eran más sencillos, uno de clasificación binaria con datos tabulares y otro de análisis de sentimiento con texto. El último y más importante reto era este, el poder mejorar la eficiencia de los servicios médicos. Al principio no teníamos claro cómo hacerlo ni qué datos utilizar, pero tras analizar nuestras posibilidades y viendo una presentación de Ibai Tamayo (de Navarrabiomed) sobre minería de procesos aplicada a los casos de diabetes nos percatamos de que sería interesante hacer lo mismo, pero para los servicios médicos. Además, nos decidimos centrar en el servicio de cirugía plástica, ya que, a priori, era el más sencillo y, si funcionaba para este servicio, podría ser extrapolable para los demás.

Para la realización del trabajo hemos contado con la ayuda del DGTD (Servicio de Tecnologías de Salud), que nos han podido proporcionar los datos necesarios para poder hacer ambas partes, tanto la minería de procesos como la minería de datos. Por último, pero no menos importante, hemos contado con la indispensable ayuda de Andrea San Martín (jefa del servicio de cirugía plástica), que nos ha ayudado verificando los resultados del trabajo y aportando importante información del servicio de cirugía plástica.

La Cirugía Plástica es una especialidad quirúrgica que se ocupa de la corrección de todo proceso congénito, adquirido, tumoral o simplemente involutivo, que requiera reparación o reposición, o que afecte a la forma y/o función corporal[13]. Esta especialidad abarca tanto la cirugía estética como la cirugía reconstructiva.

La cirugía estética tiene como objetivo mejorar la apariencia física de una persona y se realiza por razones personales, estéticas o de autoestima[14]. Los procedimientos más comunes en la cirugía estética incluyen la rinoplastia (cirugía de la nariz), la liposucción, el aumento de senos, la reducción de senos, la cirugía de párpados y la abdominoplastia (cirugía del abdomen).

La cirugía reconstructiva, por otro lado, se enfoca en corregir defectos físicos, deformidades o lesiones que afectan la función y/o la apariencia física de una persona[15]. Algunos ejemplos de cirugía reconstructiva incluyen la reconstrucción mamaria después de una mastectomía, la corrección de deformidades congénitas, la reparación de heridas faciales y la reconstrucción de los tejidos después de una lesión o enfermedad.

En general, la cirugía plástica busca mejorar la calidad de vida de las personas a través de procedimientos seguros y efectivos que pueden tener un impacto significativo en la apariencia física y la autoestima.

Una de las principales preocupaciones en la cirugía plástica es el tiempo que los pacientes deben esperar antes de ser sometidos a una intervención quirúrgica. En muchos países, las listas de espera para la cirugía plástica pueden ser bastante largas, lo que puede tener un impacto negativo en la calidad de vida de los pacientes y aumentar el riesgo de complicaciones médicas.

Hay varias razones por las que las listas de espera para la cirugía plástica pueden ser largas. Una de ellas es la escasez de cirujanos plásticos calificados. Otro factor que puede contribuir a las listas de espera es la mala gestión de estas. Además, la pandemia de coronavirus ha aumentado aún más el tiempo de espera para las cirugías electivas, incluyendo la cirugía plástica.

Para abordar esta necesidad de mejorar las listas de espera de cirugía plástica, se han implementado varias soluciones en diferentes países. Algunas de estas soluciones incluyen aumentar el número de cirujanos plásticos capacitados, reducir los costos de los procedimientos, mejorar la eficiencia de los hospitales y clínicas, y utilizar tecnología innovadora para agilizar el proceso de registro y programación de cirugías. La solución que nosotros planteamos es utilizar la minería de procesos y la minería de datos para lograr este fin.

3.2. Tipos de consulta

Cirugía plástica es una especialidad relativamente simple en cuanto a que no tiene demasiados tipos de consultas diferentes. A diferencia de otras especialidades médicas, en cirugía plástica no se realizan pruebas como radiografías, cosa que facilita el análisis.

Para poder llevar a cabo la minería de procesos es necesario conocer los tipos de consultas que se dan en los servicios médicos que, en este caso, por simplicidad, únicamente se consideran las que ocurren en el servicio de cirugía plástica.

3.2.1. Primera consulta procedente de atención primaria (1P)

Una primera consulta procedente de primaria se refiere a la primera visita de un paciente a un especialista médico, que es remitido por su médico de atención primaria o médico de cabecera. La atención primaria es la asistencia sanitaria esencial basada en métodos y tecnologías prácticos, científicamente fundados y socialmente aceptables, puesta al alcance de todos los individuos y familias de la comunidad mediante su plena participación y a un costo que la comunidad y el país puedan soportar, en todas y cada una de las etapas de su desarrollo con un espíritu de autorresponsabilidad y autodeterminación[16].

Esta consulta tiene como objetivo que el especialista evalúe la condición del paciente y determine si se necesita tratamiento adicional o si se puede manejar con cuidados en atención primaria. Durante esta consulta, el cirujano plástico realizará un examen físico detallado de la zona que se va a tratar, e indicará si es necesario intervención quirúrgica o es suficiente con los cuidados indicados en atención primaria.

3.2.2. Primera consulta procedente de atención especializada (1E)

Una primera consulta procedente de atención especializada se refiere a la primera visita de un paciente a un centro médico que proporciona servicios de atención médica especializada para una condición específica que necesita ser tratada por cirugía plástica.

Durante esta consulta, el paciente es evaluado por un especialista o equipo de especialistas en cirugía plástica en la condición médica específica que presenta, y se lleva a cabo una evaluación completa de la condición del paciente, revisión de su historial médico, y un examen físico. Es posible que se necesiten pruebas diagnósticas adicionales para determinar un diagnóstico definitivo o para ayudar en el desarrollo de un plan de tratamiento.

El objetivo principal de esta consulta es determinar el mejor plan de tratamiento para el paciente y proporcionar atención médica especializada y específica para su condición. Dependiendo de la complejidad de la afección del paciente, pueden ser necesarias varias visitas de seguimiento para garantizar que el plan de tratamiento sea efectivo y se ajuste a las necesidades del paciente.

En resumen, la primera consulta procedente de atención especializada es una visita importante para los pacientes que necesitan atención médica especializada, ya que les permite ser evaluados por expertos en su condición y recibir tratamiento específico y personalizado para su afección médica.

3.2.3. Interconsulta (I)

Una interconsulta médica se refiere a la solicitud de un médico especialista para que otro especialista o equipo médico (en este caso de cirugía plástica) evalúe y brinde recomendaciones o tratamientos adicionales para un paciente.

En términos generales, una interconsulta se realiza cuando un médico necesita la opinión o experiencia de otro médico especialista para tratar a un paciente que presenta una condición que debería ser tratada en otra especialidad.

La interconsulta puede ser solicitada por el médico tratante del paciente, o por iniciativa del propio especialista al detectar una situación que requiere de la opinión o intervención de otro especialista. El especialista consultado evalúa al paciente y proporciona una opinión o recomendaciones, y puede trabajar en conjunto con el médico tratante para desarrollar un plan de tratamiento efectivo para el paciente.

En definitiva, la interconsulta médica es una herramienta importante para brindar atención médica de alta calidad y garantizar que los pacientes reciban el tratamiento adecuado y necesario.

3.2.4. Primera consulta procedente de urgencias (1U)

Una primera consulta procedente de urgencias se refiere a la primera visita de un paciente a un servicio de urgencias o de emergencias médicas debido a una afección médica que requiere atención inmediata, en este caso de cirugía plástica.

Durante esta consulta, el personal médico del servicio de urgencias evalúa la condición del paciente, toma sus signos vitales y revisa su historial médico para determinar el mejor curso de acción. Es posible que se realicen pruebas diagnósticas para determinar el diagnóstico o para guiar el tratamiento. Dependiendo de la gravedad de la afección del paciente, se puede proporcionar tratamiento inmediato o se puede derivar al paciente a una unidad de cuidados intensivos o a un hospital para recibir tratamiento adicional.

El objetivo principal de la primera consulta procedente de urgencias es brindar atención médica inmediata a pacientes con afecciones médicas graves, con el fin de estabilizar la afección y, en algunos casos, salvar la vida del paciente.

En conclusión, la primera consulta procedente de urgencias es una visita importante para los pacientes que requieren atención médica inmediata debido a una afección médica grave o emergencia. El objetivo principal es proporcionar tratamiento inmediato y estabilizar al paciente, y luego desarrollar un plan de atención a largo plazo si es necesario.

3.2.5. Urgencia médica (U)

Una urgencia médica es una situación en la que se requiere atención médica inmediata debido a una enfermedad o lesión. En comparación, una primera consulta procedente de urgencias es una visita a la sala de urgencias para una afección médica que no pone en peligro la vida del paciente pero que requiere atención médica inmediata.

La diferencia principal entre una urgencia médica y una primera consulta en urgencias es la gravedad de la afección médica y el nivel de urgencia para el tratamiento. Las urgencias médicas requieren atención médica inmediata y pueden poner en peligro la vida del paciente si no se tratan adecuadamente, mientras que las primeras consultas procedentes de urgencias son menos urgentes y no ponen en riesgo la vida del paciente de inmediato, por lo que se puede derivar a una primera consulta en vez de tratarlo en urgencias.

Sin embargo, las urgencias (dentro de la base de datos) no representan la realidad, ya que debería haber muchas más de las que hay. Esta falta de errores de debe a una mala recogida de datos del personal de urgencias.

3.2.6. Consulta sucesiva (2)

Una consulta sucesiva se refiere a una visita posterior de un paciente a un médico después de la primera consulta. Es una continuación del proceso de atención médica que comenzó en la primera consulta y generalmente se realiza para revisar la evolución del paciente y hacer ajustes en el plan de tratamiento si es necesario. También se va a considerar como consultas sucesivas las auto derivaciones. Las auto derivaciones son primeras consultas procedentes de la misma especialidad, por lo que se cuenta como sucesiva. Es decir, aunque sea una primera consulta, como procede también de cirugía plástica lo mas probable es que sea del mismo caso, por lo que las consideramos como consultas sucesivas.

Durante una consulta sucesiva, el médico revisará el historial médico del paciente, evaluará su condición actual y, si es necesario, realizará pruebas adicionales. Generalmente en estas consultas el médico revisará la zona tratada o reconstruida para saber si es necesario otro tipo de contacto dentro del servicio. En algunos casos, el médico

puede derivar al paciente a otro especialista o a un hospital para recibir tratamiento adicional.

Las consultas sucesivas son importantes para garantizar que los pacientes reciban atención médica continua y de alta calidad. Pueden ser programadas según sea necesario, dependiendo de la condición del paciente y del tipo de tratamiento que se esté recibiendo. En algunos casos, el médico puede recomendar consultas sucesivas regulares para controlar la evolución de una afección crónica o para monitorear la salud del paciente a largo plazo. En el Hospital Universitario de Navarra hay pacientes que han necesitado hasta más de 40 consultas sucesivas para revisar el mismo caso.

En resumen, la consulta sucesiva se refiere a una visita posterior del paciente a un médico después de la primera consulta, para revisar la evolución de la condición del paciente y hacer ajustes en el plan de tratamiento si es necesario. Las consultas sucesivas son importantes para garantizar una atención médica continua y de alta calidad.

3.2.7. Consulta no presencial (4)

Una consulta no presencial es una consulta médica que se realiza sin que el paciente y el médico estén en el mismo lugar físico, utilizando medios electrónicos para la comunicación. Esto puede incluir consultas por teléfono, videollamadas, correo electrónico, mensajes de texto o chat en línea.

Durante una consulta no presencial, el médico evalúa los síntomas del paciente y discute su historial médico y cualquier otra información relevante para hacer un diagnóstico y recomendaciones de tratamiento. También puede proporcionar orientación y asesoramiento médico, recetas y referencias a otros especialistas en caso de ser necesario.

Las consultas no presenciales son especialmente útiles para situaciones en las que el paciente no puede acudir al consultorio médico debido a la distancia geográfica, problemas de movilidad o restricciones de tiempo. También pueden ser útiles en situaciones de emergencia en las que se necesita una evaluación médica rápida antes de trasladar al paciente a un centro médico. No obstante, para el servicio médico de cirugía plástica no es muy práctico, ya que es una especialidad en la que es muy necesario realizar un examen físico y pruebas de diagnóstico para obtener un diagnóstico preciso y establecer un plan de tratamiento adecuado.

3.2.8. Cirugía programada (QP)

Una cirugía programada es una operación que se planifica con antelación y se lleva a cabo en una fecha determinada. La cirugía programada se puede realizar por varias razones,

como para tratar una afección crónica, para mejorar la calidad de vida del paciente, para diagnosticar una afección o para prevenir problemas de salud futuros.

Las cirugías se pueden clasificar en cirugías mayores y menores en función de la complejidad y duración del procedimiento, así como del nivel de anestesia y cuidados postoperatorios requeridos:

- **Cirugía mayor:** es una cirugía que implica un procedimiento invasivo y extenso, generalmente realizado bajo anestesia general, que puede requerir una hospitalización prolongada y un largo período de recuperación.
- **Cirugía menor:** es una cirugía que implica un procedimiento menos invasivo y de corta duración, que generalmente se realiza bajo anestesia local y puede ser ambulatoria, lo que significa que el paciente puede regresar a casa el mismo día. Ejemplos de cirugías menores incluyen una biopsia de piel, la extirpación de una verruga o una cirugía de cataratas.

La cirugía mayor se realiza en un quirófano grande, equipado especialmente para ello, mientras que la cirugía menor se suele realizar en pequeños quirófanos, donde al lado hay una pequeña consulta. Un procedimiento habitual en las cirugías menores (en el Hospital Universitario de Navarra) es el hacer una consulta sucesiva y, cuando el cirujano plástico lo ve necesario y la operación es de cirugía menor, se suele pasar a estos pequeños quirófanos a intervenir al paciente.

En definitiva, una cirugía programada es una operación planificada con antelación, mientras que la cirugía de emergencia se realiza de inmediato debido a una afección médica urgente. Las cirugías se pueden clasificar en cirugías mayores y menores en función de la complejidad y duración del procedimiento, así como del nivel de anestesia y cuidados postoperatorios requeridos.

3.2.9. Cirugía Urgente (QU)

Una cirugía urgente es una intervención quirúrgica que se realiza de manera inmediata o en un plazo muy corto de tiempo debido a una condición médica que requiere atención inmediata para salvar la vida o prevenir complicaciones graves. La cirugía urgente se realiza en casos de emergencia médica en los que el paciente presenta una afección que no puede esperar y necesita atención médica inmediatamente.

3.2.10. Ingreso hospitalario (IH)

Un ingreso hospitalario se refiere al proceso en el que un paciente es admitido en un hospital para recibir tratamiento médico. Durante el ingreso, el paciente es asignado a una habitación y es supervisado de cerca por el personal médico y de enfermería.

El ingreso hospitalario puede ser programado o urgente:

- Un **ingreso programado** es para pacientes ingresados con orden de ingreso programada, independientemente de si proceden de lista de espera, o son de un servicio que no tiene demora de ingreso[17].
- Un **ingreso urgente** se realiza para pacientes ingresados con orden de ingreso urgente, independientemente de que procedan del área de urgencias o no[17].

Durante el ingreso hospitalario, se pueden realizar varios procedimientos médicos, como pruebas diagnósticas, cirugías, administración de medicamentos y cuidados intensivos. El personal médico y de enfermería supervisa continuamente al paciente para asegurarse de que reciba la atención necesaria y para monitorear su condición de cerca.

3.2.11. Alta hospitalaria (AH)

Se entiende por alta hospitalaria a la salida de un enfermo previamente ingresado en el hospital, dejando de ocupar una cama de hospitalización en el centro, independientemente de la forma de salida (traslado a otro centro, alta voluntaria, fuga, etc.) o el estado del paciente, curación, mejoría, éxitus (es un término médico utilizado para referirse a la muerte de un paciente[18]), etc[17]. El alta se produce cuando el paciente ha alcanzado una cierta estabilidad en su estado de salud y se considera que ya no necesita atención médica continua en un entorno hospitalario.

Durante el proceso de alta, el equipo médico proporciona al paciente información sobre su diagnóstico, tratamiento y plan de cuidados a realizar después del alta. Esto puede incluir instrucciones sobre la medicación, cambios en la dieta, terapia física u ocupacional, seguimiento con un especialista, citas de seguimiento y cualquier otro cuidado continuo que pueda necesitar.

3.3. Análisis de los tipos de consulta

En la siguiente gráfica (Ilustración 4) podemos observar que tenemos los tipos de consulta (Tabla 1) desbalanceados, ya que la gran mayoría de los contactos son consultas sucesivas, lo cual es lógico, debido a que hay pacientes que requieren de revisión continua para su tratamiento. Podemos discernir que las urgencias (primera consulta procedente de urgencias y urgencias) y las consultas no presenciales son las menos frecuentes. Esto es

normal ya que cirugía plástica es una especialidad en la que es muy necesario hacer un diagnóstico preciso al paciente y, para ello, la mayoría de las veces es necesario palpar la zona afectada.

Abreviatura	Significado
1P	Primera consulta procedente de primaria
2	Consulta Sucesiva
I	Interconsulta
1E	Primera procedente de especializada
IH	Ingreso hospitalario
QP	Cirugía programada
AH	Alta hospitalaria
1U	Primera procedente de urgencias
QU	Cirugía urgente

Tabla 1. Tabla de abreviaturas

Cabe destacar que, como se observa en la segunda gráfica (Ilustración 5), estas consultas se realizan mayoritariamente los días de entre semana, que es cuando se realizan las consultas. En los fines de semana los contactos que más suele haber son urgencias, ya que no se suelen programar citas para días no lectivos.

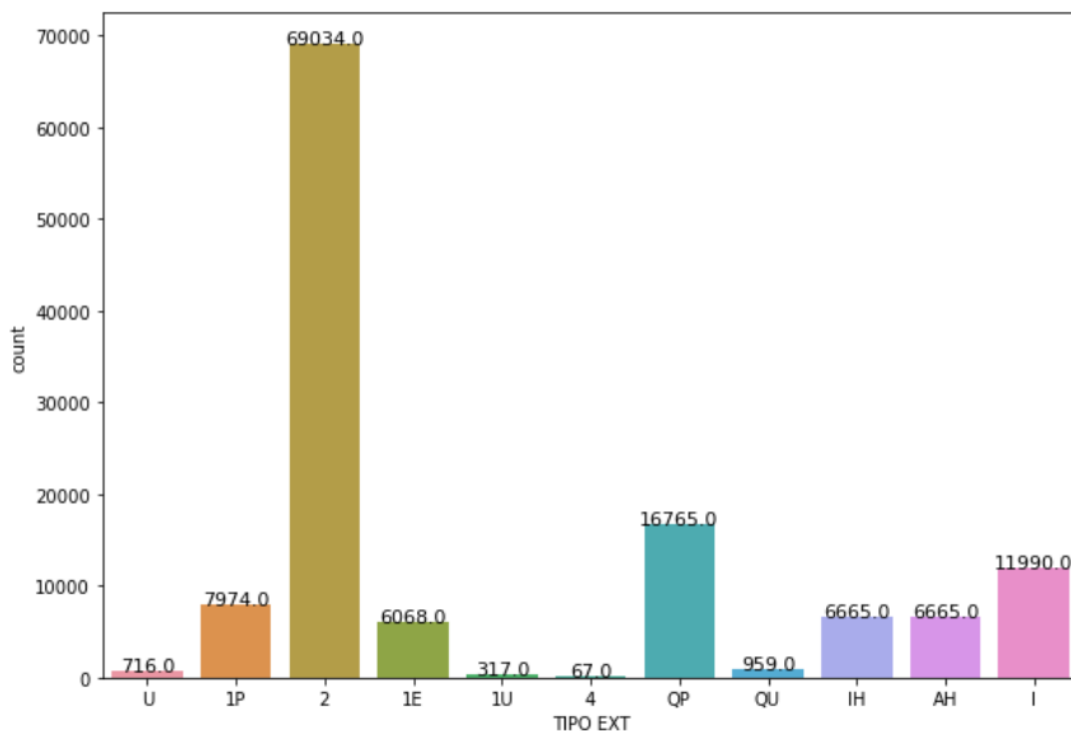


Ilustración 4. Gráfico de barras de los contactos (ver ;Error! No se encuentra el origen de la referencia.)

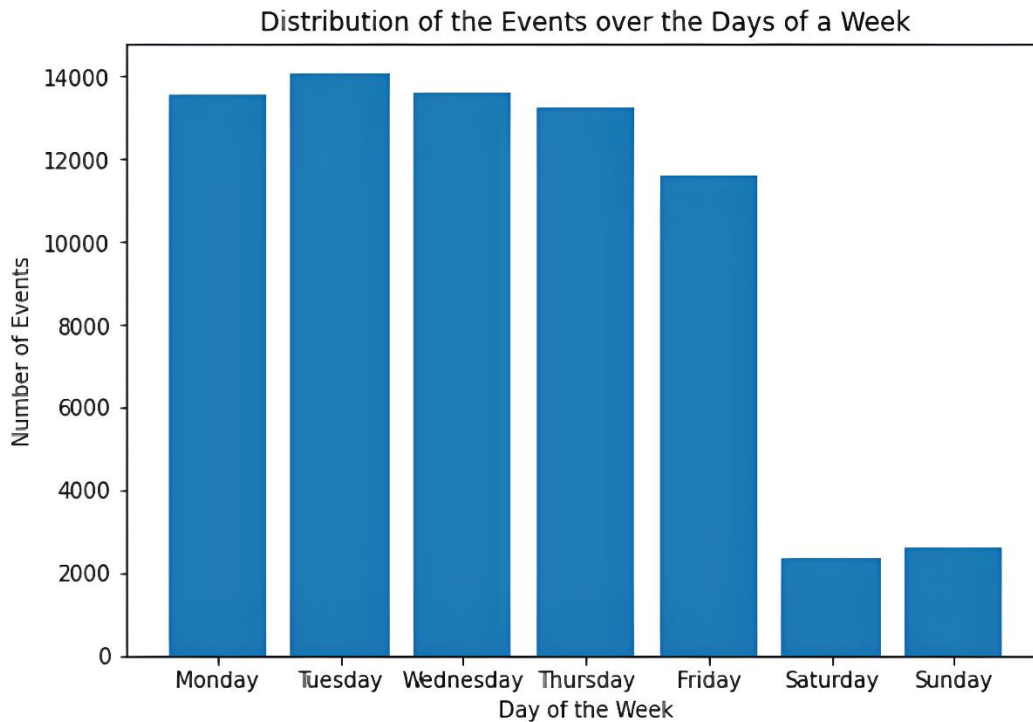


Ilustración 5. Número de comienzo de casos por día de la semana

4. Minería de procesos

En esta sección vamos a realizar la minería de procesos a nuestros contactos para organizarlos por casos. Vamos a, según una serie de contactos y unas reglas, definir los procesos de los pacientes. Los procesos son lo que comúnmente conocemos como caso médico. Es decir, cada paciente tiene varios contactos con el servicio de cirugía plástica, pero estos no tienen por qué pertenecer al mismo proceso.

4.1. Preparación de los contactos

Para llevar a cabo con éxito técnicas de minería de procesos es fundamental contar con un equipo multidisciplinar, que incluya expertos en el proceso en cuestión y especialistas en ciencia de datos, informática y matemáticas capaces de dominar las técnicas necesarias para implementar la minería de procesos de manera efectiva. La combinación de habilidades y conocimientos de diferentes áreas es esencial para lograr una comprensión profunda y exhaustiva del proceso, y así poder aplicar las herramientas adecuadas para su análisis y optimización. En resumen, una colaboración interdisciplinaria puede generar resultados más precisos y significativos en la minería de procesos para representar nuestra información.

La información que utilizaremos se llama registro de eventos y se trata de información almacenada en el sistema. Cada registro se compone de un caso, una actividad y una duración o fecha de inicio [19]. Como podemos observar en la Tabla 2 no tenemos un identificador de caso, ya que para obtenerlo vamos a tener que establecer una serie de reglas. Antes de nada, tenemos que ver qué datos tenemos por cada tipo de evento (contacto) en nuestro dataset. Vamos a mostrar algún contacto en la siguiente Tabla 2.

Pac Unif Cod	Fecha	TIPO EXT	SitAten Nom
4	17/11/2015	1P	Desconocid@
357	19/11/2012	2	ALTA
91	06/08/2016	I	
357	03/02/2012	QP	
383891	15/02/2012	IH	
3203	25/07/2015	U	
5202	30/04/2013	1U	ALTA
5206	26/07/2016	1E	Desconocid@

Tabla 2. Datos de contactos

Como podemos ver tenemos un total de 4 variables, que son las siguientes:

- **Pac Unif Cod:** Es un número que identifica unívocamente a cada paciente de manera anónima.
- **Fecha:** Fecha en la que se realizó ese contacto.
- **TIPO EXT:** Tipo de contacto realizado.
- **SitAten Nom:** Es una variable que solamente tenemos en las citas (primera consulta procedente de primaria, primera consulta procedente de especializada, primera consulta procedente de urgencias, consultas sucesivas y consultas no presenciales) y que indica si el paciente fue dado de alta o no. Como para las demás no tenemos esta información asumiremos que no fue dado de alta, lo cual tiene sentido, ya que la mayoría de las veces se les da el alta en consulta.

El registro de estos contactos puede tener varios problemas. El primero es el ruido, ya que hay algunos contactos que son muy poco frecuentes, como las consultas no presenciales y las urgencias. El segundo es la incompletitud de los datos, ya que puede haber contactos con el servicio entre medias de un caso que no está recogido. Por ejemplo, en nuestros datos ocurre que hay algún alta hospitalaria sin un ingreso previo.

Para solventar el primer problema vamos a eliminar las consultas más irrelevantes. Las consultas no presenciales las eliminamos al ser una práctica poco habitual e irrelevante

para una especialidad como cirugía plástica, donde es imprescindible el contacto directo con el paciente. Además, las urgencias también son eliminadas al estar mal representadas (debería haber muchas más de las que hay).

Para solventar el segundo problema hemos tenido que estudiar detenidamente los datos de los contactos para ver qué problemas tenía cada uno. Para solventar las altas sin ingreso previo hemos decidido solo considerar los ingresos y altas donde haya un ingreso y, en menos de un año, un alta para el mismo paciente.

El último inconveniente que encontramos fue que nuestros datos son de 10 años, por lo que hay que tener en cuenta los casos que ya han empezado en 2012 (año de nuestros primeros datos) y los que actualmente (abril de 2023) se encuentran en lista de espera. Los primeros datos no tenemos que arreglarlos, ya que el algoritmo (que se explicará más adelante) ya los elimina. Los casos que aún no han terminado los hemos arreglado eliminando los casos que, se encuentran en proceso (sin haber terminado), y que se encuentran en este momento en lista de espera de consulta o de cirugía. Estos datos son eliminados ya que nosotros queremos representar los procesos completos, por eso no vamos a tener en cuenta los incompletos.

Tras estos cambios pasamos de tener 127220 contactos a 82185 (Ilustración 6). Esto significa una reducción del 35% de los datos, algo necesario para tener datos que representen realmente los casos completos.

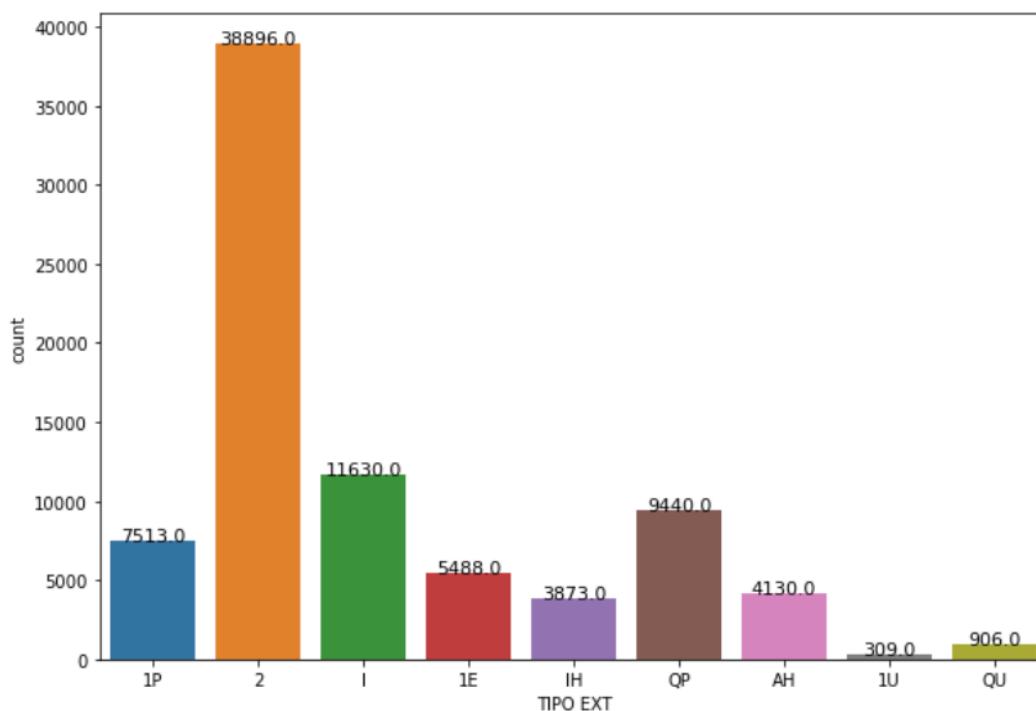


Ilustración 6. Gráfico de barras de los contactos tras la preparación (ver Tabla 1)

4.2. Definición de los procesos

Para poder realizar la minería de procesos hay que definir los procesos. Hay que definir cuando empieza un proceso y cuando acaba. Para poder definirlo consultamos a Andrea San Martín (jefa del servicio de cirugía plástica del HUN) para tener la opinión de una experta en los procesos de la cirugía plástica. Además, analizando los grafos resultantes tras establecer los inicios y los finales añadimos excepciones. Estas excepciones son casos en los que se concluiría el proceso del paciente, pero que realmente no termina ahí.

4.2.1. Criterios de inicio

Una de las partes principales de un proceso es, claramente, el inicio. Como inicio vamos a establecer varios tipos de contactos, de modo que, cuando se produzca ese tipo de consulta, se inicie un nuevo caso. Los contactos que hemos elegido como inicios son las primeras consultas (de atención primaria, especializada y de urgencias), la cirugía urgente y las interconsultas. Dentro de nuestro grafo estos nodos solo van a tener aristas que salen hacia otros nodos (que no sea ninguno de los inicios).

4.2.2. Criterios de fin

Los criterios de fin de caso son, posiblemente, la parte mas importante del proceso, ya que es lo que dictamina cuanto dura el proceso. Finalmente, tras analizar varias posibilidades, decidimos que un caso finaliza si ocurre una de estas tres situaciones:

- **Alta en consulta:** Como hemos visto antes, en las consultas hay una variable que indica que se le ha dado el alta o no, entonces si se le da el alta el caso queda cerrado.
- **Inicio de un nuevo caso:** Si ocurre una de las citas que hemos establecido como inicio de caso entonces este caso automáticamente queda cerrado.
- **Diferencia de un año entre consultas:** Si el paciente (que no se le ha dado el alta) lleva más de un año sin acudir al servicio de cirugía plástica entonces damos por concluido ese proceso.

4.2.3. Excepciones

Una vez definidos los criterios de inicio y fin de casos hicimos una primera visualización, pero pudimos ver que había procesos extraños, que no deberían terminar ahí. Por eso mismo fue que decidimos añadir 3 excepciones (las más frecuentes).

1. Había muchas veces que después de una interconsulta (inicio de caso) se producía, muy seguido, una primera consulta procedente de especializada para el mismo caso. Cuando esto ocurre este segundo contacto formará parte del proceso y no significará un inicio de caso. Por tanto, el nodo de primera consulta procedente de especializada tendrá una arista que llega a él.
2. Hay muchas veces que después de una cirugía programada se produce una interconsulta (para el postoperatorio). Entonces si hay una cirugía programada y, en menos de 3 días, se produce una interconsulta, esta última no significa inicio de caso.
3. En mitad de un ingreso hospitalario, se produce una cirugía urgente. A veces cuando un paciente está ingresado necesita urgentemente una cirugía, pero eso no significa que sea inicio de caso. Por tanto, si se produce una cirugía urgente en mitad de un ingreso no significará inicio de caso.

4.3. Algoritmo para establecer los casos

Para poder definir los casos hemos hecho un algoritmo que va a identificar los casos según nuestras reglas y a desechar los casos que no tengan inicio o que aún no hayan finalizado. Para ello primero hemos tenido que ordenar todos nuestros contactos por paciente y por fecha, de manera que, para cada paciente, estén ordenados los contactos por orden cronológico.

Una vez ordenados recorreremos por cada paciente todos sus contactos. Si es un inicio finalizamos el caso anterior y empezamos uno nuevo. Si tiene alta marcada o ha pasado un año o más desde su anterior contacto, entonces finalizamos el caso anterior y esperamos a un nuevo inicio de caso. Si, entre medias, hay algún contacto que no sea inicio se elimina, ya que no está dentro de ningún proceso.

4.4. Visualización de los procesos

Para la visualización, como he comentado anteriormente, utilizaremos la librería `bupaR` para crear los `evenlogs` (objeto con los procesos) y la librería `processmapR` para realizar los grafos con los procesos y demás gráficas.

4.4.1. Mapas de procesos

Los mapas de procesos son grafos en los que se puede ver de manera global todos los procesos. Los vértices de los grafos son los contactos y las aristas tienen un valor. Este puede ser tanto el número de pacientes (absoluto y relativo) que siguen ese camino en su proceso como el tiempo medio (o la mediana) que tarda un paciente en ir de un contacto a otro.

En el mapa de los procesos completos (Ilustración 7) se puede ver el principal cuello de botella que hay que resolver. Podemos ver como de 7266 procesos que comienzan con una primera consulta procedente de atención primaria 3287 se les da el alta en esa misma consulta. Esto supone más de un 35% de los pacientes que realmente no deberían haber acudido a la consulta, ya que inmediatamente se les da el alta. Esto se puede observar también para los demás inicios de caso.

En la Ilustración 8 e Ilustración 9 podemos ver representado de manera absoluta y relativa respectivamente como sería el mapa de procesos si solo usáramos el 80% de los procesos más frecuentes. En la Ilustración 10 podemos ver el mismo mapa de procesos que antes, pero esta vez el valor de las aristas no son personas, sino que es el tiempo medio de ese camino. Esto nos permite analizar correctamente el tiempo que se tarda entre dos consultas distintas dentro del mismo caso.

En la Ilustración 11 se puede ver esto mucho más claro, los 3 procesos más frecuentes son inicios de caso y directamente alta. Reducir esto puede suponer una significativa reducción en las listas de espera de consulta en esta especialidad.

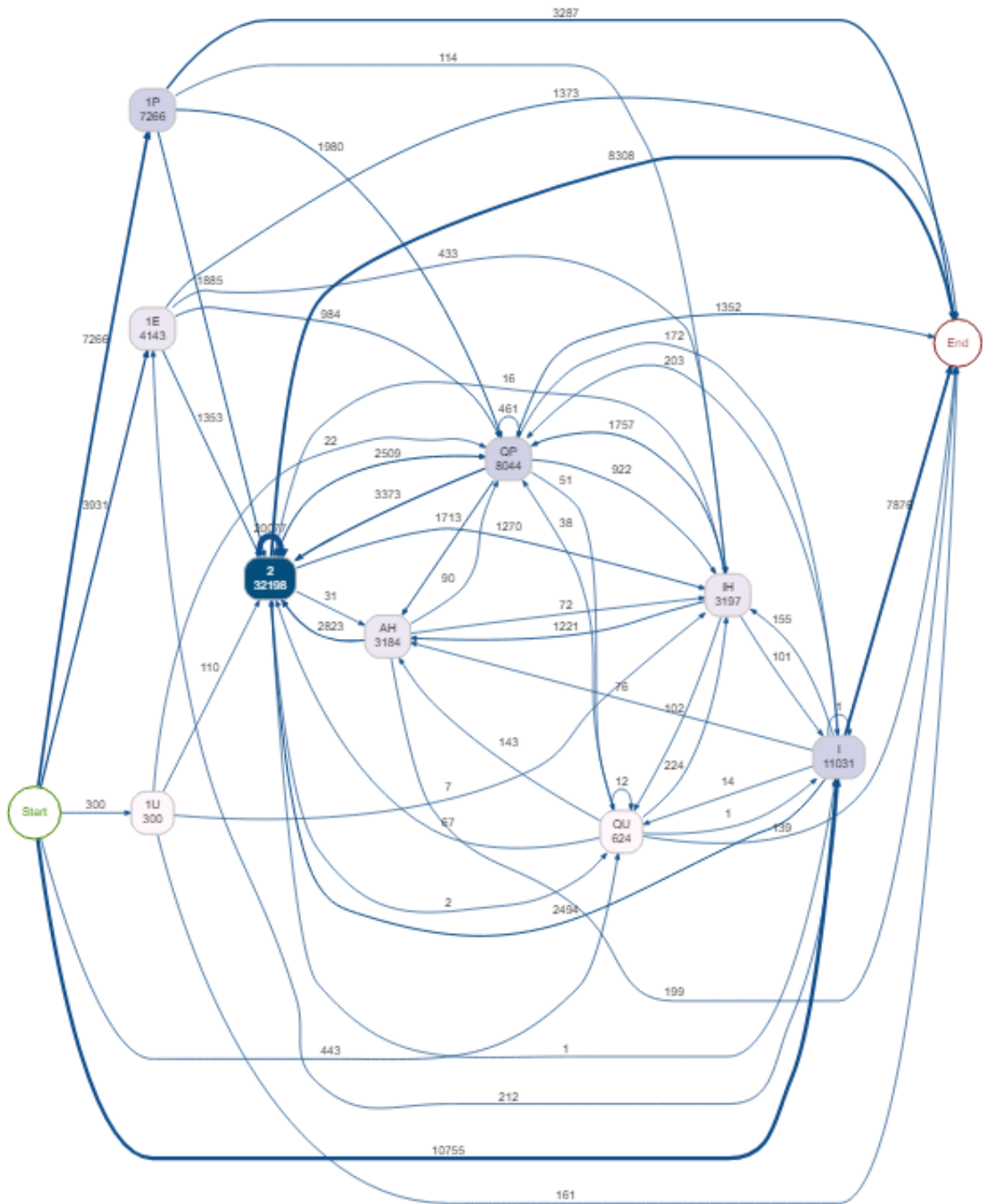


Ilustración 7. Mapa de procesos con el número de pacientes.

Grafo de los procesos donde los nodos son los tipos de contacto y el valor de las aristas es el número de casos que han seguido ese camino. Por ejemplo, el número de pacientes que, tras una consulta sucesiva, han ido a una cirugía programada son 2509.

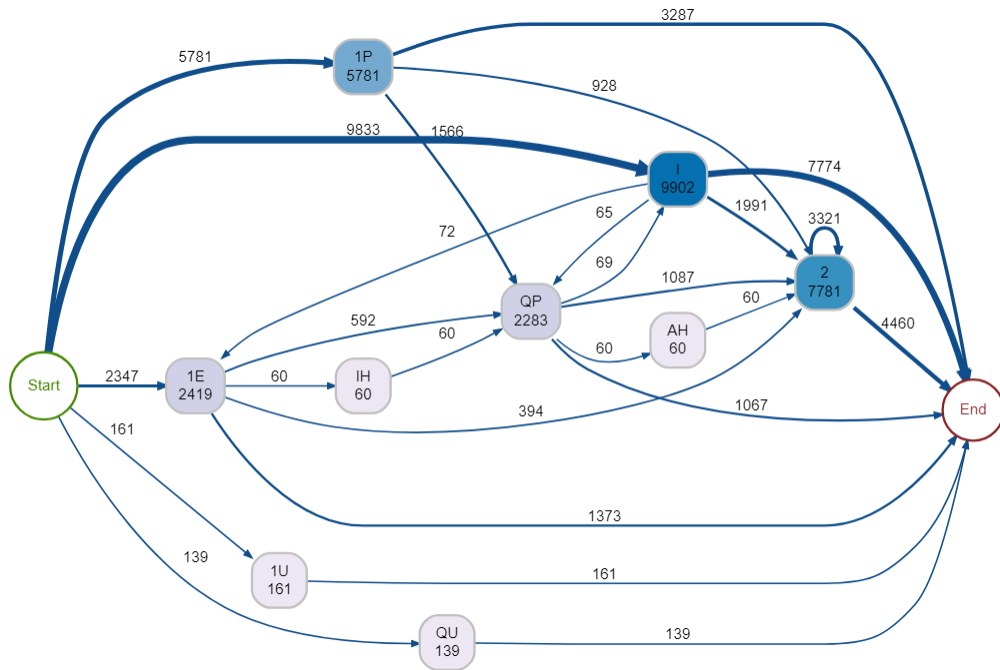


Ilustración 8. Mapa de procesos con el número de pacientes (absoluto) representando el 80% caminos más frecuentes.

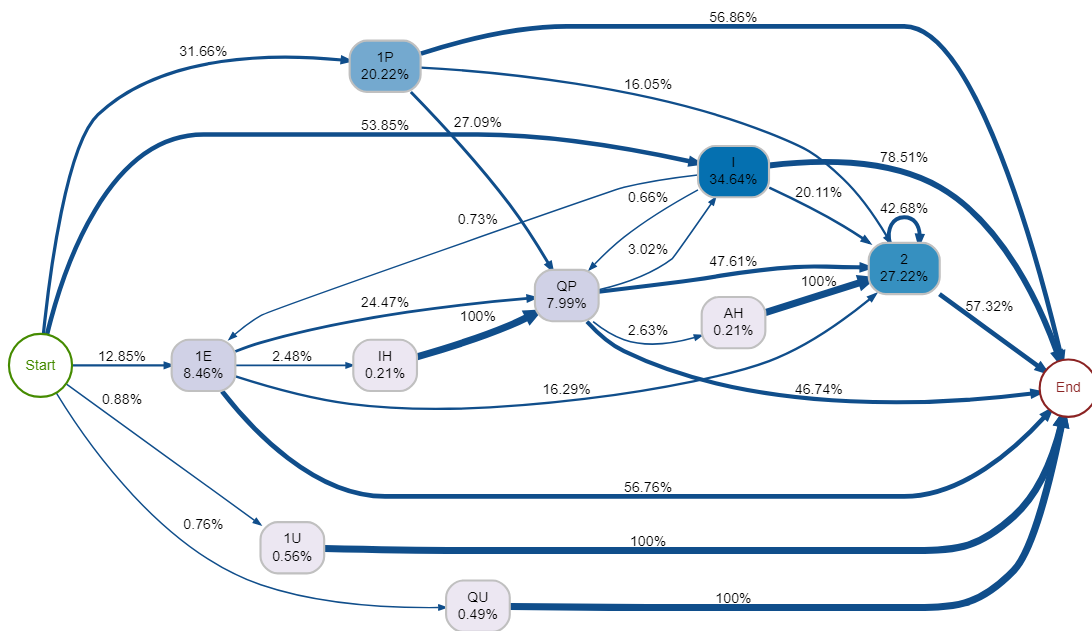


Ilustración 9. Mapa de procesos con el número de pacientes (relativo) representando el 80% caminos más frecuentes.

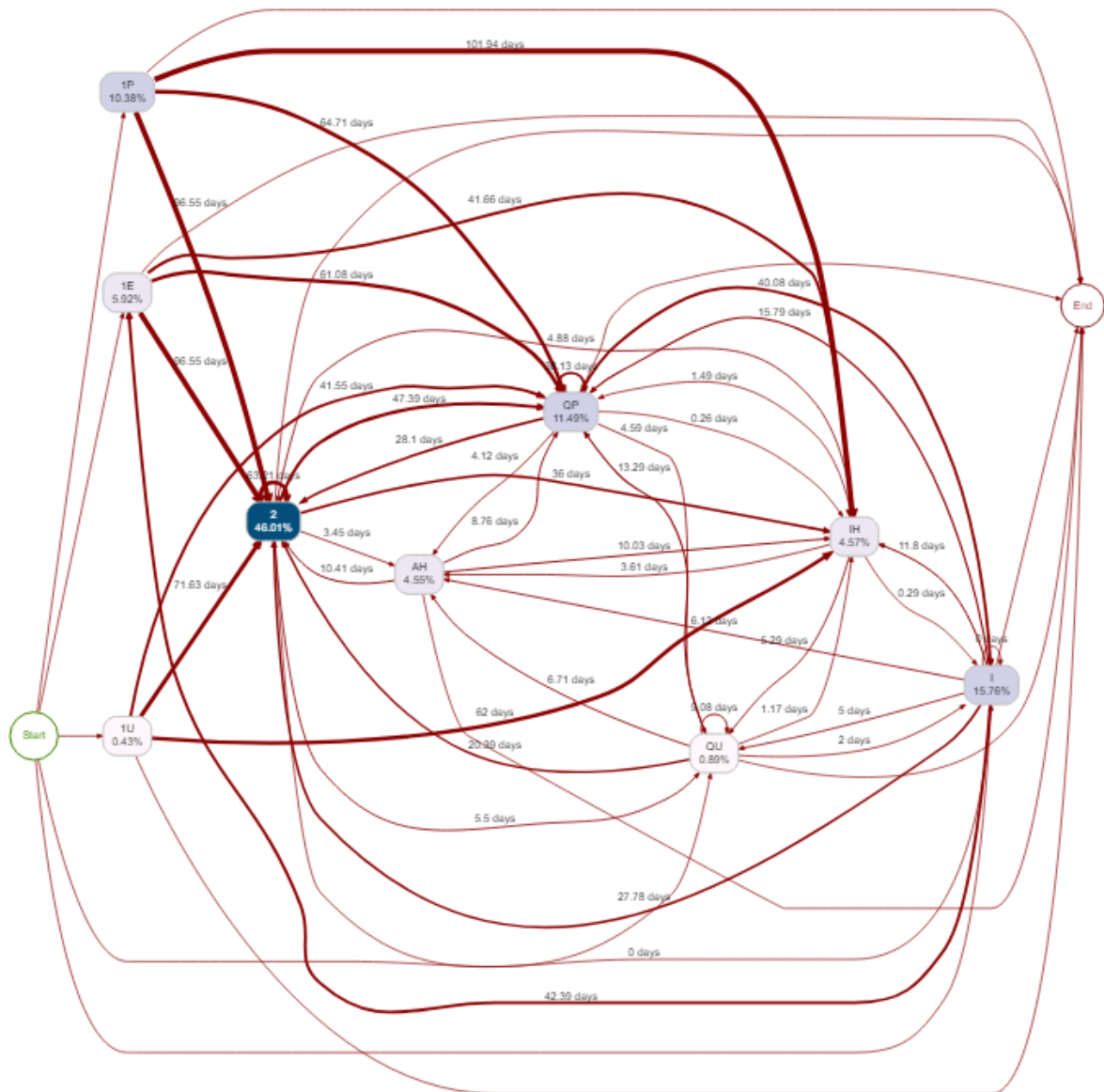


Ilustración 10. Mapa de procesos con el tiempo medio entre contactos.

Grafo de los procesos donde los nodos son los tipos de contacto y el valor de las aristas es la media de tiempo que se ha tardado en ir de un nodo a otro. Por ejemplo, la media de tiempo entre una cirugía programada y una consulta sucesiva es de 28.1 días.

4.4.2. Procesos frecuentes

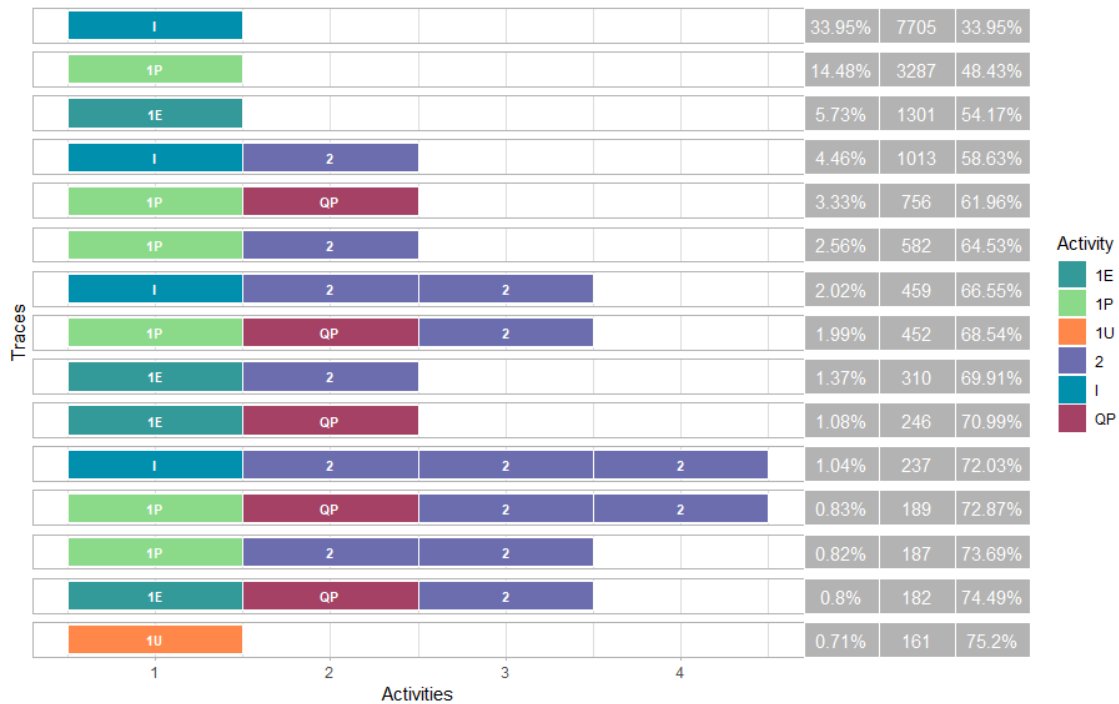


Ilustración 11. Gráfica con los procesos más frecuentes, representando el 75% de los procesos.

En la Ilustración 11 se puede ver los procesos más frecuentes. A la izquierda sale el proceso completo, la primera columna de valores es el porcentaje de procesos que sigue ese mismo camino, la segunda columna es lo mismo, pero en valores absolutos y, la última columna es el porcentaje acumulado de los procesos. Se muestran solo hasta el 75% de los procesos.

5. Minería de datos

Una vez hecha y analizada la minería de procesos, lo que mas impactó a las expertas de cirugía plástica fue esos procesos en los que solo se realiza una primera consulta y después se les da el alta. Según ellas esto son casos que no se deberían dar y, si conseguimos reducir este tipo de procesos, se produciría una reducción considerable de las listas de espera para consulta.

Ahora es donde entra en juego el aprendizaje automático. En esta sección lo que vamos a hacer va a ser construir un algoritmo de predicción que nos diga si, dados unos datos del paciente, tras una primera consulta procedente de primaria, va a ser dado de alta o va a continuar su proceso dentro del servicio. Esto lo vamos a realizar, por lo menos para este trabajo, solo para las consultas procedentes de primaria. Como líneas futuras se pretende aplicar para cirugías y para primeras procedentes de especializada.

5.1. Elección de los datos

Para escoger los datos escogimos una serie de variables genéricas y algunas también específicas. Las escogimos según el criterio de Andrea San Martín (jefa de cirugía plástica en el HUN). Todos los datos han sido obtenidos de Bardena.

5.1.1. Variables de entrada

Para las variables de entrada tratamos de escoger las variables representativas de las personas que se puedan obtener antes de que se realice ninguna consulta. Por ello las variables que elegimos fueron las siguientes:

- **Edad:** Edad de cada persona en el momento en que se produjo la consulta.
- **Sexo:** Puede ser masculino o femenino.
- **Aportación a Salud:** Es una variable socioeconómica de cada paciente, que indica cuanto aporta a salud.
- **IMC:** El índice de masa corporal (IMC) es un indicador simple de la relación entre el peso y la talla que se utiliza frecuentemente para identificar el sobrepeso y la obesidad en los adultos[20].
- **Diabetes:** Es un indicador de si el paciente ha tenido diabetes o no.
- **MUST:** Es un medidor de riesgo de desnutrición en adultos[21]. El riesgo puede ser bajo, medio o alto.
- **Tabaquismo:** Indica si el paciente es fumador activo, exfumador o no ha fumado nunca.
- **GMA_Peso:** El peso del GMA se refiere al valor numérico asignado a cada grupo de morbilidad ajustado[22].
- **Intervenciones anteriores:** Es el número de intervenciones que ha tenido el paciente el año antes de su consulta.
- **Coagulación:** Indica si el paciente ha tenido problemas de coagulación o no.
- **Fármacos:** Indica si el paciente toma fármacos o no.
- **Consultas AE:** Es el número de consultas de atención especializada que ha tenido el paciente el año anterior a la consulta. Se tienen en cuenta los servicios de cirugía vascular, dermatología, rehabilitación, traumatología y ortopedia, y hematología.
- **Creatinina:** Es un valor que mide los niveles de creatinina en la sangre o en la orina[23]. Tener un valor menor que 1 indica un buen funcionamiento renal.
- **Atenciones de urgencias:** Es el número de veces que el paciente ha asistido a urgencias en el último año.
- **Índice Charlson:** Es un indicador que dice si un paciente ha sufrido cierto problema de salud. Viene dado por la fecha en la que se produjo. Hay un total de 25 problemas de salud que hemos considerado, que son, a priori, los más determinantes en cirugía plástica.

5.1.2. Variable de salida

La variable de salida como ya hemos comentado va a ser una variable binaria, que va a indicar si el paciente se le da el alta tras una primera consulta procedente de atención primaria o no. Como vemos en la Ilustración 12 tenemos una muestra bastante balanceada. Esto se podía ver también en los mapas de procesos, donde teníamos que 3287 procesos (de 7510) fueron dados de alta tras la primera consulta.

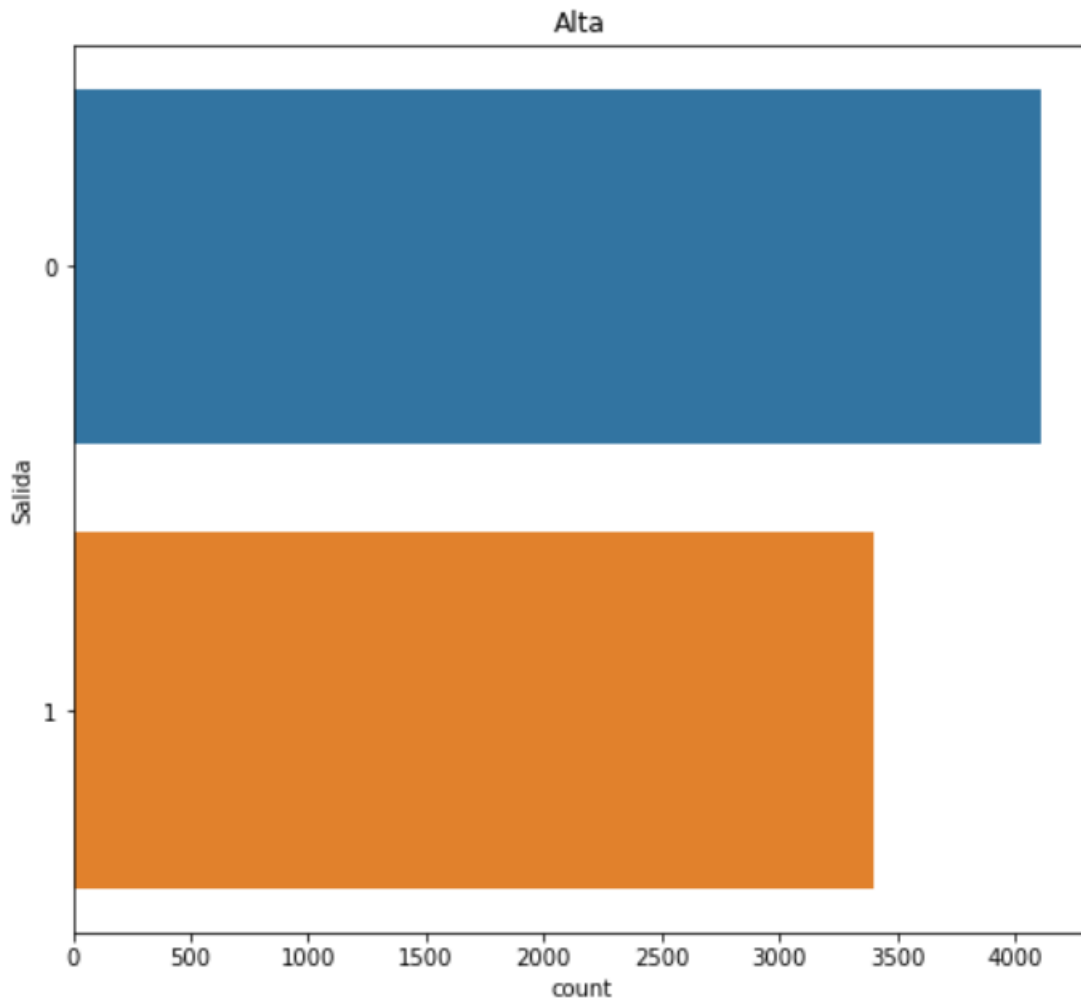


Ilustración 12. Gráfico de conteo de la variable de salida

La Ilustración 12 Gráfico de conteo que indica el balanceo de la variable de salida. En este caso el 1 indica alta y el 0 no alta.

5.2. Preprocesamiento de datos

En esta sección vamos a realizar los pasos previos a la clasificación. Dentro de las bases de datos hay veces que los datos son tomados de forma inadecuada y, por eso, es que hay que preparar los datos. Puede parecer que la parte más importante sea hacer el algoritmo de clasificación, pero lo cierto es que la preparación de los datos es la fase en la que hay que hacer mas énfasis, ya que es necesario arreglar los errores que puede haber y asegurar

la consistencia de los datos para asegurarnos que nuestros datos son buenos para hacer una buena clasificación[24].

Lo primero que tenemos que hacer en la preparación de los datos es la transformación de todas las variables a numéricas. Para las variables que solo pueden representar dos estados, como la diabetes, la transformación es fácil, asignamos 1 si lo tienen y 0 en caso contrario. Sin embargo, para las variables categóricas que pueden tener más de dos posibles valores vamos a hacer la transformación One Hot [25]. Este método consiste en crear una variable binaria (indicando presencia o ausencia de un valor) por cada valor distinto de la variable categórica. En nuestro caso lo vamos a tener que aplicar para dos variables solamente, que son el tabaquismo y la aportación a salud. Tras aplicar la transformación tenemos un dataset de 57 columnas y 7510 filas.

Después procedimos a hacer un análisis exploratorio de los datos. Con las variables numéricas como la edad o el IMC graficamos histogramas y boxplots y, para las categóricas, hicimos gráficos de barras. Durante el análisis de la visualización de los datos pudimos observar que las variables IMC y creatinina tiene varios valores anómalos, como se puede observar en la Ilustración 13. Como se puede observar hay valores del IMC superiores a 8000, lo cual es físicamente imposible.

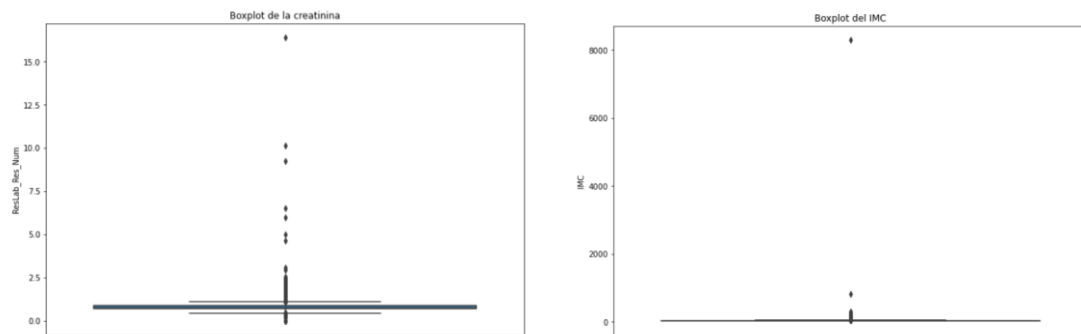


Ilustración 13. Gráficos de cajas de las variables creatinina (izquierda) y IMC (derecha)

Tras observar gráficamente estos valores anómalos calculamos mediante el rango intercuartílico. Este método consiste en comparar los datos respecto a los cuartiles de la variable. Se dice que un valor es anómalo si el dato se encuentra fuera del intervalo $(Q_1 - k * IQR, Q_3 + k * IQR)$, siendo $IQR = Q_3 - Q_1$ y k un parámetro que amplía el intervalo [26]. Con un valor de $k=3$ ya se consideraría un valor extremo, pero nosotros vamos a establecerlo en 5 para captar los muy extremos.

Con esta función obtenemos que con la variable IMC hay 19 outliers (valores anómalos), con el GMA 2 y con la creatinina 3. Respecto al IMC hay valores que superan ampliamente lo máximo visto en la historia (posiblemente debido a errores en la recogida de datos), que fue un valor de 185 [27]. Sin embargo, esto sigue sin ser un valor típico. Lo mas normal, según las expertas del HUN, es que el valor del IMC esté por debajo de 70. Por tanto, los valores que haya por encima de ese umbral no los consideraremos para nuestro estudio. Tras eliminar los registros con valores anómalos las gráficas quedan mucho mas normales, como se observa en la Ilustración 14.

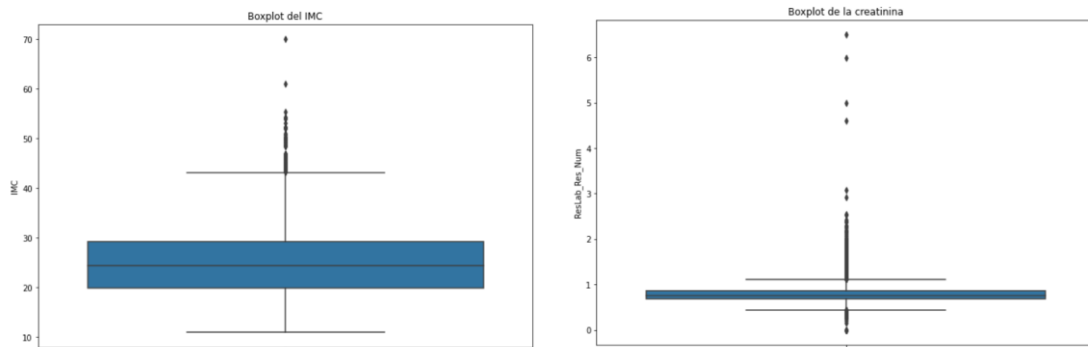


Ilustración 14. Gráficos de cajas de las variables creatinina (izquierda) y IMC (derecha) tras la eliminación de outliers

Otra cosa importante que hay que examinar en nuestros datos es si faltan datos. Y, efectivamente, en nuestros datos faltan valores en las variables de IMC, GMA y creatinina, por lo que vamos a tener que imputarlos para poder luego hacer nuestro algoritmo de clasificación. Para imputar los valores nulos de estas tres variables vamos a utilizar la técnica de imputación por regresión estocástica. Esta técnica consiste en hacer una regresión lineal con el resto de las variables (las que no tienen nulos) y, para solucionar la subestimación de los datos que produce una regresión, sumaremos un valor aleatorio, de la siguiente distribución:

$$N(0, \sigma_{Predsobredatoscompletos})$$

Para obtener datos más precisos no vamos a considerar todas las variables para hacer la regresión lineal (entrenada con los datos que no tienen valores nulos). Solo consideraremos las que tengan una correlación mayor a 0.2 (en valor absoluto) con la variable de salida. Con estos valores que obtengamos (sumándole el valor de la normal) imputaremos los datos faltantes. Tras esto, nuestro dataset ya estará listo para la clasificación.

5.3. Clasificación

Para la clasificación es importante tener en cuenta qué es lo que queremos maximizar o minimizar. En nuestro caso queremos reducir las listas de espera, por lo que nos interesaría aumentar la cantidad de gente que se les da el alta, pero siempre teniendo cuidado, ya que en medicina puede ser que se le dé el alta y luego resulta que necesitaba una cirugía, un ingreso... Los tipos de resultados que se pueden dar en la clasificación binaria vienen dados por la matriz de confusión (Ilustración 15).

En nuestro caso los positivos son las altas y los negativos son las no altas. Por tanto, lo que nos interesa es minimizar la gente que se les da el alta y que, en realidad, no estaban listos para darles de alta. Como también nos interesa minimizar el número de personas que siguen por el proceso que, en realidad, no necesitaban y se les podía dar el alta, entonces vamos a utilizar como métrica el $F1_score$.

		Predicción	
		Positivos	Negativos
Observación	Positivos	Verdaderos Positivos (VP)	Falsos Negativos (FN)
	Negativos	Falsos Positivos (FP)	Verdaderos Negativos (VN)

Ilustración 15. Matriz de confusión [Rpubs].

Una vez hecha la elección de la métrica vamos a proceder con la clasificación. Para ello hemos utilizado la librería `lazypredict` de Python. Lo que hace la función integrada en la librería (`LazyClassifier`) es clasificar nuestros datos con un total de 29 modelos de machine learning diferentes para ver, a priori, cual funciona mejor. Primero dividimos nuestros datos en train, validación y test, con un 60%, 20% y 20% respectivamente. Los resultados obtenidos en la validación se presentan en la Ilustración 16.

Model	Accuracy	Balanced Accuracy	ROC AUC	F1 Score	Time Taken
NearestCentroid	0.59	0.60	0.60	0.58	0.02
LogisticRegression	0.60	0.59	0.59	0.59	0.04
BernoulliNB	0.59	0.58	0.58	0.59	0.02
AdaBoostClassifier	0.60	0.58	0.58	0.59	14.88
SVC	0.59	0.58	0.58	0.59	2.17
LinearDiscriminantAnalysis	0.60	0.58	0.58	0.59	0.05
RidgeClassifier	0.60	0.58	0.58	0.59	0.03
RidgeClassifierCV	0.60	0.58	0.58	0.59	0.04
Perceptron	0.58	0.58	0.58	0.58	0.02
LinearSVC	0.59	0.58	0.58	0.59	0.67
CalibratedClassifierCV	0.59	0.57	0.57	0.57	2.63
LGBMClassifier	0.56	0.56	0.56	0.56	0.10
RandomForestClassifier	0.56	0.55	0.55	0.55	7.61

Ilustración 16. Resultados obtenidos con `LazyClassifier` (los 13 mejores modelos)

Podemos observar que no se obtienen buenos resultados. Nuestro mejor modelo no llega a 0.6 de `F1_score`, por lo que podemos concluir que nuestros datos de entrada no describen correctamente nuestra variable de salida. Para líneas futuras se podría explorar la posibilidad de cambiar las variables de entrada o modificarlas para ver si así se obtienen mejores resultados.

6. Conclusiones

En este trabajo hemos tratado de mejorar la eficiencia del servicio de cirugía plástica del Hospital Universitario de Navarra mediante minería de procesos y minería de datos. En primer lugar, hemos descrito su funcionamiento mediante la minería de procesos, haciendo una visualización por grafos donde se podían observar varios cuellos de botella en los que trabajar para poder mejorar la eficiencia del servicio.

Posteriormente, mediante la minería de datos, exploramos la posibilidad de predecir si un paciente va a ser dado de alta o no dentro del servicio. Aunque no hemos obtenido resultados satisfactorios con nuestros datos, este estudio sienta las bases para futuros análisis más exhaustivos de cada servicio médico. La limitación de tiempo fue un obstáculo en esta sección, ya que con más tiempo habríamos podido probar diferentes variables de entrada y salida.

También consideramos la posibilidad de desarrollar un algoritmo para predecir si un paciente requería una cirugía mayor o menor. Esta predicción podría ser útil, ya que, si se determinara que la cirugía es menor, el paciente podría ser programado en una sala especial que combina consulta y quirófano. Esto permitiría ahorrar una consulta adicional, ya que el paciente podría pasar directamente a la cirugía después de un chequeo previo.

En resumen, este estudio nos ha mostrado que los servicios médicos tienen margen de mejora significativo y que una gestión más eficiente puede reducir las listas de espera.

7. Bibliografía

1. Listas de espera [Internet]. 2022. Disponible en: http://www.navarra.es/home_es/Temas/Portal+de+la+Salud/Ciudadania/Actualidad/Listas+de+espera/Octubre+2022.htm
2. Hayashi C. Studies in Classification, Data Analysis, and Knowledge Organization. 1998;40-51.
3. Harris H. Data Science, Moore's Law, and Moneyball [Internet]. 2011. Disponible en: <https://web.archive.org/web/20150925121255/http://www.harlan.harris.name/2011/09/data-science-moores-law-and-moneyball/>
4. M.Y. Tai A. Machine learning and big data: Implications for disease modeling and therapeutic discovery in psychiatry. 2019; Disponible en: https://www.sciencedirect.com/science/article/pii/S0933365717301781?casa_token

n=62LDNMx5W7QAAAAA:VN_ytpK6nFcwTwqZyjGdHaISyE2yGUaiUwEPm
mIN6j-7hJ2k0jVgBdUFNu0cekYUYalhAe9L

5. UVa. Teoría de grafos para redefinir la esquizofrenia. 2019; Disponible en: <https://www.desdesoria.es/articulo/noticias/la-uva-y-el-hospital-clinico-emplean-la-teoria-de-grafos-para-redefinir-la-esquizofrenia/20190530162704583311.html>
6. RuiLing W. Application of Data Mining in Performance Management of Public Hospitals. 2022; Disponible en: <https://doi.org/10.1155/2022/2412928>
7. Munoz Gama J. Process mining for healthcare: Characteristics and challenges. 2022; Disponible en: <https://www.sciencedirect.com/science/article/pii/S1532046422000107>
8. Wil M P van der A. Process mining [Internet]. 2011. Disponible en: <https://link.springer.com/book/10.1007/978-3-642-19345-3>
9. Trudeau RJ. Introduction to graph theory [Internet]. 1993. 228 p. Disponible en: <https://archive.org/details/introductionto0000trud/page/n227/mode/2up>
10. Bang-Jensen J. Digraphs Theory, Algorithms and Applications [Internet]. 2007. Disponible en: <https://www.cs.rhul.ac.uk/books/dbook/main.pdf>
11. Maimon O. Data Mining and Knowledge Discovery Handbook [Internet]. 2010. Disponible en: <https://link.springer.com/book/10.1007/978-0-387-09823-4>
12. Witten IH. Data Mining: Practical Machine Learning Tools and Techniques, 3rd Edition [Internet]. 2011. Disponible en: <https://ebookcentral.proquest.com/lib/upnasp-ebooks/detail.action?docID=634862>
13. SECPRE. ¿Qué es la Cirugía Plástica? [Internet]. Disponible en: <https://secpre.org/que-es-la-cirurgia-plastica>
14. Wikipedia. Cirugía estética [Internet]. 2023. Disponible en: https://es.wikipedia.org/wiki/Cirug%C3%ADa_est%C3%A9tica#cite_note-2
15. Centro Médico T. ¿Qué es la cirugía reconstructiva? [Internet]. Disponible en: <https://www.teknon.es/es/cirurgia-plastica-estetica/preguntas-frecuentes/preguntas-frecuentes-cirurgia-reconstructiva/cirurgia-reconstructiva>
16. Atención primaria. En 1978. Disponible en: https://es.wikipedia.org/wiki/Conferencia_Internacional_sobre_Atenci%C3%B3n_Primaria_de_Salud_de_Alma_Ata
17. Ministerio de Sanidad. MANUAL DE DEFINICIONES ESTADÍSTICA DE CENTROS SANITARIOS DE ATENCIÓN ESPECIALIZADA [Internet]. 2003. Disponible en: https://www.sanidad.gob.es/en/estadEstudios/estadisticas/docs/DEFINICIONES_C1.pdf

18. Real Academia Nacional de Medicina. Definición de exitus [Internet]. Disponible en: <http://dtme.ranm.es/terminos/exitus.html?id=222>
19. van der Aalst, Eindhoven University of Technology, Eindhoven, The Netherlands, VVan der Aalst, W.M.P. Finding Structure in Unstructured Processes: The Case for Process Mining. 2007; Disponible en: <https://ieeexplore.ieee.org/document/4276259>
20. OMS. Obesidad y sobrepeso [Internet]. 2021. Disponible en: <https://www.who.int/es/news-room/fact-sheets/detail/obesity-and-overweight>
21. Marinos E. MANUAL EXPLICATIVO ‘MUST’ [Internet]. 2011. Disponible en: <https://www.bapen.org.uk/images/pdfs/must/spanish/must-exp-bk.pdf>
22. Monterde D. Los grupos de morbilidad ajustados: nuevo agrupador de morbilidad poblacional de utilidad en el ámbito de la atención primaria. 2016; Disponible en: <https://www.elsevier.es/es-revista-atencion-primaria-27-articulo-los-grupos-morbilidad-ajustados-nuevo-S0212656716302104>
23. MedlinePlus. Prueba de creatinina [Internet]. 2020. Disponible en: <https://medlineplus.gov/spanish/pruebas-de-laboratorio/prueba-de-creatinina/#:~:text=La%20prueba%20de%20creatinina%20se,de%20un%20panel%20metab%C3%B3lico%20completo.>
24. Davidson I. Data preparation using data quality matrices for classification mining. 2009; Disponible en: <https://www.sciencedirect.com/science/article/pii/S0377221708005602>
25. Brownlee J. Why One-Hot Encode Data in Machine Learning? [Internet]. 2020. Disponible en: <https://machinelearningmastery.com/why-one-hot-encode-data-in-machine-learning/>
26. Wei X. A quality control method based on improved IQR for estimating multi-GNSS real-time satellite clock offset. 2022; Disponible en: <https://www.sciencedirect.com/science/article/pii/S0263224122009034>
27. Jon Brower Minnoch. En. Disponible en: https://es.wikipedia.org/wiki/Jon_Brower_Minnoch