



**UNIVERSIDAD DE CONCEPCIÓN**  
**FACULTAD DE INGENIERÍA**  
**DEPARTAMENTO DE INGENIERÍA ELÉCTRICA**



**PROPUESTA DE AUTOMATIZACIÓN PARA PROCESO DE AGENDA-  
MIENTO DE CITAS MÉDICAS POST-ALTA A PARTIR DE EPICRISIS EN  
EL SERVICIO DE MEDICINA INTERNA DEL HOSPITAL LAS HIGUERAS**

POR

**Francisca Belén Valdés Brito**

Memoria de Título presentada a la Facultad de Ingeniería de la Universidad de Concepción para  
optar al título profesional de Ingeniera Civil Biomédica

Profesores Guía

Dr. Rosa Figueroa I.

Dr. Hernán Carrillo B.

Comisión

Dr. Esteban Pino Q.

Jean Paul Navarrete C.

Agosto 2024

Concepción  
(Chile)

© 2024 Francisca Belén Valdés Brito

© 2024 Francisca Belén Valdés Brito

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento.



## Agradecimientos

En primer lugar, quiero agradecer a mi familia por brindarme la oportunidad de estudiar esta carrera lejos de casa, asegurándose de que nunca me faltara nada. Agradezco su amor incondicional, por estar siempre presentes, por creer en mí desde el principio y por darme ánimos para perseguir mis metas y seguir adelante con todo, incluso estando a la distancia. Gracias por apoyarme en cada decisión que he tomado, cuidarme y enseñarme a dar lo mejor de mí en todo lo que me propongo hacer.

A mi pololo, por su comprensión, apoyo y compañía. Gracias por estar siempre ahí para celebrar mis logros y subirme el ánimo en los días que no estaba al 100%. Tus palabras de ánimo y tu paciencia han sido ese empujón necesario para mantenerme enfocada y motivada durante todo este proceso de crecimiento tanto académico como personal.

También quiero agradecer a las amistades que formé a lo largo de estos años, quienes han compartido conmigo este camino universitario. Gracias por los innumerables momentos de alegría, risas, carretes y apoyo mutuo, especialmente cuando necesitaba desconectarme de los estudios y disfrutar de su compañía con panoramas entretenidos.

Por supuesto, no puedo dejar de mencionar a los académicos y profesionales que hicieron posible este proyecto, agradezco su guía y enseñanzas. En particular, quiero reconocer al doctor y a mi profesora guía por su constante apoyo y dedicación ya que sus reuniones, correcciones y orientaciones fueron fundamentales para la realización de esta Memoria.

Finalmente, agradecer a quienes fueron parte de esta etapa de mi vida, ¡este logro fue gracias a ustedes!

## Resumen

El aumento exponencial de la producción de datos en el ámbito de la salud ha provocado la subutilización de la información contenida en ellos, una de las razones es debido a su naturaleza no estructurada. Este informe aborda la problemática del proceso manual de agendamiento de citas médicas post-alta, que genera ineficiencias y una carga administrativa solucionable mediante tecnologías y aprovechamiento de datos existentes.

Para resolver esta situación, se propuso un sistema automatizado basado en técnicas de Machine Learning, Data Mining y Procesamiento de Lenguaje Natural, utilizando herramientas y bibliotecas de Python y Google. Este sistema permitió clasificar y priorizar episodios de pacientes para su agendamiento en controles post-alta, transformando la información extraída de las epicrisis en un formato estructurado y utilizable para análisis y toma de decisiones, considerando variables demográficas, diagnósticas y tiempos de espera.

La finalidad del proyecto fue mejorar la eficiencia y efectividad del agendamiento, reduciendo errores en la lectura e ingreso de datos, además de mejorar la coordinación entre el personal hospitalario desde la emisión de la epicrisis hasta la programación de la cita del paciente. Esto resulta en una disminución de los desplazamientos innecesarios del personal médico hacia los mesones de los policlínicos de interconsulta para coordinar controles post-alta, facilitando una transmisión rápida y precisa de la información clave.

La metodología incluyó el desarrollo de un algoritmo para extraer y analizar datos de las epicrisis de la base de datos del Hospital Las Higueras de Talcahuano, así como su integración con aplicaciones de Google para visualización y gestión. Los resultados demostraron que el sistema implementado procesa los datos de manera eficiente y coherente en el entorno hospitalario, proporcionando información sobre los episodios y su flujo para el agendamiento en diferentes policlínicos. Donde la herramienta, desde el punto de vista de un usuario experto, responde a las necesidades de información del proceso de alta y controles del Servicio de Medicina Interna. Esto destaca el potencial de las tecnologías de la información en la gestión médica actual, sirviendo como modelo para innovar y mejorar entornos dinámicos.

## Abstract

The exponential increase in data production in the healthcare sector has led to underutilization of the information contained within them, primarily due to their unstructured nature. This report addresses the issue of manual scheduling processes for post-discharge medical appointments, which create inefficiencies and administrative burdens that can be resolved through technology and leveraging existing data.

To address this situation, an automated system was proposed based on Machine Learning, Data Mining, and Natural Language Processing techniques, utilizing Python tools and libraries and Google. This system enabled the classification and prioritization of patient episodes for scheduling post-discharge appointments, transforming information extracted from discharge summaries into a structured and usable format for analysis and decision-making, considering demographic variables, diagnoses, and waiting times.

The project aimed to enhance scheduling efficiency and effectiveness by reducing errors in data reading and entry, as well as improving coordination among hospital staff from discharge summary issuance to patient appointment scheduling. This results in fewer unnecessary trips by medical personnel to outpatient clinic desks for post-discharge appointment coordination, facilitating rapid and precise transmission of key information.

The methodology involved developing an algorithm to extract and analyze data from discharge summaries at Hospital Las Higueras de Talcahuano's database, integrating it with Google applications for visualization and management. The results demonstrated that the implemented system efficiently and consistently processes data within the hospital environment, providing insights into patient episodes and their flow for scheduling across various outpatient clinics. From the perspective of an expert user, the tool meets the information needs of the discharge and follow-up process in the Internal Medicine Service. This underscores the potential of information technologies in current medical management, serving as a model for innovation and improvement in dynamic environments.

## Tabla de Contenidos

<b>CAPÍTULO 1. INTRODUCCIÓN.....</b>	<b>1</b>
1.1. INTRODUCCIÓN GENERAL .....	1
1.2. OBJETIVOS.....	3
1.2.1 Objetivo General.....	3
1.2.2 Objetivos Específicos .....	3
1.3. ALCANCES Y LIMITACIONES.....	3
1.4. METODOLOGÍA .....	4
1.4.1 Selección de Datos y Variables.....	5
1.4.2 Selección del Tamaño Muestral .....	6
1.5. TEMARIO .....	7
<b>CAPÍTULO 2. MARCO TEÓRICO.....</b>	<b>8</b>
2.1. INTRODUCCIÓN .....	8
2.2. ANTECEDENTES DEL PROCESO DE AGENDAMIENTO DE CITAS MÉDICAS EN HOSPITAL LAS HIGUERAS .....	9
2.3. EPICRISIS MÉDICA .....	10
2.4. CIE-10.....	11
2.5. SISTEMAS DE INFORMACIÓN Y TECNOLOGÍAS DIGITALES EN EL ÁMBITO DE LA SALUD Y ATENCIÓN MÉDICA .	11
2.5.1 Sistemas y Tecnología de la Información .....	11
2.5.2 Aprendizaje Automático .....	12
A. Sistemas Basados en Conocimiento.....	13
2.5.3 Minería de Datos.....	13
A. Procesamiento de Lenguaje Natural.....	14
B. Expresiones Regulares .....	15
C. Medidas de distancia .....	15
2.5.4 Heurística.....	15
2.5.5 Lenguajes de programación.....	15
2.5.6 Google.....	16
A. Google Workspace .....	16
B. Google Cloud .....	17
2.6. ESTIMADOR DE HORVITZ-THOMPSON PARA LA MEDIA EN DISTRIBUCIÓN BINOMIAL NEGATIVA.....	17
<b>CAPÍTULO 3. REVISIÓN BIBLIOGRÁFICA .....</b>	<b>21</b>
3.1. INTRODUCCIÓN .....	21
3.2. INTRODUCCIÓN AL AGENDAMIENTO Y PROGRAMACIÓN DE CITAS EN EL ÁMBITO MÉDICO .....	21
3.2.1 Agendamiento de Citas Médicas Ambulatorias .....	22
3.3. TECNOLOGÍAS DE INTELIGENCIA ARTIFICIAL PARA AGENDAMIENTO DE CITAS MÉDICAS.....	23
3.3.1 Extracción de Información Clínica.....	23
3.3.2 Implementación de Google Workspace en el ámbito médico .....	25
3.3.3 Sistemas Basados en Reglas.....	26
3.4. DISCUSIÓN.....	26
<b>CAPÍTULO 4. LEVANTAMIENTO DE INFORMACIÓN .....</b>	<b>29</b>

4.1. INTRODUCCIÓN A LA PROBLEMÁTICA .....	29
4.1.1 <i>Recopilación de Información</i> .....	29
<b>CAPÍTULO 5. PROPUESTA DE AUTOMATIZACIÓN.....</b>	<b>31</b>
5.1. FUNCIONALIDAD DEL SISTEMA AUTOMATIZADO .....	31
5.1.1 <i>Extracción de Información de Epicrisis</i> .....	31
5.1.2 <i>Clasificación y Priorización de Pacientes</i> .....	32
5.2. ARQUITECTURA Y DISEÑO DEL SISTEMA.....	34
5.2.1 <i>Arquitectura del Sistema</i> .....	34
5.2.2 <i>Diseño del Sistema</i> .....	35
5.2.3 <i>Esquema del Sistema y componentes</i> .....	37
<b>CAPÍTULO 6. RESULTADOS.....</b>	<b>38</b>
6.1. INTRODUCCIÓN.....	38
6.2. EXTRACCIÓN DE INFORMACIÓN DE EPICRISIS .....	38
6.3. CLASIFICACIÓN Y PRIORIZACIÓN DE PACIENTES.....	39
<b>CAPÍTULO 7. CONCLUSIONES.....</b>	<b>43</b>
7.1. INTRODUCCIÓN.....	43
7.2. CONCLUSIONES .....	43
7.3. AVANCES FUTUROS.....	44
<b>GLOSARIO.....</b>	<b>46</b>
<b>REFERENCIAS.....</b>	<b>48</b>
<b>ANEXO A. APROBACIÓN COMITÉ ETICO CIENTÍFICO.....</b>	<b>54</b>



## Lista de Tablas

Tabla 1.1. Tamaños Mínimos Muestrales para el proyecto .....	6
Tabla 2.1. Ejemplo de registro para el agendamiento de horas .....	10
Tabla 4.1. Preguntas para el Levantamiento del Proceso .....	29
Tabla 5.1. Reglas, cálculos y ejemplos de variables para el cálculo del Puntaje de Prioridad .....	33



## Lista de Figuras

Figura 1.1. Diagrama general del proceso .....	5
Figura 5.1. Esquema del sistema y componentes .....	37
Figura 6.1. Extracción de variables estructuradas .....	38
Figura 6.2. Extracción sección 'Indicaciones de Egreso' .....	38
Figura 6.3. Extracción de policlínicos, temporalidades y adición de variables .....	39
Figura 6.4. Sección 'Información General' AppSheet.....	40
Figura 6.5. Sección 'Indicadores' AppSheet.....	41
Figura 6.6. Sección 'Policlínico Broncopulmonar' AppSheet.....	41
Figura 6.7. Sección 'Policlínico Broncopulmonar' y datos extra del episodio en AppSheet. ....	42



## Capítulo 1. Introducción

---

### 1.1. Introducción General

La producción mundial de datos está en aumento, y la introducción de sistemas de información en apoyo a los procesos clínicos, administrativos y de gestión colaboran con este aumento constante de información que, generalmente sirve solo para los fines primarios para los que fue creada [1]. Estos sistemas están en instituciones de salud, como hospitales, centros de salud comunitarios y clínicas. La información no estructurada que generan dificulta su análisis inmediato con herramientas convencionales [2], lo que lleva a una subutilización de datos que podrían ser valiosos para la toma de decisiones en procesos clínicos y/o la gestión de recursos [3]. Problema relevante por la demanda constante en los servicios de salud.

Adicionalmente, los sistemas de salud deben tener en cuenta la eficiencia y eficacia de la distribución y utilización de sus recursos para dar un servicio de calidad con tal de promover, restablecer o mantener la salud de quien recurra a ellos [4]. Entre estos servicios del sistema está el recurso de la cita médica, fundamental en el tratamiento y seguimiento de patologías, en donde la programación en tiempo razonable y organizada es vital tanto para los recursos limitados de los servicios de salud como para los usuarios de estos. De esta manera, no solo se asegura la sostenibilidad a largo plazo, sino que impulsa el avance hacia una economía circular al abordar las demandas actuales sin descuidar el futuro. Este enfoque minimiza el desperdicio, fomenta la conservación de recursos y promueve la responsabilidad ambiental y social en todos los ámbitos [5]. De manera similar, la integración de la inteligencia artificial en la práctica clínica puede transformar completamente la eficiencia en la prestación de atención médica al mejorar la atención al paciente y su calidad de vida [6].

Los niveles de atención en salud suelen estar divididos en 3 o 4 en el mundo [7], en Chile se dividen en 3 [8]: primaria, secundaria y terciaria, donde la complejidad de estos va en aumento respecto a las necesidades de la atención [9]. En el tercer nivel de atención del sistema de salud se atienden casos de hospitalización cerrada donde hay personal médico especializado con equipos sofisticados de mayor rigor técnico para las variadas patologías, no obstante, los recursos son limitados y deben ser asignarlos con tal de entregar una atención de salud de calidad y segura para sus usuarios además de entregarles información oportuna y comprensible sobre sus estados de salud [10]. Gracias a esto se refleja la importancia de la asignación de citas médicas para garantizar tratamientos oportunos y facilitar el acceso a la atención en estas instituciones de salud.

En el Hospital Las Higueras de Talcahuano, el procedimiento actual tras dar de alta a un paciente en el departamento de Medicina Interna implica la generación de la epicrisis, un documento que resume eventos en la estadía hospitalaria y que contiene secciones en formato de texto libre y lenguaje natural como la sección de indicaciones de egreso relacionada con medicamentos, cuidados y consultas futuras a las que debe asistir el paciente para el seguimiento de su salud y recuperación.

La coordinación y consolidación de los controles post-alta indicados en la epicrisis recae en el personal de enfermería técnica (TENS), ya que se dirigen personalmente con la epicrisis en mano hacia el mesón de la interconsulta señalada en el documento. Para esto, el TENS o el personal en mesón leen la epicrisis para identificar la información relevante respecto al control solicitado en esta. Según los horarios de los médicos y sus agendas en el policlínico, se registra la consulta del paciente en agenda o se incluye en lista de espera. Estas últimas se manejan en archivos de hojas de cálculo que carecen de priorización basada en el historial clínico del paciente para la asignación de controles, al no existir un método que permita atribuir esta cualidad. Por eso la agenda se organiza solo según la disponibilidad de cada policlínico [11].

En este sistema de trabajo que carece de estandarización del lenguaje de algunas de las secciones que se utilizan de la epicrisis, dificulta el procesamiento eficiente de la información. Además, se observa un bajo uso de tecnología de la información (IT) en este proceso de agendamiento, lo que implica que haya una existencia limitada de la aplicación de herramientas y procesos para gestionar y organizar el flujo de datos [12].

Este proyecto se ve motivado por la necesidad de automatizar este proceso de agendamiento de citas médicas en el Servicio de Medicina Interna del Hospital Las Higueras como respuesta a las limitaciones del procedimiento actual. La utilización de documentos de epicrisis en formato físico para el proceso conduce a una coordinación ineficaz y se ve que la escasa integración de la IT ha resultado en una subutilización de herramientas que podrían mejorar la comunicación entre el personal hospitalario y de la gestión administrativa.

El análisis de los resultados de este proyecto permitirá determinar ajustes o mejoramientos a futuro en esta área hospitalaria, ya que se busca modernizar y eficientizar un aspecto fundamental de la atención médica implementando soluciones tecnológicas y estrategias de automatización. El sistema basado en técnicas de Aprendizaje Automático o Machine Learning (ML); Data Mining (DM) como Text Mining (TM) y Procesamiento de Lenguaje Natural (NLP) ayudaron a reducir la carga

administrativa, minimizar errores de tipeo para el agendamiento, evitar el desplazamiento innecesario del personal técnico de enfermería y agilizar el sistema para agendamiento de pacientes que necesitan seguir controlados después del alta médica.

## **1.2. Objetivos**

### **1.2.1 Objetivo General**

Proponer un sistema de clasificación y priorización de episodios para agendar controles médicos luego del alta médica, considerando variables demográficas, diagnósticas y tiempos de espera, a partir de información extraída de la base de datos del Servicio de Medicina Interna del Hospital Las Higueras.

### **1.2.2 Objetivos Específicos**

- Extraer información relevante a partir de epicrisis (edad, diagnóstico de egreso, fecha del alta, etapa GES y días de hospitalización) para posterior análisis.
- Clasificar y priorizar, para cada policlínico de la interconsulta del Servicio de Medicina Interna, a los episodios de pacientes.
- Proporcionar información sobre el flujo de pacientes agendados en los distintos policlínicos a través del sistema implementado.

## **1.3. Alcances y Limitaciones**

Este trabajo se enfocó en el desarrollo de un algoritmo para un sistema que clasifique y priorice episodios para controles médicos después del alta, de manera eficiente, considerando variables demográficas, diagnósticas y tiempos de espera, a partir de información extraída de la base de datos del Servicio de Medicina Interna del Hospital Las Higueras. Esta información fue proporcionada por el hospital luego de la aprobación del Comité Ético Científico del Servicio de Salud Talcahuano, según Acta N° 15 de la sesión ordinaria del día 28.03.2024, ver Anexo A.

Los resultados concretos son pacientes clasificados y priorizados en cada policlínico, permitiendo al personal administrativo agendar las citas de manera organizada evitando el desplazamiento innecesario de TENS u otro personal en este contexto, además de información complementaria para mejoras del sistema de agendamiento de citas médicas post-alta a futuro.

Durante el desarrollo de este proyecto, se enfrentaron diversas limitaciones que incidieron en su alcance y ejecución. Primero, el periodo asignado por la Universidad de Concepción para realizar el proyecto definido, imponiendo restricciones temporales que condicionaron la profundidad de análisis y desarrollo del sistema propuesto. Le siguen los desafíos de la utilización de algoritmos de programación para el idioma español, ya que la mayoría de los avances en la literatura se han visto para textos en inglés. Otro desafío fue el obtener la aprobación del Comité Ético Científico del Servicio de Salud Talcahuano para la utilización de las epicrisis anonimizadas del servicio de medicina. Este proceso implicó una limitación en cuánto al periodo de tiempo en el que se pudo acceder a los datos. La colaboración del personal hospitalario, fundamental en varios aspectos del proyecto, constituyó otra limitación significativa, ya que dependió de la disponibilidad y disposición de horarios de dicho personal. Estas limitaciones se abordaron con precaución durante el proceso para mitigar su impacto en la efectividad y validez de los resultados del proyecto.

#### **1.4. Metodología**

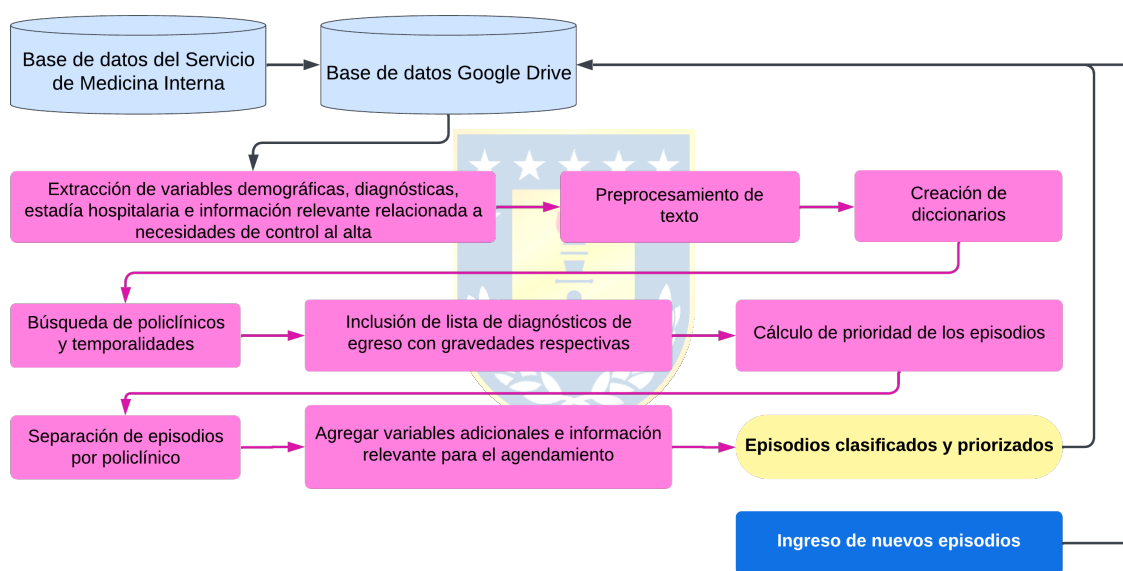
Con el desarrollo de este proyecto, se espera mejorar la eficiencia y efectividad del agendamiento de citas médicas post-alta, reduciendo la carga administrativa y minimizando errores y demoras en el proceso. Dado el contexto hospitalario en el que se desarrolla el proyecto, resulta crucial obtener datos reales para evaluar la metodología propuesta. Por esta razón, se buscó su implementación con datos e información del Servicio de Medicina Interna del Hospital Las Higueras de Talcahuano.

Luego de levantar el proceso existente en el Servicio, se tiene claro que hasta la fecha el procedimiento implica la generación de epicrisis en formato físico, seguido de un proceso manual de coordinación y consolidación de controles post-alta llevados a cabo por el personal de enfermería técnica y de la interconsulta del Servicio. Por esto, se realizará extracción de información automatizada y se utilizará esta misma para clasificar y priorizar a los episodios en los distintos policlínicos. A partir de ello, se generará una política adecuada para el agendamiento, la cual estará complementada con información relevante para el personal administrativo que lo gestiona.

Para trabajar con los datos presentes en la base de datos del hospital se procedió a desarrollar un algoritmo basado en técnicas y algoritmos de Aprendizaje Automático o Machine Learning (ML), utilización de Data Mining (DM) como Text Mining (TM) y Procesamiento de Lenguaje Natural (NLP) para la integración de datos y extracción de la información tanto de campos estructurados como

campos narrativos en texto libre. Estos algoritmos, en conjunto con sus respectivas librerías, y las aplicaciones que los contienen, se utilizaron para manipular, estructurar, analizar y evaluar los datos y resultados relevantes. La elección del lenguaje principal para la implementación del sistema fue Python utilizando Anaconda, ya que este lenguaje ofrece la versatilidad necesaria para la aplicación de los mencionados. Además, se integró la intercomunicación con aplicaciones de Google Cloud y Google Workspace como AppSheet y Google Drive, proporcionando herramientas útiles como la conexión de fuentes de datos y una interfaz amigable con el usuario para visualizar rápidamente los resultados obtenidos en Python.

Un diagrama del proceso general se puede ver en la Figura 1.1.



**Figura 1.1. Diagrama general del proceso**

### 1.4.1 Selección de Datos y Variables

La base de datos proporcionada del Servicio de Medicina Interna del Hospital Las Higueras, contenía las epicrisis de los episodios de hospitalización ocurridos en dicho servicio en los años 2020, 2021 y 2022 de pacientes adultos. Para efectos y finalidad de este proyecto se extrajeron los campos que a continuación, se nombran: edad (campo estructurado), fecha de egreso (campo estructurado), días de hospitalización (campo estructurado), variable GES (campo estructurado), diagnóstico de

egreso (campo narrativo en texto libre) e indicaciones de egreso incluyendo policlínico y temporalidad del próximo control.

### 1.4.2 Selección del Tamaño Muestral

Para poder estimar un tamaño mínimo adecuado, se recurre al cálculo mínimo de muestras a través de estadísticas para que el algoritmo funcione correctamente. Dada las características del problema y el tamaño y características de la base de datos, se observa que el factor de variabilidad que podía afectar la representatividad del set de datos era el diagnóstico de egreso.

Dado lo anterior, se calcula un tamaño muestral en base a las estadísticas de egresos del Servicio de Medicina Interna del Hospital Las Higueras. Lo primero fue agrupar las clasificaciones por sección de diagnósticos principales según el código CIE-10, debido a la gran cantidad de diagnósticos existentes y así poder lograr una representatividad significativa en los cálculos para el tamaño muestral. Con las agrupaciones de las clasificaciones efectuadas que estuvieron sobre el 2% de representatividad respecto al mismo año, se calculó el tamaño mínimo de muestra con el estimador de Horvitz-Thompson para la media en distribución binomial negativa con muestreo estratificado.

Con estas fórmulas y los datos de representatividad (diagnósticos por sobre 2% de representatividad anual) de un año, extraídos de los egresos por diagnóstico principal en el Servicio de Medicina Interna del Hospital Las Higueras, se concluye que los resultados del mínimo tamaño muestral deben ser los ilustrados en la Tabla 1.1 donde los parámetros de confianza y representatividad fueron de 0.95 y 0.05 respectivamente para el mejor resultado.

**Tabla 1.1. Tamaños Mínimos Muestrales para el proyecto**

Nivel de Confianza	Precisión	Tamaño Muestral Mínimo Año 2020	Tamaño Muestral Mínimo Año 2021	Tamaño Muestral Mínimo Año 2022
0.95	0.05	99	88	79

## 1.5. Temario

- (i) Capítulo 1: Introducción general del trabajo. Se indican los objetivos, alcances y limitaciones del proyecto junto a la metodología.
- (ii) Capítulo 2: En esta sección se tiene el marco teórico, el cual entrega contexto sobre el tema y las tecnologías que se han desarrollado en esta área.
- (iii) Capítulo 3: Se proporciona un análisis bibliográfico del estado del arte. Además, se tiene una discusión en la que se describe la información destacable de este análisis.
- (iv) Capítulo 4: Se muestra la información recopilada del levantamiento de información para análisis de la problemática.
- (v) Capítulo 5: Se describe la propuesta de automatización detallando su funcionalidad, diseño y arquitectura del sistema.
- (vi) Capítulo 6: En este capítulo se mencionan los resultados de la propuesta de automatización.
- (vii) Capítulo 7: Se podrá encontrar las conclusiones del proyecto.



## Capítulo 2. Marco Teórico

---

### 2.1. Introducción

La creciente adopción de tecnologías digitales en salud ha marcado un hito significativo en la manera de abordar la atención médica a nivel global. En la Estrategia Mundial sobre Salud Digital 2020-2025 se habla de tecnologías digitales, en la Agenda 2030 para el Desarrollo Sostenible se enfatiza la expansión de las tecnologías de la información y las comunicaciones e interconexión en el campo de la salud y, además, en la Estrategia Nacional de Salud para los Objetivos Sanitarios al 2030 la importancia de la información en salud es central. Esta evolución hacia la transformación digital utilizando tecnologías puede mejorar la atención centrada en las personas, ampliar los conocimientos basados en la evidencia, además de las habilidades y competencias de los profesionales de la salud [13].

Según los lineamientos de la Organización Mundial de la Salud (OMS), la información en salud es una de las seis funciones esenciales del sistema de salud, refiriéndose a establecer un sistema de información capaz de generar, analizar y dispersar datos, ayudando a la toma de decisiones y gestión efectiva del sistema de salud. Para 2030 en la Estrategia Nacional para Objetivos Sanitarios, el objetivo es contar con Sistemas de Información (IS) de salud sostenibles, suficientes, integrados y alineados con las exigencias del sector. Esto surge de la insuficiente interoperabilidad entre los distintos actores, estándares insuficientes, datos limitados, infraestructura tecnológica y recursos insuficientes e inadecuados [14]. En resumen, los sistemas de salud deben tener en cuenta la eficiencia y eficacia de sus recursos para brindar un servicio de calidad continuo y oportuno con tal de promover, restablecer o mantener la salud de quien recurre a ellos.

En los servicios del sistema de salud está el recurso de la cita médica, fundamental en el tratamiento y seguimiento de patologías. Un sistema de agendamiento debe entregar una programación en tiempo razonable y conveniente para no agravar ni retrasar las atenciones de los tratamientos demandados, en especial de episodios urgentes que lo requieran. A pesar de su importancia, hay centros de salud con alta demanda estocástica y recursos humanos limitados que no aprovechan herramientas que podrían mejorar un agendamiento manual, a menudo inoportuno y poco intuitivo.

Así, lo que mencionan los documentos de la Estrategia Mundial y Nacional de Salud sobre lo que se quiere alcanzar, se relaciona con el contexto específico del Hospital Las Higueras al identificarse un área de oportunidad en el proceso de agendamiento de episodios del Servicio de Medicina Interna, concordando en la ayuda para tomar decisiones y gestionar el sistema de salud mencionado en las Estrategias.

En este caso se usará la epicrisis médica como base de información ya que provee los conocimientos relevantes para llevar a cabo el proyecto. Según el Glosario de términos frecuentes de la Superintendencia de Salud [15], es el documento que emite el establecimiento de salud cuando el paciente recibe el alta o egresa de una internación u hospitalización, correspondiendo a un resumen de su historia o antecedentes clínicos y de las medidas terapéuticas adoptadas durante la permanencia en el recinto.

En el Hospital Las Higueras, el proceso objetivo se basa en la generación de epicrisis en formato físico, documento con secciones en texto libre, que incluye información sobre controles futuros a los que los pacientes deben asistir para seguir su salud. Sin embargo, la coordinación y consolidación de estas citas presenta desafíos significativos. Entre estos, que el personal de enfermería técnica (TENS) y el administrativo del área clínica usan un sistema manual sin normalización del lenguaje y que no emplea herramientas tecnológicas ideales para asignar citas, situación que conlleva una coordinación ineficiente y priorización inadecuada. Resultando en una experiencia subóptima para los usuarios del sistema de salud.

## **2.2. Antecedentes del Proceso de Agendamiento de Citas Médicas en Hospital Las Higueras**

En el ámbito hospitalario, el proceso de agendamiento de citas médicas es crucial para garantizar una buena coordinación y atención médica.

En el Hospital Las Higueras de Talcahuano, Servicio de Medicina Interna, la toma de horas para episodios que requieren consultas posteriores en policlínicos, comienza el día del alta cuando el TENS toma posesión de la epicrisis en formato físico y se dirige a hacer entrega de esta al personal del mesón del área correspondiente señalada en el mismo documento [16]. Una vez hecho esto el personal lee en la epicrisis las indicaciones de egreso donde se indica el policlínico y temporalidad para el control post-alta para observar las disponibilidades de fechas en las hojas de cálculo en donde se manejan las agendas, ver ejemplo en la Tabla 2.1. El paciente puede quedar anotado en la agenda

provisoria, es decir que queda en la agenda, pero luego se debe llamar para confirmar esa hora o, ser anotado en lista de espera cuando en la agenda ya no existe disponibilidad.

**Tabla 2.1. Ejemplo de registro para el agendamiento de horas**

Estado de res- puesta	Fecha para la consulta	Nombre	RUT	Fecha del alta	Exámenes	Prioritario
PENDIENTE	09/11	Nombre1 Apellido1	12.345.678- 1	01/10	NO RE- QUIERE	
ENTREGADA	07/11	Nombre2 Apellido2	12.345.678- 2	01/10	NO RE- QUIERE	PRIORITARIO
NO CON- TESTA	07/11	Nombre3 Apellido3	12.345.678- 3	10/10	RE- QUIERE	

**Fuente:** Elaboración propia y [17].

Cada día 20 aproximadamente de cada mes, se actualizan las listas de espera y se modifica la agenda para que quede la agenda definitiva de cada uno de los profesionales de los policlínicos. Se debe esperar hasta esa fecha para saber qué día se agendarán las consultas médicas o si es que quedan pendientes para el mes siguiente. Si el paciente queda en lista de espera, se anota con prioridad para que sea agendado entre los primeros para la siguiente oportunidad de agendamiento.

Los policlínicos presentes en la interconsulta del Servicio de Medicina Interna del Hospital Las Higueras son Medicina Interna, Gastroenterología, Reumatología, Cardiología, Endocrinología Adulto, Diabetes, Nefrología, Broncopulmonar, Infectología, Inmunología, Geriátría, Hematología y Oncología.

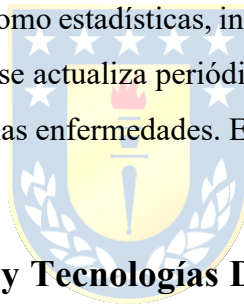
### 2.3. Epicrisis Médica

Este es un documento clave en el agendamiento post-alta. Como menciona el documento de la Superintendencia de Salud, la epicrisis es un documento proveniente de la ficha clínica que recoge información detallada sobre la hospitalización y recomendaciones de seguimiento del paciente una vez se le da el alta médica. Es escrita por profesionales de la salud y juega un rol fundamental en la planificación de consultas posteriores al alta, ya que contiene la información sobre medicamentos, cuidados o controles posteriores.

La periodicidad o temporalidades de los controles post-alta en el Servicio de Medicina Interna del Hospital Las Higueras son especuladas por los mismos profesionales del área del policlínico respectivo, es decir, la dejan mencionada en la ficha clínica. Al momento de emitir la epicrisis, el médico tratante o personal encargado revisa la ficha y completa la sección “Indicaciones del alta”, subsecciones “Conjunto de indicaciones, medicamentos y controles”, “Próximos controles” junto a la sección “Indicaciones de egreso” en la epicrisis. Las fechas que indican los profesionales (4 días, 2 semanas, 3 semanas...) es la que ellos estiman conveniente para realizar un control posterior en el policlínico que respecte al caso.

## **2.4. CIE-10**

La Clasificación Internacional de Enfermedades (CIE) clasifica y agrupa las enfermedades y otros problemas de salud utilizando una codificación específica. Cada código representa una enfermedad o condición específica y se agrupan en categorías más amplias [18]. La CIE se utiliza en diversos contextos del ámbito de la salud como estadísticas, investigación, gestión en servicios de salud, seguros médicos y educación [19]. Esta se actualiza periódicamente para mantener al día los conocimientos médicos y posibles cambios de las enfermedades. En este contexto, CIE-10 hace referencia a la décima edición.



## **2.5. Sistemas de Información y Tecnologías Digitales en el Ámbito de la Salud y Atención Médica**

Es fundamental comprender los Sistemas de Información (IS) y tecnologías digitales que pueden intervenir en el proceso de agendamiento de citas médicas. Estos ya fueron mencionados en la Estrategia Nacional de Salud para los Objetivos Sanitarios al 2030 y, conforme a los lineamientos de la Organización Mundial de la Salud (OMS), la información en salud se posiciona como una de las seis funciones esenciales que componen el sistema de salud, refiriéndose a establecer un sistema de ayuda a la toma de decisiones y gestión efectiva del sistema de salud.

### **2.5.1 Sistemas y Tecnología de la Información**

En [20], se resalta que los IS tienen una presencia significativa en diversas áreas, como negocios, finanzas, contabilidad, gestión de operaciones, marketing y recursos humanos, entre otras. El libro argumenta que la aplicación de sistemas de información y tecnologías es crucial para el éxito de

empresas y organizaciones. Los autores enfatizan que las Tecnologías de la Información (IT) desempeñan roles vitales al mejorar la eficiencia y efectividad de los procesos, siendo esenciales en entornos dinámicos y en conjuntos de componentes interrelacionados con objetivos comunes.

Es así como las IT juegan un papel fundamental en la reingeniería de procesos, debido a la velocidad, capacidades de procesamiento de la información y conectividad de las computadoras y tecnologías de Internet. Al poseer estas características es que pueden aumentar la eficiencia, mejoran las comunicaciones y las colaboraciones entre las personas responsables de la operación y gestión de procesos.

Su arquitectura sigue un diseño conceptual que contine componentes para operaciones eficientes en las organizaciones. La Plataforma Tecnológica constituye la infraestructura fundamental, englobando cómputo, comunicaciones, internet/intranets, extranet, sistemas informáticos y software integrado de aplicaciones. En paralelo, los Recursos de Datos comprenden diversos tipos de bases de datos, desde operativas hasta especializadas, como almacenes de datos y bases de datos de Internet/intranet, destinadas a almacenar y proveer información para respaldar procesos comerciales y decisiones.

Según lo mencionado, la Tecnología de la Información puede actuar como el cimiento sobre el cual se construye y se impulsan diversos algoritmos, proporcionando herramientas para la manipulación y el análisis de datos a gran escala, como lo son la recopilación, almacenamiento y procesamiento eficiente de grandes conjuntos de datos.

### **2.5.2 Aprendizaje Automático**

Comenzando con el Aprendizaje Automático, o Machine Learning (ML), este constituye una rama de la inteligencia artificial y ciencias de la computación que se enfoca en el uso de datos y algoritmos para imitar el aprendizaje del ser humano. En esencia, permite que las máquinas aprendan autónomamente a partir de los datos, mejorando progresivamente su rendimiento en el tiempo [21].

Existen dos categorías principales de aprendizaje automático: el aprendizaje supervisado y el aprendizaje no supervisado. En el aprendizaje supervisado, el modelo se capacita y genera una función a partir de datos de entrenamiento etiquetados, lo que implica conocer el resultado deseado para cada conjunto de datos, el objetivo es poder predecir el valor de cualquier atributo a partir de los datos entregados. En cambio, en el aprendizaje no supervisado, el modelo se entrena con datos no etiquetados, donde no se tiene información previa sobre el resultado deseado para ningún dato en particular e

implica que muchas veces se trabaja con variables aleatorias. Cabe mencionar que existe una categoría intermedia denominada aprendizaje semisupervisado el cual requiere solamente una parte del set de datos etiquetado [22] y otra categoría de aprendizaje por refuerzo [23].

El ámbito de aplicación del ML puede ser diverso ya que puede abarcar desde el reconocimiento de imágenes y el procesamiento del lenguaje natural hasta el diagnóstico médico y la conducción autónoma [24].

En el ámbito de salud podemos nombrar su uso para predecir mortalidad intrahospitalaria, readmisión de pacientes, diagnóstico de enfermedades u otras problemáticas. Los datos administrativos también pueden ayudar a mejorar y optimizar la gestión hospitalaria como predicción de inasistencias de pacientes, organización de horas médicas y pabellón, etc. [25].

#### ***A. Sistemas Basados en Conocimiento***

Dentro de estos sistemas se encuentran los sistemas basados en reglas (SBR) o RBML por sus siglas en inglés, utilizados cuando se desean tomar decisiones o procesar información de forma consistente y repetible. Se utilizan cuando se tiene un conjunto de reglas lógicas, reglas determinísticas y conocimientos previos, de esta manera se asimila el pensamiento humano en la resolución de los problemas, por esta razón en un área de estudio relevante en ML [26]. Las reglas corresponden a sentencias if-then (si-entonces), lo que quiere decir que, si pasa un acontecimiento específico, se debe ejecutar una determinada acción.

En este tipo de sistemas se tiene una base de conocimientos en el que se tiene un dominio del tema en cuestión. En [27] se habla de que lo esencial es el conocimiento específico de alta calidad del que está formada la base de conocimientos del sistema. El conocimiento se adquiere por dos vías principales: el conocimiento factual, documentado en la literatura, y el heurístico, basado en la experiencia y el juicio.

### **2.5.3 Minería de Datos**

Es la Minería de Datos (DM) que se utiliza a menudo para modelos de ML. La tecnología de adquisición y el almacenamiento de datos aumenta notablemente lo que provoca bases de datos de tamaño significativo. Esto está presente en toda intervención humana, como datos de transacciones de supermercados, registros de tarjetas de crédito, llamadas telefónicas, incluso imágenes astronómicas, bases de datos moleculares y expedientes médicos. Por ello, cuanto más se exploten y analicen

estos datos, mayor será la información valiosa que se podrá obtener para quien los posee. La disciplina relacionada con la tarea de extraer esta información relevante de los datos es llamada minería de datos [28].

Esta información puede usarse para mejorar la toma de decisiones, identificar patrones y tendencias, y predecir resultados permitiendo que se enfoque en aplicaciones incluidas la ciencia, la ingeniería, los negocios y/o la salud. En salud, el uso de registros médicos y datos administrativos es una fuente valiosa de información; por ejemplo, en las atenciones clínicas de los pacientes, los registros clínicos contienen el conocimiento del profesional de salud, fundamental para desarrollar herramientas para las bases de estos datos.

El texto es uno de los tipos de datos más comunes de las bases de datos. La minería de datos específica para textos se denomina minería de textos, la cual consiste en analizar colecciones textuales para capturar conceptos clave, tendencias y relaciones entre palabras transformando texto no estructurado en formato estructurado para identificar patrones significativos e información nueva. Los datos estructurados, como nombres y direcciones, se presentan en un formato tabular con filas y columnas, facilitando su almacenamiento y análisis para ML. Los datos no estructurados carecen de un formato predefinido, abarcando texto de redes sociales o multimedia.

Las herramientas de esta práctica permiten transformar estos corpus de textos para habilitar el análisis y generación de información de calidad, que normalmente consta de un previo preprocesamiento donde se realiza identificación del lenguaje, tokenización, etiquetado y el posterior análisis, con el objetivo y fin de obtener mejores resultados [29].

### ***A. Procesamiento de Lenguaje Natural***

El Procesamiento del Lenguaje Natural (PLN), también conocido como Natural Language Processing (NLP), se dedica al análisis de la estructura y gramática de los textos posibilitando la síntesis de textos de manera concisa y coherente para el lector. Este campo aprovecha métodos de inteligencia artificial, ciencia informática, lingüística y ciencia de datos para que las máquinas comprendan el lenguaje humano y puedan manipularlo, permitiendo la categorización o clasificación de textos en categorías predefinidas.

Obtener acceso a datos de texto clínico representa un desafío debido a la naturaleza privada de la información médica. Por esta razón, es crucial establecer colaboraciones entre el ámbito académico y la industria de la salud. Estas colaboraciones son necesarias para avanzar en el campo del NLP

clínico, buscando métodos que permitan extraer información relevante sin comprometer la privacidad de los pacientes.

### ***B. Expresiones Regulares***

También conocidas como REGEX, son utilizadas para la búsqueda en texto, al ser patrones de secuencias de caracteres, permiten encontrar coincidencias dentro de estas cadenas o texto [30]. Esto las convierte en herramientas valiosas para buscar y manipular información específica en cualquier contexto que implique el procesamiento de texto. Su utilidad se ha demostrado en varios campos y su uso continúa expandiéndose hoy [31].

### ***C. Medidas de distancia***

Estas son herramientas fundamentales en el análisis de los textos para poder cuantificar la similitud o diferencia entre cadenas de caracteres.

La distancia de Levenshtein es una de ellas que calcula el número mínimo de operaciones necesarias para transformar una cadena en otra, lo que es útil para la comparación y corrección ortográfica de texto. Además, ha desempeñado un papel central en la alineación de secuencias y en la búsqueda de similitud en bases de datos biológicas. Se ha utilizado en la historia de la bioinformática y continúa siendo relevante en el presente, contribuyendo al desarrollo de su empleo en software [32].

## **2.5.4 Heurística**

La heurística es un enfoque de resolución de problemas que se basa en el uso de reglas prácticas, métodos aproximados o estrategias de búsqueda para encontrar soluciones satisfactorias y razonables donde no siempre se garantiza lo óptimo, pero son útiles para no realizar búsquedas exhaustivas de todas las soluciones posibles [33]. Estos algoritmos pueden ayudar a proporcionar soluciones más rápidas y aceptables cuando se necesite eficiencia en el tiempo de la resolución del problema, como es la optimización enfocada en el área médica [34].

## **2.5.5 Lenguajes de programación**

Existe una amplia variedad de lenguajes de programación, cada uno diseñado con propósitos específicos. Entre los más populares se encuentra Python, un lenguaje de alto nivel (fácil de leer y escribir para el humano) de código abierto orientado a objetos, que no necesita ser compilado y que se compara con Perl, Ruby, Scheme o Java. Python destaca por su sintaxis que facilita la lectura de

programas, bibliotecas estandarizadas y un modo interactivo que simplifica la prueba de extractos de código. Además, es versátil, ya que puede integrarse en aplicaciones para proporcionar interfaces programables en diversos sistemas operativos [35].

Para llevar a cabo algunas implementaciones de DM, se emplean librerías especializadas de Python como SpaCy y pandas. Hay muchas más, pero en general estas dos son ocupadas para la primera parte de preprocesamiento de los datos. La primera es una librería de procesamiento de NLP de código abierto que proporciona herramientas y recursos para el procesamiento automatizado de texto, permitiendo tareas como Tokenización, lematización, NER, entre otros [36]. Pandas es una librería de código abierto que proporciona estructuras de datos flexibles y herramientas para el análisis y manipulación de datos, como limpieza y preparación, exploración y visualización, y procesamiento de información en general [37]. Es especialmente útil para trabajar con conjuntos de datos tabulares.

## 2.5.6 Google

### A. *Google Workspace*

Google Workspace es un entorno integral de herramientas de productividad y colaboración basado en la nube ofrecido por Google dirigido a empresas, estudiantes, organizaciones sin fines de lucro y particulares, con el propósito de facilitar la organización, colaboración y comunicación en estos [38]. Este conjunto de herramientas incluye correo electrónico, calendario, video llamadas y conferencias, así como almacenamiento en la nube y editores de documentos, hojas de cálculo y presentaciones. Además, ofrece funcionalidades para la comunicación interna, formularios y administración, todo ello con características en tiempo real que brindan flexibilidad y seguridad para los datos de los usuarios. Entre las medidas de seguridad se incluye el cifrado de datos y la protección contra el malware. Google Workspace es escalable para adaptarse a cualquier tamaño de organización o empresa, con opciones de precios ajustables a las necesidades individuales [39].

Esta plataforma alberga sus aplicaciones en donde más de 3 mil millones de usuarios disfrutan de sus herramientas integradas e intercomunicadas, dentro de este conjunto se encuentran:

- Herramientas de almacenamiento y gestión de archivos: Google Drive
- Herramientas de edición de documentos: Google Docs
- Herramientas de edición de hojas de cálculo: Google Sheets
- Herramientas de edición de presentaciones: Google Slides

- Herramientas de comunicación y colaboración: Gmail, Calendar, Meet, Chat [40]

Además de esto, este conjunto de aplicaciones ofrece la capacidad de integrar otras aplicaciones personalizadas o permitir que estas personalizadas accedan a archivos dentro de Google Workspace, como aquellas creadas con secuencias de comandos que utilicen la API de Google Drive. Los usuarios pueden emplear estas aplicaciones para una amplia gama de actividades, como la edición de imágenes y videos, la firma y envío de documentos por fax, la gestión de proyectos, la creación de diagramas de flujo, y más. Asimismo, es posible gestionar si los miembros de la organización pueden utilizar la API de Drive para acceder de manera programática al contenido de Drive [41].

### ***B. Google Cloud***

Otro de los servicios proporcionados por Google es Google Cloud, que incluye infraestructura, plataformas y software para computación, almacenamiento de datos, análisis de datos y aprendizaje automático. Mientras que Google Workspace proporciona herramientas de productividad y colaboración esenciales para el día a día, Google Cloud ofrece la infraestructura y las capacidades avanzadas de computación necesarias para esas herramientas, donde hay veces en que ambos trabajan de forma colaborativa [42].

Dentro de Google Cloud se encuentra AppSheet, plataforma de desarrollo sin código que permite a los usuarios crear aplicaciones móviles y web sin necesidad profunda de conocimientos de programación. Se integra perfectamente con Google Workspace y los servicios de Google Cloud, permitiendo automatizar flujos de trabajo y mejorar la productividad utilizando datos almacenados en diversos servicios en la nube. AppSheet funciona permitiendo a los usuarios definir la lógica de la aplicación, las reglas del sistema y los flujos de trabajo a través de una interfaz gráfica, sin necesidad de escribir código extenso. La plataforma convierte estas configuraciones en aplicaciones funcionales que pueden ejecutarse en dispositivos móviles y web, facilitando la automatización y gestión de procesos [43].

## **2.6. Estimador de Horvitz-Thompson para la Media en Distribución Binomial Negativa**

Este es un estimador general que sirve para estimar parámetros poblacionales lineales, como el total o la media poblacional. Es útil en situaciones en las que se conoce la probabilidad de selección de cada unidad de la población y no es igual para todas las unidades [44]. Y la distribución binomial

negativa consideremos pruebas Bernoulli independientes con la misma probabilidad de éxito en cada una de ellas. Fijando el número de éxitos que queremos obtener, se realizan cierto número de pruebas adicionales hasta obtener el éxito  $n$ -ésimo [45].

En cuanto a tamaño muestral, los algoritmos de ML no suelen calcular un tamaño muestral previo, ya que se trabaja con los datos disponibles. Lo anterior puede causar que los algoritmos se sobre-ajusten y no puedan generalizar. La estimación del tamaño mínimo de muestra para estudios con distribución binomial negativa se puede realizar mediante el estimador de Horvitz-Thompson, una herramienta estadística que permite obtener una estimación no sesgada de la media poblacional a partir de datos muestrales y las probabilidades de selección de las unidades en la población. Este método resulta útil cuando se dispone de información limitada sobre una población y se desea conocer su media, cuando la variable de interés sigue una distribución binomial negativa. Esto consiste en las siguientes definiciones y notaciones:

- $N$ : Tamaño total de la población.
- $n_h$ : Tamaño del estrato  $h$ .
- $y_i$ : Valor observado para las variables de interés de la unidad  $i$ .
- $\bar{Y}_h$ : Media del estrato  $h$ .
- $t_h$ : Total poblacional del estrato  $h$ ,  $t_h = n_h \cdot \bar{Y}_h$ .



Estimador de Horvitz-Thompson para la media en un estrato:

$$\mu_h = \frac{t_h}{n_h}$$

Estimador de Horvitz-Thompson para la media poblacional:

$$\mu_h = \sum_h w_h \cdot \mu_h$$

donde  $w_h$  es el peso del estrato  $h$ :

$$w_h = \frac{N - n_h}{N \cdot n_h}$$

Respecto a la aplicación a la distribución binomial negativa, se presenta de la siguiente manera:

- $r$  : Número de éxitos requeridos.
- $p$  : Probabilidad de éxito en el ensayo.

La media del estado  $h$  en una distribución binomial negativa es  $E_h[Y] = \frac{r}{p}$ , el total poblacional es:  
 $t_h = n_h \cdot E_h[Y] = \frac{r \cdot n_h}{p}$ .

El estimador de Horvitz-Thompson para la media el estrato  $h$  :

$$\mu_h = \frac{r}{p}$$

El estimador de Horvitz-Thompson para la media poblacional en una distribución binomial negativa:

$$\mu = \sum_h w_h \cdot \mu_h = \sum_h w_h \cdot \frac{r}{p}$$

Entonces, se supone que se tiene una población dividida en  $H$  estratos en donde es de interés el estimar la media de una variable de interés que sigue una distribución binomial negativa, esta consiste en un modelo adecuado para tratar aquellos procesos en los que se repite un determinado ensayo o prueba hasta conseguir un número determinado de resultados favorables (por vez primera) [46]. Se denotan los parámetros de interés como  $r$  (número de éxitos requeridos en un ensayo) y  $p$  (probabilidad de éxito en un ensayo).

En el diseño de muestreo estratificado, el tamaño total de la población es  $N$  y el tamaño del estrato  $h$  es  $n_h$ . Se define  $y_i$  como el valor observado para la variable de interés en la unidad  $i$ .

El estimador de Horvitz-Thompson para la media en el estrato  $h$ ,  $\mu_h = \frac{t_h}{n_h}$ , donde  $t_h = n_h \cdot E_h[Y]$ , y  $E_h[Y]$  es la media del estrato  $h$ . Para una distribución binomial negativa,  $E_h[Y] = \frac{r}{p}$ , por lo que  $\mu_h = \frac{r}{p}$ .

El peso del estrato  $h$  en un diseño de muestreo estratificado es:  $w_h = \frac{N - n_h}{N \cdot n_h}$ .

El estimador de Horvitz-Thompson para la media poblacional  $\mu$  se obtiene ponderando los estimadores de los estratos según sus pesos:

$$\mu = \sum_h w_h \cdot \mu_h$$

Para la distribución binomial negativa, esto se simplifica a:

$$\mu = \sum_h w_h \cdot \frac{r}{p}$$

Por lo tanto, el cálculo del tamaño mínimo de muestra para un diseño de muestreo estratificado con distribución binomial negativa y el estimador de Horvitz-Thompson indica la consideración del nivel de confianza y la precisión deseada. Se utiliza la fórmula del error estándar para el estimador y se resuelve para el tamaño de la muestra:

$$n_h = \frac{r \cdot N \cdot w_h}{p \cdot (1 + p \cdot r) \cdot se^2}$$

donde  $se$  es la precisión deseada y se calcula a partir del nivel de confianza.



## Capítulo 3. Revisión Bibliográfica

---

### 3.1. Introducción

La producción de datos a nivel mundial aumenta día a día, y el ámbito de la salud no es una excepción. La ficha clínica electrónica, también conocida como Registro Electrónico de Salud, es un documento donde se registran los antecedentes relacionados con la salud de las personas [1]. Estos registros, al igual que la epicrisis que deriva de ella, quedan almacenadas en hospitales, centros de salud comunitarios, clínicas u otros. Dada la gran cantidad de documentación por paciente, las técnicas de procesamiento de datos o de ML pueden ser utilizadas para extraer información relevante y con el potencial necesario para crear conocimiento y ayudar a los profesionales que trabajan en el área de la salud [2].

El campo del ML ha tenido un notable progreso en el campo del apoyo a las decisiones clínicas, principalmente por los avances en recursos computacionales y la disponibilidad de datos extraíbles, pero se estima que aproximadamente el 80% de estos datos están en formato no estructurado o presentado en forma de texto libre, esta modalidad se prefiere porque facilita la comunicación en lenguaje natural, siendo una forma más expresiva de documentar los eventos clínicos [47]. El problema es la falta de estructura, errores gramaticales y uso de dialectos. Es por esto por lo que es de mucha ayuda procesar los datos para que sean precisos, completos y consistentes [48], adicionando que se puede mejorar la precisión y la transmisión de información específica en comparación con los espacios estructurados, que a veces no permiten este nivel de detalle.

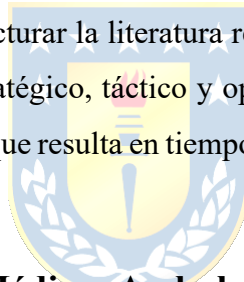
### 3.2. Introducción al Agendamiento y Programación de Citas en el Ámbito Médico

En 1952 se comienza a utilizar el término “scheduling” o programación en español y se inician los estudios sobre cómo funcionan los sistemas de citas médicas. El responsable de esto fue Norman T. J Bailey, el primero en analizar el sistema de agendamiento y proponer una recomendación para el procedimiento de citas de pacientes (el cual presentaba problemas de tiempo), que consistió en que los pacientes asistan cada cierto tiempo y por intervalos a las consultas [49].

Otra visión general de los desafíos y oportunidades de los sistemas de programación fue abordada en [50] para clínicas de atención primaria y especializada. Presentaron una organización metódica y jerarquizada de factores que dificultaban la programación de citas en el entorno de la retención

médica y mencionaron que la prioridad de garantizar un acceso rápido y oportuno, adaptando reglas de acceso y programación anticipada, es esencial para optimizar el tiempo, y que considerar la variabilidad en la urgencia de los pacientes permite un manejo efectivo de la oferta y demanda de citas. La capacidad para acomodar flujos de pacientes entre clínicas especializadas puede inspirar estrategias para asignar eficientemente citas médicas y la colaboración entre proveedores de atención médica una perspectiva colaborativa para abordar la demanda variable en áreas específicas. Estos principios proporcionan relevancia para mejorar la eficiencia operativa y satisfacción de los pacientes en el ámbito de citas médicas.

El problema de agendamiento de citas médicas en sistemas complejos como lo son los servicios de salud es revisado por Ala y Chen en [51] donde pudieron examinar más de 150 artículos científicos caracterizados según varias especificaciones de problemas presentes en estos servicios; como el flujo de pacientes, su llegada a servicios aleatorios y publicaciones relacionadas con la programación de citas médicas publicadas entre 2000 y 2021 en la base de datos Web of Science Core Collection. Ellos seleccionaron varios criterios para estructurar la literatura reciente que aborda los problemas de programación de citas médicas a nivel estratégico, táctico y operativo, para proponer varios métodos y reducir el tiempo de espera del paciente que resulta en tiempos de inactividad más cortos en los centros de salud.



### **3.2.1 Agendamiento de Citas Médicas Ambulatorias**

Respecto a citas ambulatorias Fuenzalida Del Valle en [2] propone una solución de corto plazo que consiste en la categorización de pacientes dependiendo de su gravedad en el Hospital del Salvador. Esta propuesta viene a abordar las extensas listas de espera que superan la oferta y capacidad de especialidades o tratamientos por la alta demanda sostenida en el tiempo de estos mismos. Fuenzalida Del Valle utiliza un Proceso de Decisión Markoviano para escenarios específicos de agenda y hace referencia al proceso de Poisson, que modela la frecuencia de eventos durante algún intervalo fijo de tiempo [52] en el contexto del estudio con el que concluye aumentar horas médicas para ciertas especialidades del hospital.

Otro acercamiento al agendamiento de citas ambulatorias se ve en [53] en el cual los autores abordan el desafío de programar pacientes de manera secuencial considerando las ausencias y diferentes grupos de pacientes. Para eso proponen una programación de citas en tres pasos: obtención de datos del paciente, limpieza de datos y clasificación del riesgo de ausencia mediante algoritmos de

aprendizaje automático a partir de datos de la ficha electrónica y fuentes externas. Finalmente, proponen y evalúan reglas de programación de citas basadas en las predicciones entregadas por algoritmos de aprendizaje automático.

### **3.3. Tecnologías de Inteligencia Artificial para Agendamiento de Citas Médicas**

#### **3.3.1 Extracción de Información Clínica**

El estudio en [54] se desarrolló un algoritmo llamado Regular Expression Discovery Extractor (REDEx), entrenado en un corpus de notas de atención primaria ambulatoria. El objetivo fue extraer medidas de peso corporal desde las notas clínicas del Registro de Salud Electrónico de la Administración de Veteranos (VA), aproximadamente 2 mil millones de documentos clínicos de proveedores con acceso a datos médicos estructurados y no estructurados de 17 millones de veteranos. El estudio no especifica las librerías o tipo de computación utilizados para implementar REDEx, sin embargo, menciona haber desarrollado un módulo de procesamiento NLP para la extracción de información sobre peso corporal, permitiendo estimar trayectorias en datos de signos vitales importantes y reducir posibles sesgos por datos faltantes o mal temporizados. Los resultados resaltan la efectividad de REDEx para extraer datos desde las notas clínicas, lo que tiene implicancias importantes para la atención clínica, epidemiología y esfuerzos de mejora de calidad.

Otro enfoque se da en [55] donde minimizan el tiempo de espera para pacientes ambulatorios con mayor prioridad utilizando Machine Learning y Data Mining ocupando prioridades “alta” y “baja”. Su algoritmo de Modelo de Aprendizaje Supervisado (MAS) y No Supervisado, propone la utilización de K-means Clustering y agrupamiento jerárquico aglomerativo para los pacientes de clases prioritarias, además de utilizar el algoritmo Wrapper para poder encontrar las mejores características para emplearlas en el entrenamiento de clusters. Un Proceso de Decisión de Markov o Markoviano (MDP) también fue utilizado obteniendo finalmente la priorización y programación efectiva de los pacientes en el Centro Médico y de Investigación Shaheed Rajaei en Irán.

Un alcance más que menciona el aprendizaje automático fue en [56] donde se propone extracción de datos estructurados a partir de otros que no lo están, específicamente registros médicos relacionados con partos prematuros con el propósito de predecir la probabilidad de que ocurra un parto prematuro en intervalos de tiempo específicos. Para ello, se identificaron conceptos, entidades, eventos y las relaciones o atributos que comparten, mediante el uso de herramientas de NLP, en archivos

de literatura médica que suelen centrarse en áreas como enfermedades, fármacos u optimización de flujos de trabajo. SpaCy fue una de las bibliotecas más utilizadas por su popularidad, rapidez, facilidad de uso y rendimiento en NLP. Asimismo, se añadieron tres módulos: de-identificación, vinculación de entidades y análisis semántico. En este proceso, se emplea Aprendizaje Automático, específicamente CatBoost, un método que se basa en árboles de decisión de Machine Learning impulsados por gradientes. Esta elección se justifica por su buen rendimiento con datos tabulares de alta dimensión y con muchas interacciones entre características. Además, establece parámetros de manera heurística en lugar de depender de valores predeterminados estáticos.

Adicionando la combinación de técnicas de minería de textos y aprendizaje automático para el análisis de textos médicos, se realiza en [57] una implementación y simulación realista de flujo de pacientes con patologías de Síndrome Coronario Agudo (SCA). Para ello utilizaron bibliotecas de Python, método K-means con distancia de Levenshtein, clustering y árbol de decisiones, mostrando una simulación de la duración de la estadía en hospitales de pacientes con SCA obtenida de las fichas clínicas de pacientes del Centro Nacional de Investigación Médica Almazov en San Petersburgo, Rusia.

Respecto a textos médicos en [58] podemos ver a los autores enfocarse de igual forma en la extracción de información clínica, pero de los resúmenes de alta de pacientes utilizando técnicas de extracción de información automatizadas. Su objetivo fue comparar el rendimiento de 5 sistemas de anotación clínica de código abierto entre estas MedTagger, Gate, cTAKES, NCBO Annotator y Clamp para extracción de entidades médicas y así destacar la importancia de la extracción automatizada de información clínica para mejorar la eficiencia y calidad de la atención médica. Para obtener buenos resultados utilizaron 108 resúmenes de alta de pacientes obtenidos del sitio web mtsamples para evaluar las extracciones. Para trabajos futuros mencionan la potencialidad del ámbito de la extracción de entidades de contextos clínicos.

Una propuesta actual reciente del año 2023 en [59] introduce MedTem2.0, un sistema diseñado para la clasificación temporal de eventos de tratamiento en resúmenes de alta médica mediante el uso de prompts. Su principal objetivo fue automatizar la extracción y recopilación de la lista de medicamentos de un paciente, tarea laboriosa que consume mucho tiempo. La implementación de MedTem2.0 se basó en métodos avanzados de NLP, aprovechando el aprendizaje basado en prompts en modelos como los Generative Pretrained Transformers (GPTs) y el ajuste fino en modelos de lenguaje pre entrenados (PLMs) como BERT. Estos enfoques se aplican para clasificar las relaciones

temporales entre los tratamientos y los periodos de hospitalización en los resúmenes de alta médica. El artículo también aborda los desafíos asociados con extracción y clasificación de eventos temporales en textos clínicos, explorando el potencial del aprendizaje basado en prompts en el ámbito del procesamiento del lenguaje natural clínico. En la metodología, se describe el proceso de etiquetas manuales a partir de conjuntos de datos anotados, seguido de diversos pasos de preprocesamiento, como aprendizaje de pocos ejemplos y segmentación de oraciones. Se utilizaron bibliotecas de procesamiento del lenguaje natural, incluyendo Hugging Face Transformers, PyTorch y TensorFlow. Finalmente, el sistema MedTem2.0 se implementó en una máquina virtual de Google Cloud para su ejecución.

### 3.3.2 Implementación de Google Workspace en el ámbito médico

A pesar de que Google Workspace no se considera una herramienta de aprendizaje automático, sí es posible integrar funcionalidades de ML en sus aplicaciones creadas con esta plataforma [60], como en el artículo [59] mencionado en el párrafo anterior.

En el ámbito médico, esta herramienta ha sido utilizada para digitalizar tareas manuales y repetitivas como la introducción de datos y la programación, con el fin de evitar el agotamiento, ayudando a médicos y empleados a maximizar el tiempo que dedican a sus pacientes.

En Hunterdon, Nueva Jersey, se encuentra Hunterdon Healthcare, un sistema de salud con un hospital que ofrece servicios preventivos, diagnósticos y terapéuticos tanto hospitalarios como comunitarios para pacientes hospitalizados y ambulatorios. El centro médico Hunterdon Medical Center, trata a más de 8,600 pacientes hospitalizados anualmente y ha adoptado herramientas de colaboración basadas en la nube que ayudan a los profesionales médicos y de IT de la organización a trabajar de manera más rápida e inteligente. Las aplicaciones ocupadas permiten, por ejemplo, al departamento de enfermería automatizar procesos que antes eran manuales para ayudar a los enfermeros a pasar más tiempo con los pacientes, haciendo seguimiento de los incidentes del paciente en Hojas de cálculo, pero compartidas.

Otras organizaciones de asistencia sanitaria y ciencias biológicas que utilizan Google Workspace hoy en día son CommonSpirit, Cambridge Health Alliance, Roche, Hurley Medical Center, National Institute for Health Research y KPJ Healthcare [61].

### 3.3.3 Sistemas Basados en Reglas

En el artículo [62] mencionan que, en el ámbito de los sistemas de citas para pacientes ambulatorios se suelen implementar enfoques basados en reglas, ya que ofrecen buenas soluciones para las decisiones operativas. Ellos presentaron un modelo basado en reglas donde introdujeron criterios de programación basados en diferentes reglas, entre estas la prioridad de los pacientes y el tiempo de espera de los pacientes para manejar el agendamiento de pacientes ambulatorios en un escenario real.

Otro caso donde se ocupa este sistema en salud fue en [63] donde se utilizó para la búsqueda de medicamentos. Al iniciar el proceso, el usuario proporcionaba sus coordenadas geográficas y el nombre de medicamento deseado. El sistema consultaba la base de datos para establecer reglas específicas basadas en la disponibilidad del medicamento y la ubicación de farmacias cercanas para seguir con la verificación de existencias reales del medicamento en estas. La salida del sistema incluía información como el nombre de la farmacia, dirección, coordenadas geográficas y cantidad disponible del medicamento. Finalmente, el sistema utilizó criterios para seleccionar y presentar al usuario las mejores opciones, priorizando la cercanía geográfica y la disponibilidad del medicamento.

### 3.4. Discusión

De la revisión de la literatura del análisis bibliográfico se puede apreciar que existe una constante necesidad que ha perdurado a lo largo de los años de mejorar diversos procesos con diferentes metodologías dependiendo de las situaciones de cada área y caso en particular. Se observa que se proponen soluciones para reducir los tiempos de espera mediante el desarrollo de sistemas de apoyo a la toma de decisiones para la gestión de servicios, y se ejemplifica que los sistemas de atención médica han enfrentado desafíos para ofrecer servicios de alta calidad. El tiempo se presenta como un elemento esencial para garantizar la seguridad y el rendimiento de los pacientes, y juega un papel crucial en su satisfacción. La literatura identifica problemas en el agendamiento de pacientes ambulatorios, admisión de pacientes hospitalizados y asignación de recursos [64]. Dada la complejidad del problema de agendamiento, normalmente se recurre a algoritmos de heurística o aprendizaje automático para encontrar una solución. Dado que este proyecto busca resolver el agendamiento de citas médicas ambulatorias para pacientes dados de alta, se tratará como un subproblema del agendamiento ambulatorio de pacientes.

En este contexto, se identifica la necesidad de perfeccionar los procesos de agendamiento de citas médicas debido a diferentes factores, como el aumento de la demanda de servicios médicos, la

falta de eficiencia en estos mismos y la falta de estandarización en el lenguaje de los documentos de seguimiento. Esto hace que sean áreas de mejora tanto para los sistemas existentes como para la implementación de nuevas formas de operación. Además, se ha observado el uso de técnicas de aprendizaje automático y minería de datos para la clasificación donde autores sugieren la escalabilidad de modelos para instancias mayores.

La literatura proporciona una perspectiva valiosa sobre las prácticas y métodos relacionados con la programación de citas en el ámbito de la atención médica que fortalece y complementa la investigación al proporcionar una revisión de enfoques, modelos, algoritmos de programación y técnicas de inteligencia artificial, consolidando así la base de conocimientos para comprender los retos y las soluciones en la gestión de citas.

Es importante destacar la utilidad del artículo [58], ya que en él se lleva a cabo la extracción de medicamentos a partir de las epicrisis de pacientes. Aunque este enfoque difiere del objetivo del proyecto actual, al utilizar documentos clínicos la información proporcionada en dicho artículo es relevante.

En la literatura revisada, se observa el uso de herramientas de Google, como Google Cloud, que ofrece servicios en la nube para analizar datos, aprovechar la inteligencia artificial y almacenar datos de manera segura. Esto se evidencia en el marco teórico, donde se discute la utilidad de las diversas herramientas de Google Workspace para trabajar con aplicaciones externas.

En cuanto a los lenguajes más utilizados, se destaca el uso de C++ y Python en la literatura revisada. Sin embargo, la mayoría de los estudios no mencionaban la población objetivo, ya que se centraban en mediciones o simulaciones teóricas.

La revisión bibliográfica revela una falta de estudios centrados en el contexto posterior al alta médica, una faceta de gran relevancia para mejorar el sistema en el ámbito hospitalario. Aunque se mencionan aplicaciones para agendamiento de citas, la literatura examinada no da ejemplos específicos destinados a pacientes dados de alta y que necesitan programar una cita después.

Asimismo, la ausencia de investigaciones que aborden la extracción de la prioridad del paciente directamente desde las fichas clínicas o epicrisis para facilitar su agendamiento en el momento pertinente es un notable vacío identificado en la revisión.

Es fundamental destacar que, a pesar de los avances y las diversas aplicaciones identificadas en la literatura revisada, la mayoría de estos logros y desarrollos relevantes están predominantemente

centrados en el contexto del lenguaje inglés. Esta observación sugiere la necesidad de una mayor atención y dedicación a entornos médicos de habla hispana, subrayando la importancia de considerar particularidades lingüísticas y culturales en el desarrollo de herramientas y sistemas para el sector hospitalario. Este aspecto, por lo tanto, representa un área potencial de investigación y mejora que podría beneficiar la eficacia de los sistemas de agendamiento médico.



## Capítulo 4. Levantamiento de Información

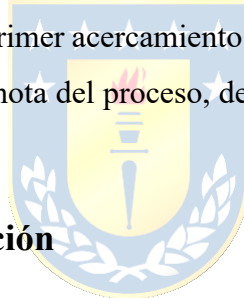
La gestión de citas médicas en el sistema de salud enfrenta numerosos desafíos. La gran cantidad de datos sin estructura generados por los sistemas de información, el aumento de la demanda, la necesidad de mejorar la programación de citas y la limitación de recursos son solo algunos de los obstáculos que deben superarse para mejorar la calidad de atención.

Para mejorar la eficiencia y la eficacia de la gestión de citas médicas, fue necesario realizar un análisis de la situación actual, identificar los principales desafíos e implementar soluciones innovadoras respecto al contexto actual.

A continuación, se mostrarán aspectos clave de la implementación del proyecto.

### 4.1. Introducción a la Problemática

Para el levantamiento de información se realizaron dos videos conferencias con el doctor jefe del Servicio de Medicina Interna para comprender la problemática existente en el agendamiento de controles médicos en su servicio como primer acercamiento. Luego, se llevaron a cabo visitas presenciales al hospital para observar y tomar nota del proceso, desde la preparación del alta hasta la inclusión de los episodios en las agendas.



#### 4.1.1 Recopilación de Información

Previo a las visitas, se crea un documento con una serie de preguntas, ver Tabla 4.1, relacionadas con el agendamiento en la interconsulta de Medicina Interna, las que fueron respondidas por el doctor jefe del Servicio de Medicina Interna y el personal del mesón de la interconsulta.

**Tabla 4.1. Preguntas para el Levantamiento del Proceso**

Pregunta	Respuestas
1. ¿Se pueden pedir citas para consultas ambulatorias en el mesón de interconsulta?	No, las citas solo se les dan a pacientes dados de alta de Medicina Interna o derivados del mismo hospital.
2. Para el agendamiento, ¿hay diferencias entre un paciente ambulatorio y un paciente que estuvo hospitalizado?	No existen diferencias para el agendamiento, se les trata como iguales.

---

3. ¿Hay algún horario especial para los pacientes que estuvieron hospitalizados?	Doctor: Sí, hay diferencias. Secretarias: No hay horarios especiales, quedarán categorizados como "consulta repetida" y eso no cambia en nada.
--	---

---

4. Cuando se agendan a los pacientes, ¿hay algún tipo de priorización?	No.
--	-----

---

**Fuente:** Elaboración propia, [11] y [17].

En las visitas, se recompila información detallada sobre el proceso de agendamiento, las agendas de los doctores en policlínicos y otros aspectos relevantes que ayudarán al proyecto como número de policlínicos y formatos de agendas.



## Capítulo 5. Propuesta de Automatización

---

### 5.1. Funcionalidad del Sistema Automatizado

#### 5.1.1 Extracción de Información de Epicrisis

El sistema automatizado extrae la siguiente información relevante de las epicrisis de los pacientes:

- Edad
- Diagnósticos de Egreso
- Fecha de Egreso
- Días de Hospitalización
- Etapas GES
- Fecha de control post-alta sugerida (si corresponde)
- Policlínico de interconsulta indicado (si corresponde)

Estos datos son los que se utilizan como base para la priorización y clasificación de episodios, ya que, según lo conversado con el doctor jefe del servicio, proporcionan información precisa y adecuada para el objetivo del proyecto. Al ofrecer el estado de salud al momento del alta, facilitan la toma de decisiones médicas informadas para la asignación acertada de recursos médicos.

Las variables se extraen utilizando expresiones regulares debido al formato de texto estructurado. El código contempla la posibilidad de múltiples diagnósticos y etapas GES. Para las últimas dos variables, se extrae la sección de indicaciones de egreso porque, aunque las variables no estén en formato de texto estructurados, la sección sí lo está.

Los policlínicos se extraen de las indicaciones de egreso utilizando distancia de Levenshtein junto a un método de búsqueda con diccionarios, tokens y bigramas. Para el umbral de distancia se considera la existencia de faltas ortográficas o abreviaciones al ingresarlos a las indicaciones en campos narrativos en formato de texto libre. Los diccionarios empleados contienen los nombres de los policlínicos de la interconsulta del Servicio de Medicina Interna, donde se utilizan bigramas para el policlínico de medicina interna (compuesto por dos palabras) y tokens para los demás. La fecha de control post-alta se obtiene utilizando expresiones regulares y diccionarios. Primero, se delimita el texto de indicacio-

nes de egreso buscando palabras clave como "control", "policlínico" o "ambulatorio". Luego, se identifican las expresiones que indican en cuánto tiempo más debe ser el control, estas expresiones de duración incluyen términos como "un", "dos", etc., seguidos de unidades de tiempo como "días", "semanas" o "meses". Para facilitar la lectura, los números escritos en letras se convierten a números (por ejemplo, "tres semanas" se convierte en "3 semanas"). Además, las indicaciones que no mencionan la temporalidad del control post-alta se clasifican como 'según disponibilidad' y se les asigna un valor equivalente a 14 días. Finalmente, todas las temporalidades extraídas se convierten a valores numéricos equivalentes en días para adicionar variables de fechas al código.

### 5.1.2 Clasificación y Priorización de Pacientes

El sistema automatizado prioriza a los episodios y los clasifica según policlínicos de la interconsulta del Servicio de Medicina Interna, considerando los siguientes criterios:

- **Prioridad clínica:**
  - Pacientes con enfermedades con gravedad alta, media o baja: una clasificación preliminar realizada en función de la generalidad de la enfermedad y nivel de gravedad de esta, considerando aspectos de letalidad, capacidad de transmisión, existencia de vacunas, tratamientos efectivos, etc.
  - Pacientes con fecha de controles determinados: corresponde cuando se requiere una atención en una temporalidad específica indicada por el médico.
- **Prioridad demográfica:**
  - Pacientes adultos en grupos etarios: se consideran grupos etarios diferentes ya que a medida que la edad avanza, la salud se vuelve más inestable y se tiende a ser más vulnerable a ciertas enfermedades o complicaciones que otros.
  - Pacientes GES: los que forman parte del sistema de Garantías Explícitas en Salud, lo que les permite el acceso oportuno y de calidad para ciertos contextos en salud.
- **Prioridad histórica:**
  - Pacientes que aún no se les ha podido concretar el agendamiento: cuando no se ha tenido éxito para agendar una cita o aún no se lleva a cabo la coordinación para esta.

Para iniciar el cálculo de la prioridad, se carga un archivo con diagnósticos en formato CIE-10 junto con su gravedad. Para comparar estos diagnósticos con los indicados en la epicrisis, ambos

se limpian y se tokenizan permitiendo implementar distancia de Levenshtein normalizada para identificar coincidencias y seleccionar el diagnóstico correspondiente al CIE-10 identificado del junto a su gravedad.

Antes de proceder a la clasificación por policlínico, se calcula un puntaje de priorización basado en un SBR, aplicando reglas específicas al cumplirse ciertos criterios. Esto incluye las variables categóricas de etapa GES y la urgencia del diagnóstico, a las cuales se les asignan valores numéricos para facilitar cálculos. Adicionalmente, se aplican reglas de divisiones enteras a otras variables involucradas en los cálculos para mejorar la interpretación del puntaje de prioridad usando intervalos claros, sumando los valores resultantes al puntaje final. Se añadió una condición donde los pacientes sin control requerido post-alta en la interconsulta del Servicio de Medicina Interna obtienen un puntaje de prioridad de 0. Estos detalles se pueden ver en la Tabla 5.1.

**Tabla 5.1. Reglas, cálculos y ejemplos de variables para el cálculo del Puntaje de Prioridad**

Variable	Regla y/o Cálculo	Ejemplo
Etapa GES	Mapeo: 'Ninguno' y 'No Aplica' = 0, 'Caso en Proceso' y 'Caso en Proceso Diagnóstico' = 1, 'Caso en Seguimiento', 'Caso en Tratamiento', y 'Caso Confirmado' = 2.	'Caso Confirmado' = 2.
Urgencia/Gravedad	Mapeo: 'Ninguno' = 0, 'BAJA' = 1, 'MEDIA' = 2, 'ALTA' = 3.	'ALTA' = 3.
Edad	Se divide por 10 (agrupando en décadas: 0-9, 10-19, 20-29, etc.) y se suma al puntaje de prioridad.	Edad = 35 años; $35 / 10 = 3.5$ , redondeado a 3.
Días de Hospitalización	Se divide por 7 (agrupando en semanas) y se añade al puntaje.	Días hospitalizado = 14; $14 / 7 = 2$ .
Días desde el Egreso	Se divide por 7 (agrupando en semanas) y se añade al puntaje.	Días desde el egreso = 21; $21 / 7 = 3$ .
Días para Fecha Indicada	Se resta el número de días para la fecha indicada del valor máximo de días para	Máximo de días para fecha indicada entre todos los pacientes = 30,

		la fecha indicada entre todos los pacientes, y luego se divide por 7 para agrupar en semanas, dando mayor prioridad a los pacientes con fechas de control más próximas.	Días hasta fecha indicada = 10; $(30 - 10) / 7 \approx 2.86$ , redondeado a 3.
Control no (condición)	Requerido	Si el policlínico es 'Ninguno', el puntaje de prioridad es 0.	Policlínico = 'Ninguno'; Puntaje de Prioridad = 0.

**Fuente:** Elaboración propia.

Cada una de estas operaciones representa una regla lógica del sistema. Cuando se cumple la condición de existencia de un policlínico, se ejecuta la acción correspondiente para asignar un puntaje específico de prioridad. Estas operaciones matemáticas y transformaciones de datos aplican reglas if-then, formadas a base de conocimiento documentado en la literatura y a la experiencia contextual, para automatizar la priorización.

Posterior a la prioridad, se clasifican los pacientes según policlínico y dentro de estos por el puntaje de prioridad calculado. Con esto finalmente se almacena la información en AppSheet y se visualizan los resultados.

## 5.2. Arquitectura y Diseño del Sistema

### 5.2.1 Arquitectura del Sistema

El sistema automatizado está compuesto por los siguientes módulos que componen la estructura global y organización fundamental del sistema:

- Extracción de información: encargado de extraer la información relevante de la epicrisis, ya sea en formato de texto estructurado o en campos narrativos de texto libre de formato no estructurado.
- Procesamiento de la información: la información extraída se limpia y procesa para obtener lo elemental de esta y lo que es útil para la construcción del algoritmo.
- Módulos de diccionarios: creación y gestión de diccionarios que incluyen los policlínicos de la interconsulta y temporalidades para las citas (días, semanas, meses, según disponibilidad, y urgente son algunas).

- Información complementaria: ingreso de variables CIE-10 y sus clasificaciones preliminares por gravedad (alta, media, baja).
- Priorización de episodios: para asignar el nivel de prioridad numérico a los episodios en función de su contexto médico post-alta. Basado en SBR.
- Clasificación de episodios: en donde se clasifican los episodios en cada uno de los policlínicos y se ordenan según nivel de necesidad de atención medica dependiendo del puntaje de priorización para la asignación de cita post-alta.
- Intercomunicación con APIs: se utiliza la intercomunicación entre Python, Google Sheets, Google Drive y AppSheet de Google Cloud para almacenamiento de datos y manipulación de estos, además, de la visualización de resultados.

## 5.2.2 Diseño del Sistema

El diseño explica cómo se implementó la arquitectura descrita en la sección anterior incluyendo aspectos técnicos y funcionales, como la elección de tecnología, interacción entre los módulos y los algoritmos utilizados en el procesamiento de la información, entre otros.

Partiendo por la implementación de la arquitectura, se utilizó Python como lenguaje de programación principal para el desarrollo de los módulos del sistema. Para la extracción de información desde las epicrisis se emplearon bibliotecas como PyPDF2, unicodedata, NLTK, re y os. Estas bibliotecas permitieron trabajar con los archivos PDF de la base de datos del Servicio de Medicina Interna del Hospital de Talcahuano, manejar caracteres Unicode, realizar análisis de texto mediante NLP, implementar uso de expresiones regulares y operaciones con el sistema de archivos del sistema operativo, respectivamente.

Además, se utilizaron bibliotecas adicionales como textdistance, spacy e intertools para tareas más avanzadas en el procesamiento y análisis de texto, Pandas para manipulación de datos tabulares, openpyxl para trabajar con archivos Excel, Google OAuth 2.0 y biblioteca cliente para las API basadas en descubrimiento de Google para la interacción con servicios de Google como Sheets y Drive.

Las conexiones con Google Sheets, Google Drive y AppSheet de Google Cloud se establecieron utilizando APIs y keys proporcionadas por estas plataformas con el fin de sincronización y manipulación de datos entre el sistema y estas plataformas utilizando el protocolo OAuth 2.0 y las keys de API para autenticación y autorización de acceso a los datos con la finalidad de realizar operaciones

de lectura y escritura de datos. Esto permitió importar los episodios desde fuentes de datos externa y exportar los episodios priorizados y clasificados para poder visualizarlos de mejor manera en un entorno externo a Python, AppSheet.



### 5.2.3 Esquema del Sistema y componentes

En la Figura 5.1 se muestra un esquema que ofrece una visión del funcionamiento del sistema, incluyendo sus principales componentes y operaciones.

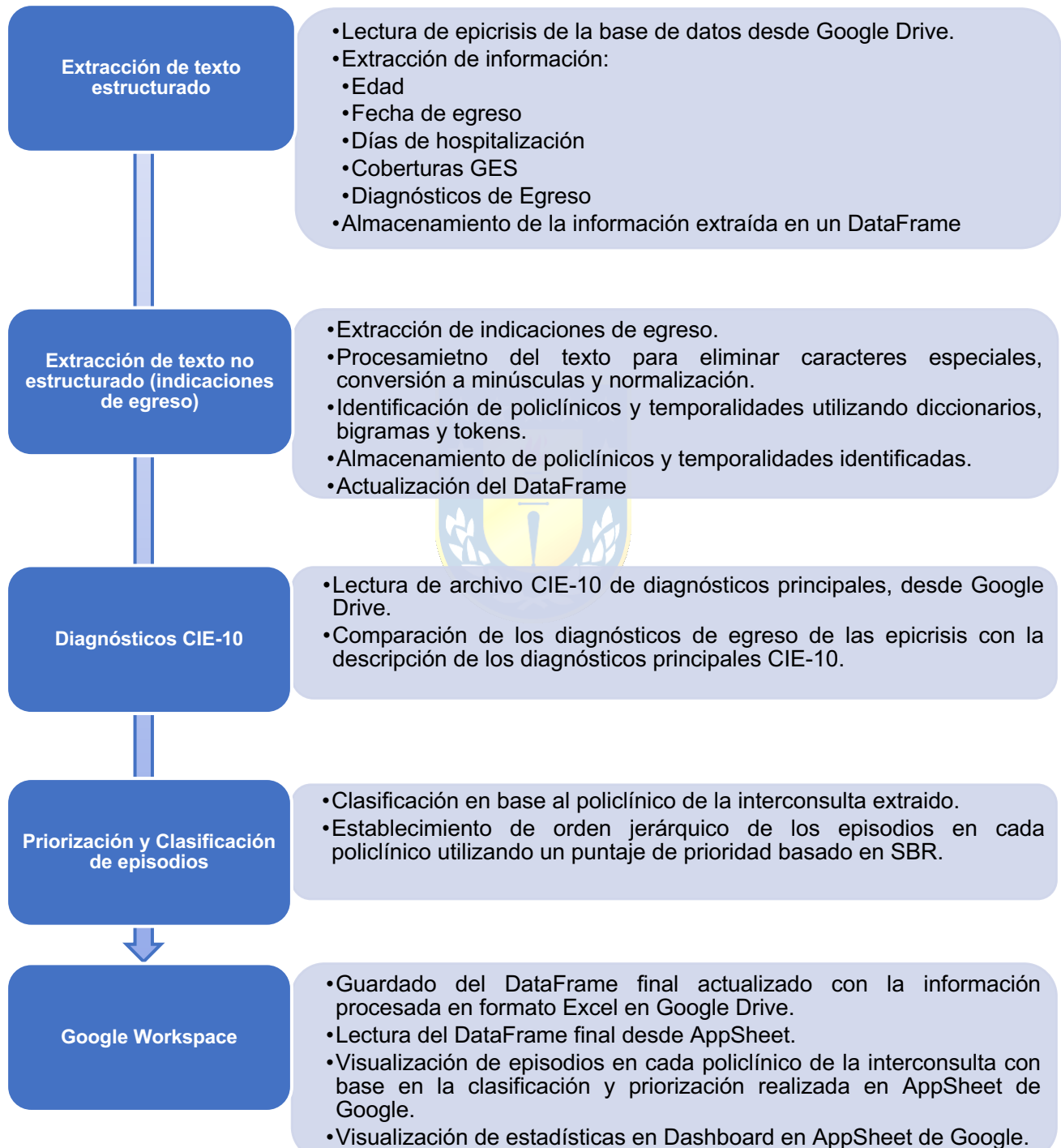


Figura 5.1. Esquema del sistema y componentes

## Capítulo 6. Resultados

### 6.1. Introducción

En este capítulo se presentan los resultados obtenidos en el proyecto de propuesta de automatización para proceso de agendamiento de citas médicas post-alta a partir de información extraída de epicrisis.

### 6.2. Extracción de Información de Epicrisis

Se procesaron 266 epicrisis de la base de datos del Servicio de Medicina Interna del Hospital Las Higueras de Talcahuano, emitidas entre 2020 y 2022. Estas se almacenaron en Google Drive, desde donde se realizaron sus lecturas y procesamiento con código en Python, permitiendo acceder a ellas de manera segura.

A partir de estas epicrisis, se extrajeron las variables de formato de texto estructurado, como se ejemplifica en la Figura 6.1 que muestra parte del DataFrame. Estas variables incluyeron 'Edad', 'Fecha de Egreso', 'Días de Hospitalización', 'Diagnóstico de Egreso N' y 'Etapa GES N', donde N representa el número del diagnóstico y la etapa GES correspondiente, ya que existieron más de uno de estos.

ID	Edad	Fecha de Egreso	Días de Hospitalización	Diagnóstico de Egreso 1	Etapa GES 1	Diagnóstico de Egreso 2	Etapa GES 2
9400636.pdf	79	28/09/2022	5	EDEMA PULMONAR	No Aplica	Ninguno	Ninguno
94003227.pdf	82	31/08/2020	11	EDEMA PULMONAR	No Aplica	Ninguno	Ninguno
94005123.pdf	71	27/03/2020	1	CEFALEA	No Aplica	Ninguno	Ninguno
94005668.pdf	92	17/01/2020	21	INSUF. RESPIRATORIA AGUDA	No Aplica	Ninguno	Ninguno
94006025.pdf	88	23/03/2022	11	COVID-19,	No Aplica	Ninguno	Ninguno

Figura 6.1. Extracción de variables estructuradas

Se logró extraer la sección de 'Indicaciones de Egreso' de las epicrisis, como se muestra en la Figura 6.2. Esta sección contiene las indicaciones generales para el post-alta que se encontraban en

Key	Type	Size	Value
9400636.pdf	str	603	reposo semisentada regimen diabetico hiposodico amlodipino 10 mg dia v ...
94003227.pdf	str	1103	1 2ministerio de salud fecha impresion 29 04 2024 servicio salud talca ...
94005123.pdf	str	592	reposo relativo regimen diabetico papilla fraccionada aspirina 100 mg ...
94005668.pdf	str	607	1 reposo relativo asistido 2 regimen papilla en 90 grado liquidos con ...
94006025.pdf	str	1106	regimen hiposodico reposo relativo losartan 25mg x dia vo 1 2ministeri ...

Figura 6.2. Extracción sección 'Indicaciones de Egreso'

un campo narrativo en formato de texto no estructurado, donde se encuentra el policlínico del control post-alta y el tiempo en que se debe asistir a este.

Para realizar la extracción de policlínicos, el mejor rendimiento para la distancia de Levenshtein fue con un umbral de 0.5. Con las temporalidades extraídas, se añadieron y calcularon las siguientes variables al código: 'Días para agendar' (temporalidad indicada en días), 'Fecha Indicada' (fecha calculada sumando la temporalidad a la fecha de egreso), 'Días para Fecha Indicada' (días desde el presente hasta la fecha indicada para el control) y también, 'Días desde el Egreso' (días desde el egreso hasta el presente), las que se pueden visualizar en el DataFrame actualizado de la Figura 6.3.

ID	Edad	Fecha de Egreso	Días de Hospitalización	Indicación Control post-alta	Policlínico	Temporalidad
94000636.pdf	79	28/09/2022	5	control policlinico nefrologia ...	Nefrología	segun disponibilidad
94003227.pdf	82	31/08/2020	11	control poli ucra 3 dias con ex...	Cardiología	segun disponibilidad
94005123.pdf	71	27/03/2020	1	reposo relativo regimen diabeti...	Ninguno	Ninguno
94005668.pdf	92	17/01/2020	21	1 reposo relativo asistido 2 re...	Ninguno	Ninguno
94006025.pdf	88	23/03/2022	11	control con nuevo perfil fierro...	Hematología	segun disponibilidad

ID	Días para agendar	Fecha Indicada	Días desde el Egreso	Días para Fecha Indicada
94000636.pdf	14	12/10/2022	646	-632
94003227.pdf	14	14/09/2020	1404	-1390
94005123.pdf	0	Ninguno	1561	0
94005668.pdf	0	Ninguno	1631	0
94006025.pdf	14	06/04/2022	835	-821

Figura 6.3. Extracción de policlínicos, temporalidades y adición de variables

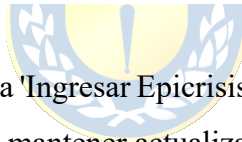
### 6.3. Clasificación y Priorización de Pacientes

Al realizar el cálculo del 'Puntaje Prioridad' mediante el SBR, se obtuvo una priorización efectiva de los pacientes según las variables definidas por las reglas del sistema. Esta metodología facilitó la asignación de puntajes precisos para los casos que requerían control post-alta, asegurando prioridad en la programación de citas. Para los pacientes sin necesidad de agendamiento, se les asignó un puntaje de 0, específicamente cuando la variable "Policlínico" era "Ninguno". Cada episodio se clasificó según el policlínico correspondiente y se ordenó dentro del mismo según el puntaje de prioridad calculado. Además, se añadió la columna 'Estado' para indicar si el paciente está 'Pendiente' de agendamiento o 'Agendado'. Estos resultados se ingresaron como base de datos en AppSheet para su visualización ordenada.

La Figura 6.4 muestra la interfaz de AppSheet, específicamente la visualización 'Información General'. En esta sección, aparece toda la información extraída de las epicrisis utilizando el código. Se adicionaron símbolos y colores para resaltar visualmente las variables 'Puntaje Prioridad', 'Días para Fecha Indicada' y 'Estado' para el agendamiento.

Información General											
	Puntaje Prioridad...	Fecha Indicada del Control...	Días para Fecha Indicada...	ID	Edad...	Policlínico	Fecha del Alta...	Días de Hospitalización...	Días desde el Egreso...	Diagnóstico de Egreso 1	Etapa GES 1
● PENDIENTE	196										
📅	472	29/01/2020	-1.619	2007150003.pdf	55	Endocrinología	15/01/2020	7	1633	Hipoglucemia, no especificada	No Aplica
📅	469	15/02/2020	-1.602	94024576.pdf	83	Hematología	01/02/2020	3	1616	Anemia por deficiencia de hi...	No Aplica
📅	469	15/02/2020	-1.602	94024576.pdf	83	Nefrología	01/02/2020	3	1616	Anemia por deficiencia de hi...	No Aplica
📅	467	03/02/2020	-1.614	2019286118.pdf	20	Broncopulmonar	20/01/2020	13	1628	Otras tirotoxicosis	No Aplica
📅	446	23/04/2020	-1.534	2017274676.pdf	33	Nefrología	09/04/2020	1	1548	Otros tipos de hipertensión s...	No Aplica
📅	444	07/05/2020	-1.520	2012207348.pdf	68	Medicina Interna	27/04/2020	5	1530	Acidosis	No Aplica
📅	442	11/05/2020	-1.516	2008163500.pdf	70	Medicina Interna	27/04/2020	2	1530	Bronquitis aguda, no especifi...	No Aplica
📅	441	16/05/2020	-1.511	99142799.pdf	79	Broncopulmonar	15/05/2020	11	1512	Tumor de comportamiento i...	No Aplica
📅	441	29/05/2020	-1.498	2018277827.pdf	84	Cardiología	15/05/2020	1	1512	Angina inestable	No Aplica
📅	441	29/05/2020	-1.498	2018277827.pdf	84	Diabetes	15/05/2020	1	1512	Angina inestable	No Aplica
📅	437	12/06/2020	-1.484	2018280262.pdf	70	Cardiología	29/05/2020	8	1498	Infarto transmural agudo deL...	No Aplica
📅	437	18/05/2020	-1.509	96069492.pdf	32	Medicina Interna	04/05/2020	2	1523	Migraña, no especificada	No Aplica
📅	436	17/06/2020	-1.479	2008155832.pdf	61	Infectología	03/06/2020	3	1493	COVID - 19	No Aplica
📅	436	04/06/2020	-1.492	2020297357.pdf	56	Nefrología	21/05/2020	8	1506	Enfermedad renal crónica, e...	No Aplica
📅	435	11/06/2020	-1.485	2020297357.pdf	56	Reumatología	21/05/2020	8	1506	Enfermedad renal crónica, e...	No Aplica

Figura 6.4. Sección 'Información General' AppSheet



Se añadió una sección denominada 'Ingresar Epicrisis', que permite la carga de nuevos archivos .pdf con episodios adicionales, pudiendo mantener actualiza la información y permitiendo su visualización posteriormente en AppSheet.

La sección 'Indicadores', posterior a la de ingresos de nuevas epicrisis, incorporó un Dashboard, ver Figura 6.5. En este se presentan los episodios agendados y pendientes, acompañados de información adicional, que incluye la demanda en los policlínicos del Servicio de Medicina Interna, un gráfico circular que muestra el porcentaje de los estados 'Pendientes' o 'Agendados' de agendamiento, un histograma de las edades y los 10 diagnósticos más frecuentes entre los episodios.

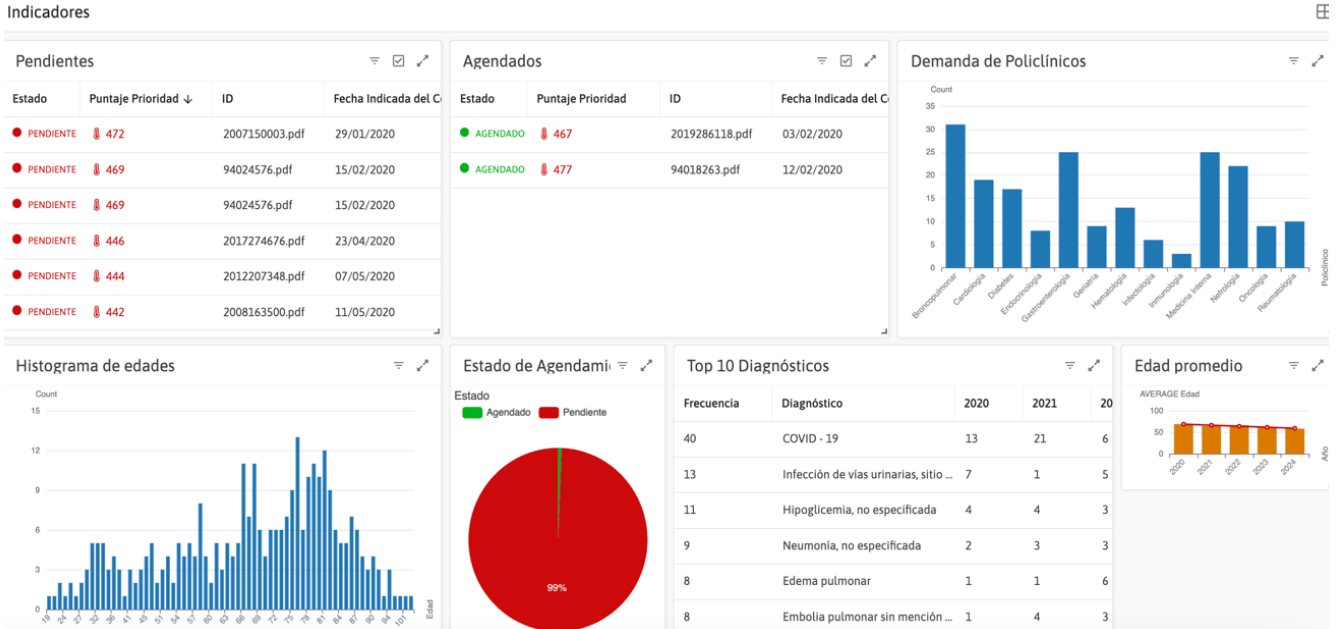


Figura 6.5. Sección 'Indicadores' AppSheet.

Cada policlínico tiene una sección dedicada específicamente en AppSheet, un ejemplo de esto se muestra en la Figura 6.6.



Policlínico Broncopulmonar

Estado	Puntaje Prioridad	ID	Fecha del Alta	Días desde el Egre...	Fecha Indicada del Contr...	Días para agenda...	Indicación Control post-alta
AGENDADO	467	2019286118.pdf	20/01/2020	1628	03/02/2020	14	control policlinico broncopulmonar moreno para d
PENDIENTE	441	99142799.pdf	15/05/2020	1512	16/05/2020	1	control poli broncopulmonar los primeros dias juni
PENDIENTE	431	2000010305.pdf	16/06/2020	1480	23/06/2020	7	control cesfam sala era para evaluar historia asma
PENDIENTE	425	2003090312.pdf	30/06/2020	1466	10/07/2020	10	control policlinico broncopulmonar tapia 10 dias
PENDIENTE	420	2020299012.pdf	10/07/2020	1456	09/08/2020	30	control 14 dias con broncopulmonar control 1 mes
PENDIENTE	418	95046203.pdf	14/08/2020	1421	28/08/2020	14	control telefonico con broncopulmonar segun prot
PENDIENTE	409	95052117.pdf	25/08/2020	1410	08/09/2020	14	control 14 dias con broncopulmonar urgencias sos
PENDIENTE	401	2020300059.pdf	17/09/2020	1387	01/10/2020	14	control fonoconsulta con broncopulmonar rivas au
PENDIENTE	399	98123777.pdf	29/09/2020	1375	06/10/2020	7	control broncopulmonar covid 7 dias mas cuarente

Figura 6.6. Sección 'Policlínico Broncopulmonar' AppSheet.

Se añadió una forma para acceder a más detalles del episodio, que permite ver los datos de 'Información General' al hacer clic sobre el episodio, facilitando así el acceso a toda la información

extraída, además de poder cambiar el estado del episodio a 'pendiente' o 'agendado' como se muestra en Figura 6.7.

Policlínico Broncopulmonar							2019286118.pdf
Estado	Puntaje Prioridad	ID	Fecha del Alta	Días desde el Egre...	Fecha Indicada del Contr...	Días para ag...	
● AGENDADO	467	2019286118.pdf	20/01/2020	1628	03/02/2020	14	● AGENDADO
● PENDIENTE	441	99142799.pdf	15/05/2020	1512	16/05/2020	1	● PENDIENTE
● PENDIENTE	431	2000010305.pdf	16/06/2020	1480	23/06/2020	7	● PENDIENTE
● PENDIENTE	425	2003090312.pdf	30/06/2020	1466	10/07/2020	10	● PENDIENTE
● PENDIENTE	420	2020299012.pdf	10/07/2020	1456	09/08/2020	30	● PENDIENTE
● PENDIENTE	418	95046203.pdf	14/08/2020	1421	28/08/2020	14	● PENDIENTE
● PENDIENTE	409	95052117.pdf	25/08/2020	1410	08/09/2020	14	● PENDIENTE
● PENDIENTE	401	2020300059.pdf	17/09/2020	1387	01/10/2020	14	● PENDIENTE
● PENDIENTE	399	98123777.pdf	29/09/2020	1375	06/10/2020	7	● PENDIENTE
● PENDIENTE	399	2000001111.pdf	15/09/2020	1389	15/10/2020	30	● PENDIENTE
● PENDIENTE	384	94019443.pdf	07/12/2020	1306	21/12/2020	14	● PENDIENTE
● PENDIENTE	379	2001047328.pdf	10/12/2020	1303	24/12/2020	14	● PENDIENTE
● PENDIENTE	381	99150329.pdf	18/12/2020	1295	01/01/2021	14	● PENDIENTE
● PENDIENTE	372	2001038041.pdf	28/12/2020	1285	11/01/2021	14	● PENDIENTE
● PENDIENTE	370	2021302003.pdf	08/01/2021	1274	22/01/2021	14	● PENDIENTE
● PENDIENTE	373	99138621.pdf	21/01/2021	1261	04/02/2021	14	● PENDIENTE

ID	2019286118.pdf
Policlínico	Broncopulmonar
Edad	20
Fecha del Alta	20/01/2020
Días de Hospitalización	13
Días desde el Egreso	1628
Indicación Control post-alta	control policlinico broncopulmonar moreno para derivacion policlinico tep cronico hospital del torax 1 semana pendiente dlco solicitada como compra servicios 15 01 20 control poli taco brevedad control sos urgencias
Puntaje Prioridad	467
Fecha Indicada del Control	03/02/2020
Días para agendar	14
Días para Fecha Indicada	-1.614

**Figura 6.7. Sección 'Policlínico Broncopulmonar' y datos extra del episodio en AppSheet.**

Debido a restricciones del protocolo presentado al Comité Ético Científico, la aplicación no se pudo someter a una evaluación de funcionalidad por parte del personal del Servicio de Medicina Interna. Sin embargo, el jefe de Medicina Interna expresó su conformidad con el desarrollo del sistema, indicando que cumple con lo solicitado. Según sus observaciones, los dashboards de la herramienta proporcionan funciones que se ajustan a las necesidades del proceso, permitiendo una visualización clara del comportamiento general de las interconsultas solicitadas por el Servicio de Medicina Interna.

## Capítulo 7. Conclusiones

---

### 7.1. Introducción

Se alcanzó el objetivo de proponer un sistema de clasificación y priorización de episodios para agendar controles médicos luego del alta, considerando variables demográficas, diagnósticas y tiempos de espera, a partir de la información extraída de la base de datos del Servicio de Medicina Interna del Hospital Las Higueras. Esto integrando soluciones tecnológicas y estrategias de automatización, incluyendo herramientas de la IT y heurística, para ayudar a reducir cargas administrativas, minimizar el tiempo en el proceso de agendamiento y evitar desplazamientos innecesarios del personal, agilizando el seguimiento de los pacientes después del alta médica.

### 7.2. Conclusiones

Los resultados demuestran que el sistema puede procesar un volumen razonable de datos con eficiencia y coherencia en el entorno hospitalario. Además, se cumplieron los objetivos específicos de extracción de información relevante desde las epicrisis, clasificación y priorización de los episodios de pacientes en cada policlínico de la interconsulta del Servicio de Medicina Interna, y provisión de información sobre el flujo de pacientes agendados en los distintos policlínicos mediante el sistema implementado.

Este sistema extrajo datos clínicos y demográficos de las epicrisis con precisión. La implementación técnica avanzada de procesamiento de texto y datos, utilizando bibliotecas y herramientas de Python, fueron cruciales para transformar la información extraída a un formato estructurado y utilizable para análisis y toma de decisiones. El uso de NLP y otros algoritmos de manejo de datos garantizaron la calidad y exactitud de esta información procesada. Adicionando la aplicación de criterios clínicos, demográficos e históricos, el sistema logró priorizar y clasificar los episodios de manera efectiva. Esto permitió identificar y asignar niveles de prioridad adecuados, asegurando que los casos más críticos y urgentes reciban la atención correspondiente. Además, la integración del sistema con plataformas intercomunicadas de Google mediante APIs permitió una visualización clara y accesible de los resultados, facilitando un seguimiento activo de estos.

La implementación de esta propuesta genera un impacto positivo en varias áreas. En primer lugar, puede reducir los tiempos desde la emisión de la epicrisis hasta el agendamiento del paciente,

disminuyendo desplazamientos innecesarios del personal médico hacia los mesones de los policlínicos de interconsulta. La automatización permite una transmisión rápida y precisa de la información, eliminando posibles errores de interpretación o ingreso de datos. Además, al ser tareas repetitivas donde se procesan los datos de manera eficiente, los profesionales médicos pueden dedicar más tiempo a la atención directa de los pacientes y utilizar mejor su tiempo en otras tareas relevantes. Esto no solo mejora el uso del tiempo del personal sanitario, sino que también lo hace para el personal administrativo al proporcionarle información completa, sin errores y en un menor tiempo.

El jefe del servicio al revisar su aplicación expresa que la aplicación desarrollada cumple de manera efectiva con la necesidad identificada en el problema planteado. Según sus observaciones, los distintos dashboards integrados en la herramienta proporcionan funciones detalladas y específicas que se ajustan perfectamente a las necesidades de cada etapa del proceso, permitiendo una visualización clara y comprensiva de los datos, lo que facilita la apreciación del comportamiento general y la identificación de patrones. Como sugerencias a futuro, se propone trabajar en la optimización del diseño de la aplicación para incorporar aún más estética, además seguir afinando los KPI o Indicadores Clave de Desempeño para asegurar que reflejen de manera precisa las métricas más relevantes. A pesar de estos aspectos a mejorar, destaca el potencial prometedor de la herramienta, considerando que tiene la capacidad de abordar de manera efectiva uno de los principales desafíos del proceso de citas posteriores a la hospitalización, con la posibilidad de escalar la solución a un nivel mucho mayor y ofrecer un impacto significativo en la eficiencia y efectividad del proceso hospitalario en general.

### **7.3. Avances Futuros**

Aunque el proyecto enfrenta limitaciones, como la dependencia de la calidad y la presentación de los datos, así como la interoperabilidad con el sistema hospitalario, destaca el potencial de las tecnologías de la información y los algoritmos de inteligencia artificial en la gestión médica actual. Los resultados obtenidos pueden servir como modelo para innovar y mejorar otros entornos dinámicos, especialmente en reingeniería de procesos, al beneficiar la comunicación y colaboración entre operadores y gestores de servicios. Además, la plataforma tecnológica y los recursos de datos contribuyen significativamente a una toma de decisiones más informada y a una asignación más eficiente de los recursos disponibles, brindando un servicio de calidad continuo y oportuno.

En el futuro, se espera que, junto a estas u otras tecnologías, se amplíe el alcance del sistema, se mejore la interfaz de usuario y se profundice en el análisis de la información recopilada. En particular, desarrollar una clasificación generalizada de la gravedad de los diagnósticos en colaboración con un equipo especializado, lo que no existe y sería de ayuda para priorización de casos. Además, se espera que escale el uso del sistema a otras áreas para mejorar el rendimiento de los sistemas de gestión en salud.

Con estas mejoras, se anticipa que los sistemas de agendamiento y gestión de citas médicas ganarán robustez y perfección a futuro.



## Glosario

---

AI	: Inteligencia Artificial (Artificial Intelligence)
BERT	: Representaciones de Codificador Bidireccional de Transformadores (Bidirectional Encoder Representations from Transformers).
CatBoost	: Algoritmo de aprendizaje automático basado en árboles de decisión.
CIE	: Clasificación Internacional de Enfermedades.
DM	: Minería de Datos (Data Mining).
GES	: Régimen General de Garantías Explícitas en Salud.
GPTs	: Modelos Preentrenados Generativos (Generative Pretrained Transformers).
IS	: Sistemas de Información (Information System)
IT	: Tecnología de la Información (Information Technology).
KPI	: Indicadores Clave de Desempeño (Key Performance Indicators).
MAS	: Modelo de Aprendizaje Supervisado.
MDP	: Proceso de Decisión de Markov o Markoviano (Markov Decision Process).
ML	: Aprendizaje Automático (Machine Learning).
NLP	: Procesamiento del Lenguaje Natural (Natural Language Processing).
NER	: Named Entity Recognition, en español, Reconocimiento de Entidades Nombradas.
OMS	: Organización Mundial de la Salud
PLMs	: Modelos de Lenguaje Pre-entrenados (Pre-trained Language Models).
PoS	: Etiquetado gramatical (Part-of-Speech Tagging).
Python	: Lenguaje de programación Python.
REGEX	: Expresiones Regulares.
REDEx:	: Regular Expression Discovery Extractor.
SBR	: Sistema Basado en Reglas. (RBML).
SCA	: Síndrome Coronario Agudo.
SpaCy	: Biblioteca de procesamiento del lenguaje natural para Python.

TENS : Técnico de nivel superior en Enfermería.  
TM : Minería de Textos (Text Mining).  
VA : Administración de Veteranos (Veterans Affairs).



## Referencias

---

- [1] J. Godoy Olave y J. Barraza Mesquida, “La ficha clínica mirada desde la legislación chilena actual”, *Acta Bioeth.*, vol. 24, núm. 2, pp. 181–188, 2018.
- [2] G. M. Fuenzalida Del Valle, “Agendamiento dinámico de citas médicas a través de un proceso de decisión markoviano”, *Magíster en Ciencias de la Ingeniería*, 2016. doi: 10.7764/tesisUC/ING/16841.
- [3] D. Sarkar, “Chapter1: Text Analytics ,” in *Text Analytics with Python A Practitioner’s Guide to Natural Language Processing*, 2nd ed, Berkeley, CA: Apress, 2019, pp. 66–67.
- [4] Alliance for Health Policy and Systems Research - WHO. Systems thinking for health systems strengthening [En línea] Don de Savigny and Taghreed Adam, editor. 2009. 112 p. Disponible en: [http://apps.who.int/iris/bitstream/handle/10665/44204/9789241563895\\_eng.pdf;jsessionid=C6B4FD203D8738F5BD82A5C10DA0541A?sequence=1](http://apps.who.int/iris/bitstream/handle/10665/44204/9789241563895_eng.pdf;jsessionid=C6B4FD203D8738F5BD82A5C10DA0541A?sequence=1). [Consultado: 30-nov-2023].
- [5] K. Topuz, T. L. Urban, R. A. Russell, and M. B. Yildirim, “Decision support system for appointment scheduling and overbooking under patient no-show behavior,” *Ann. Oper. Res.*, 2024.
- [6] S. A. Alowais *et al.*, “Revolutionizing healthcare: the role of artificial intelligence in clinical practice,” *BMC Med. Educ.*, vol. 23, no. 1, 2023.
- [7] P. Van der Stuyft and P. De Vos, “La relación entre los niveles de atención constituye un determinante clave de la salud,” *Rev Cub Sal Públ*, vol. 34, no. 4, pp. 0–0, 2008.
- [8] “Salud Pública y las etapas de la atención,” *Postgrados UNAB*, 29-Dec-2022. [En línea]. Disponible en: <https://www.postgradounab.cl/noticias/salud-publica-y-las-etapas-de-la-atencion/>. [Consultado: 01-Apr-2024].
- [9] J. Vignolo, M. Vacarezza, C. Álvarez, and A. Sosa, “Niveles de atención, de prevención y atención primaria de la salud,” *Arch Med Int*, vol. 33, no. 1, pp. 7–11, 2011.
- [10] “¿Cuáles son los Derechos de las personas en su atención de salud? - ¿Cuáles son los Derechos de las personas en su atención de salud?,” *Preguntas Frecuentes*. Superintendencia de Salud, Gobierno de Chile. [En línea]. Disponible en: <http://www.supersalud.gob.cl/consultas/667/w3-article-7960.html>. [Consultado: 01-Apr-2024].

- [11] H. C. Bestagno y C. C. Rodriguez, ago-2023.
- [12] Edwin Nieto, “Tecnologías de la información”, Blog - Bootcamps, 05-abr-2023. [En línea]. Disponible en: <https://bootcampvirtual.udd.cl/blog/tecnologias-de-la-informacion/>. [Consultado: 21-sep-2023].
- [13] “Estrategia mundial sobre salud digital 2020–2025”, Who.int, 18 de agosto de 2021. [En línea]. Disponible en: <https://www.who.int/es/publications/i/item/9789240020924>. [Consultado: 06-oct-2023].
- [14] M. de Salud, “Estrategia Nacional de Salud para los Objetivos Sanitarios al 2030”, Minsal.cl, feb-2022. [En línea]. Disponible en: <https://www.minsal.cl/wp-content/uploads/2022/03/Estrategia-Nacional-de-Salud-2022-MINSAL-V8.pdf>. [Consultado: 06-oct-2023].
- [15] S. de Salud, “Glosario de términos de uso frecuente”, Gob.cl. [En línea]. Disponible en: [https://www.supersalud.gob.cl/664/articles-18119\\_recurso\\_1.pdf](https://www.supersalud.gob.cl/664/articles-18119_recurso_1.pdf). [Consultado: 30-nov-2023].
- [16] Servicio De Salud Talcahuano – Hospital Las Higueras, Subdirección Médica De Gestión Clínica, “Cartera De Servicios Hospital Las Higueras - Talcahuano”, 2019.
- [17] TENS Hospital Las Higueras, 01-nov-2023.
- [18] “Clasificación Estadística Internacional de Enfermedades y Problemas Relacionados con la Salud,” Minsal.cl. [En línea]. Disponible en: <https://repositoriodeis.minsal.cl/ContenidoSitioWeb2020/uploads/2018/03/CIE10-2015-Vol-2.pdf>. [Consultado: 14-Apr-2024].
- [19] “International Classification of Diseases (ICD),” Who.int. [En línea]. Disponible en: <https://www.who.int/es/standards/classifications/classification-of-diseases>. [Consultado: 13-Apr-2024].
- [20] O'Brien, J. A., & Marakas, G. M. *Management Information Systems*. 10th ed. McGraw-Hill: 2011.
- [21] Ibm.com. What is Machine Learning? [En línea]. Disponible en: <https://www.ibm.com/topics/machine-learning>. [Consultado: 06-oct-2023].
- [22] Kelleher, D. J. and Tierney, B. *Ciencia de datos* (MIT Press / Conocimientos Esenciales) (Spanish Edition). Ediciones UC: 2021.

- [23] Elsevier. Inteligencia Artificial en entornos sanitarios. Tipos de algoritmos de ‘machine learning’ [En línea]. Disponible en: <https://www.elsevier.com/es-es/connect/inteligencia-artificial-en-entornos-sanitarios-tipos-de-algoritmos-de>. [Consultado: 06-oct-2023].
- [24] I. B. M. Data y AI Team. 10 everyday machine learning use cases. IBM Blog. [En línea]. Disponible en: [www.ibm.com/blog/10-everyday-machine-learning-use-cases](http://www.ibm.com/blog/10-everyday-machine-learning-use-cases). [Consultado: 06-oct-2023].
- [25] C. Aracena, F. Villena, F. Arias, y J. Dunstan, “Aplicaciones de aprendizaje automático en salud”, *Rev. médica Clín. Las Condes*, vol. 33, no. 6, pp. 568–575, 2022.
- [26] M. J. Taboada Iglesias and A. Gómez-Pérez, “Sistemas basados en reglas,” in *Inteligencia artificial: métodos, técnicas y aplicaciones*, J. T. Palma Méndez and R. Marín Morales, Eds. Madrid, España: Facultad de Informática (UPM), 2008, p. 1022.
- [27] J. L. Moya-Rodríguez, A. M. Becerra-Ferreiro, and C. A. Chagoyén-Méndez, “Utilización de Sistemas Basados en Reglas y en Casos para diseñar transmisiones por tornillo sinfín,” *Ingeniería Mecánica*, vol. 15, no. 1, pp. 01–09, 2012.
- [28] Hand, D. J., Mannila, H., & Smyth, P. *Principles of data mining*. MIT Press: 2001.
- [29] Ibm.com. ¿Qué es la minería de texto?. [En línea]. Disponible en: <https://www.ibm.com/es-es/topics/text-mining>. [Consultado: 06-oct-2023].
- [30] “Expresiones regulares - MATLAB & Simulink - MathWorks América Latina,” *Mathworks.com*. [En línea]. Disponible en: [https://la.mathworks.com/help/matlab/matlab\\_prog/regular-expressions.html](https://la.mathworks.com/help/matlab/matlab_prog/regular-expressions.html). [Consultado: 11-Apr-2024].
- [31] M. A. Murtaugh, B. S. Gibson, D. Redd, and Q. Zeng-Treitler, “Regular expression-based learning to extract bodyweight values from clinical notes,” *J. Biomed. Inform.*, vol. 54, pp. 186–190, 2015.
- [32] B. Berger, M. S. Waterman, and Y. W. Yu, “Levenshtein distance, sequence comparison and biological database search,” *IEEE Trans. Inf. Theory*, vol. 67, no. 6, pp. 3287–3294, 2021] [F. P. Miller, A. F. Vandome, and J. McBrewster, Eds., “Levenshtein Distance”. VDM Publishing, 2009. ISBN: 6130216904, 9786130216900.

- [33] Khanacademy.org. Usar heurísticas. [En línea]. Disponible en: <https://es.khanacademy.org/computing/ap-computer-science-principles/algorithms-101/solving-hard-problems/a/using-heuristics>. [Consultado: 06-oct-2023].
- [34] Z. A. Abdalkareem, A. Amir, M. A. Al-Betar, P. Ekhan, y A. I. Hammouri, “Healthcare scheduling in optimization context: a review”, *Health Technol.*, vol. 11, n° 3, pp. 445–469, may 2021.
- [35] Python.org. BeginnersGuide/Overview - Python Wiki. [En línea]. Disponible en: <https://wiki.python.org/moin/BeginnersGuide/Overview>. [Consultado: 06-oct-2023].
- [36] D. G. Es, “Probamos SpaCy: mucho más que una librería para crear proyectos reales de procesamiento del lenguaje natural”, *datos.gob.es*, 23-jun-2022. [En línea]. Disponible en: <https://datos.gob.es/es/blog/probamos-spacy-mucho-mas-que-una-libreria-para-crear-proyectos-reales-de-procesamiento-del>. [Consultado: 05-oct-2023].
- [37] A. S. Alberca, “La librería Pandas”, *Aprende con Alf*. [En línea]. Disponible en: <https://aprendeconalf.es/docencia/python/manual/pandas/>. [Consultado: 05-oct-2023]
- [38] Google, “Google Workspace Updates,” *Google Workspace Updates*. [En línea]. Disponible en: <http://workspaceupdates.googleblog.com/>. [Consultado: 04-Apr-2024].
- [39] G. Workspace, “Soluciones empresariales para colaborar y optimizar operaciones,” *Google Workspace*. [En línea]. Disponible en: <https://workspace.google.com/intl/es-419/business/>. [Consultado: 04-Apr-2024].
- [40] G. Workspace, *Google Workspace | Business Apps & Collaboration Tools*. 2016.
- [41] “Introducción: Integración de aplicaciones de terceros con Google Workspace,” *Google.com*. [En línea]. Disponible en: [https://support.google.com/a/answer/10010706?hl=es#apps\\_drive](https://support.google.com/a/answer/10010706?hl=es#apps_drive). [Consultado: 04-Apr-2024].
- [42] D. Greene, “Introducing Google Cloud,” *Google*, 29-Sep-2016. [En línea]. Disponible en: <https://blog.google/products/google-cloud/introducing-google-cloud/>. [Consultado: 07-Ago-2024].
- [43] “AppSheet,” *Google Cloud*. [En línea]. Disponible en: <https://cloud.google.com/ap-sheet?hl=es-419>. [Consultado: 07-Ago-2024].
- [44] J. A. Mayor Gallego, “Muestreo Estadístico, Estimadores de Horvitz-Thompson y Hajek. Estimador de Hansen-Hurwitz,” *Personal.us.es*. [En línea]. Disponible en: [https://personal.us.es/jmayor/ficheros/me\\_P1\\_03.pdf](https://personal.us.es/jmayor/ficheros/me_P1_03.pdf). [Consultado: 05-Apr-2024].

- [45] F. M. Guillermo Ayala, “Probabiliad Básica, UNIVERSITAT DE VALÈNCIA,” *Www.uv.es*. [En línea]. Disponible en: <https://www.uv.es/ayala/docencia/probabilidad/prob.pdf>. [Consultado: 05-Apr-2024].
- [46] J. M. Hilbe, *Negative Binomial Regression*. Cambridge University Press, 2011. [En línea]. Disponible en: [https://books.google.cl/books?id=0Q\\_ijxOEBjMC](https://books.google.cl/books?id=0Q_ijxOEBjMC). [Consultado: 05-oct-2023].
- [47] L. Sterckx et al., “Clinical information extraction for preterm birth risk prediction”, *J. Biomed. Inform.*, vol. 110, p. 103544, oct. 2020, doi: 10.1016/j.jbi.2020.103544.
- [48] W. Sun, Z. Cai, Y. Li, F. Liu, S. Fang, y G. Wang, “Data Processing and Text Mining Technologies on Electronic Medical Records: A Review”, *J. Healthc. Eng.*, vol. 2018, pp. 1–9, 2018, doi: 10.1155/2018/4302425.
- [49] N. T. J. Bailey, “A study of queues and appointment systems in hospital out-patient departments, with special reference to waiting-times”, *J. R. Stat. Soc.*, vol. 14, núm. 2, pp. 185–199, 1952.
- [50] D. Gupta y B. Denton, “Appointment scheduling in health care: Challenges and opportunities”, *IIE Transactions*, vol. 40, n° 9, pp. 800–819, jul. 2008, doi: 10.1080/07408170802165880.
- [51] A. Ala y F. Chen, “Appointment Scheduling Problem in Complexity Systems of the Healthcare Services: A Comprehensive Review”, *Journal of Healthcare Engineering*, vol. 2022, p. e5819813, mar. 2022, doi: 10.1155/2022/5819813.
- [52] P. Rodó, “Proceso de Poisson”, *Economipedia*. [En línea]. Disponible en: <https://economipedia.com/definiciones/proceso-de-poisson.html>. [Consultado: 04-oct-2023].
- [53] S. Srinivas y A. R. Ravindran, “Optimizing outpatient appointment system using machine learning algorithms and scheduling rules: A prescriptive analytics framework”, *Expert Systems with Applications*, vol. 102, pp. 245–261, jul. 2018, doi: 10.1016/j.eswa.2018.02.022.
- [54] M. A. Murtaugh, B. S. Gibson, D. Redd, y Q. Zeng-Treitler, “Regular expression-based learning to extract bodyweight values from clinical notes”, *Journal of Biomedical Informatics*, vol. 54, pp. 186–190, abr. 2015, doi: 10.1016/j.jbi.2015.02.009.
- [55] N. Yousefi, F. Hasankhani, y M. Kiani, “Appointment scheduling model in healthcare using clustering algorithms”, 2020.
- [56] L. Sterckx et al., “Clinical information extraction for preterm birth risk prediction”, *J. Biomed. Inform.*, vol. 110, p. 103544, oct. 2020, doi: 10.1016/j.jbi.2020.103544.

- [57] S. V. Kovalchuk, A. A. Funkner, O. G. Metsker, y A. N. Yakovlev, “Simulation of patient flow in multiple healthcare units using process and data mining techniques for model identification”, *J. Biomed. Inform.*, vol. 82, pp. 128–142, jun. 2018, doi: 10.1016/j.jbi.2018.05.004.
- [58] S. L. Sophie, S. S. Sathya, y C. Deepesh, “Analyzing the Performance of Information Extraction System for Annotation of Patient Discharge Summary”, en *2022 IEEE Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, dic. 2022, pp. 1–4. doi: 10.1109/IATMSI56455.2022.10119418.
- [59] Y. Cui, L. Han, y G. Nenadic, “MedTem2.0: Prompt-based Temporal Classification of Treatment Events from Discharge Summaries”, en *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop)*, V. Padmakumar, G. Vallejo, y Y. Fu, Eds., Toronto, Canada: Association for Computational Linguistics, jul. 2023, pp. 160–183. doi: 10.18653/v1/2023.acl-srw.27.
- [60] “Introducción: Integración de aplicaciones de terceros con Google Workspace,” *Google.com*. [En línea]. Disponible en: [https://support.google.com/a/answer/10010706?hl=es#apps\\_drive](https://support.google.com/a/answer/10010706?hl=es#apps_drive). [Consultado: 04-Apr-2024].
- [61] G. Workspace, “Soluciones de asistencia sanitaria e innovación en ciencias biológicas,” *Google Workspace*. [En línea]. Disponible en: <https://workspace.google.com/intl/es/industries/healthcare/>. [Consultado: 04-Apr-2024].
- [62] R. Guido, G. Ielpa, y D. Conforti, “Scheduling outpatient day service operations for rheumatology diseases,” *Flex. Serv. Manuf. J.*, vol. 32, no. 1, pp. 102–128, 2020.
- [63] D. Pérez Aguada, M. Milanés Luque, O. Mar Cornelio, and A. Orellana García, “Diseño de sistema basado en reglas para apoyar la toma de decisiones de la población en la obtención de los medicamentos,” *RCIM*, vol. 15, no. 1, 2023.
- [64] Z. A. Abdalkareem, A. Amir, M. A. Al-Betar, P. Ekhan, y A. I. Hammouri, “Healthcare scheduling in optimization context: a review”, *Health Technol. (Berl.)*, vol. 11, núm. 3, pp. 445–469, 2021].

## ANEXO A. APROBACIÓN COMITÉ ÉTICO CIENTÍFICO

A continuación, se presenta el documento de aprobación del Comité Ético Científico del Servicio de Salud Talcahuano para adquirir la información de las epicrisis desde la base de datos del Servicio de Medicina Interna del Hospital Las Higueras de Talcahuano.

### Anexo A.1. Aprobación de uso de base de datos del Servicio de Medicina Interna - Página 1.

  
 Ministerio de Salud  
 Servicio de Salud Talcahuano  
 Hospital Las Higueras  
 Unidad de Apoyo a la Investigación Biomédica  
 PSK/C.M./Ipl

ORD. N° 002366 04.04.24 /  
 ANT: Acta sesión ordinaria N°15/28.03.2024 CEC\_SST  
 MAT: AUTORIZA TRABAJO DE INVESTIGACIÓN

DE : DRA. PATRICIA SÁNCHEZ KRAUSE  
 DIRECTORA HOSPITAL LAS HIGUERAS

A : DRA. ROSA FIGUEROA  
 INVESTIGADORA RESPONSABLE

Estimada Investigadora:

Junto con saludar, es grato dirigirme a Ud., para informarle que habiendo completado los requisitos administrativos y la revisión ética del estudio, según Acta N° 15 de la sesión ordinaria del día 28.03.2024 del Comité Ético Científico del Servicio Salud Talcahuano, que aprueba el Trabajo de Investigación: "Propuesta de optimización del agendamiento de citas médicas post alta, a partir de una base de datos de epicrisis del Servicio de Medicina Interna del Hospital Las Higueras", esta Directora:

1. Autoriza el inicio de las actividades contempladas, en su Investigación, a partir de esta fecha.
2. Ratifica la autorización de uso de base de datos anonimizada de mi Ord. N°1962 del 14.03.2024
3. Finalmente, solicito a UD., mantener informada a esta Dirección del avance de la investigación que dirige, especialmente, en casos de eventos adversos o desviaciones de Protocolo. Además de hacer llegar a esta Dirección una copia del Informe Final de su trabajo.

Sin otro particular y deseándole éxito en la investigación que dirigirá, le saluda muy cordialmente,

  
 DRA. PATRICIA SÁNCHEZ KRAUSE  
 DIRECTORA  
 HOSPITAL LAS HIGUERAS

Distribución:
 

- La indicada
- Jefe Servicio Medicina Interna
- Comité Ético Científico del Servicio Salud Talcahuano
- Unidad de Apoyo a la Investigación
- Oficina de Partes

 DSFH 4022

1

 HOSPITAL LAS HIGUERAS TALCAHUANO ALTO HORNO 777 FONO 2687331

Anexo A.3. Aprobación de uso de base de datos del Servicio de Medicina Interna - Página 3.

Lista de Asistencia Reunión Extraordinaria CEC SST

Jueves 28 de marzo de 2024

Natalia Manríquez

José Quiroga

Olivia Bohamondes P.

Pedro Bravo E.

Marcela Cortés J.

Wendy Henkel P.

Nataly Cordero

Mariana Flores Mora



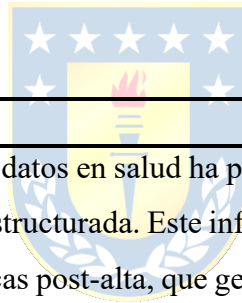
*[Handwritten signatures and initials corresponding to the names listed on the left]*

**UNIVERSIDAD DE CONCEPCION – FACULTAD DE INGENIERIA  
RESUMEN DE MEMORIA DE TITULO**

**Departamento** : Departamento de Ingeniería Eléctrica  
**Carrera** : Ingeniería Civil Biomédica  
**Nombre del memorista** : Francisca Belén Valdés Brito  
**Título de la memoria** : “PROPUESTA DE AUTOMATIZACIÓN PARA PROCESO DE AGENDAMIENTO DE CITAS MÉDICAS POST ALTA A PARTIR DE EPICRISIS EN EL SERVICIO DE MEDICINA INTERNA DEL HOSPITAL LAS HIGUERAS”

**Fecha de la presentación oral** : 12/08/2024

**Profesor(es) Guía** : Rosa Figueroa I. y Hernán Carrillo B.  
**Profesor(es) Revisor(es)** : Esteban Pino Q. y Jean Paul Navarrete C.  
**Concepto** :  
**Calificación** :



**Resumen**

El aumento de la producción de datos en salud ha provocado subutilización de información, una de las razones es su naturaleza no estructurada. Este informe aborda la problemática del proceso manual de agendamiento de citas médicas post-alta, que genera ineficiencias y carga administrativa solucionable mediante tecnologías y aprovechamiento de datos existentes. Por ello, se propuso un sistema automatizado basado en técnicas de Machine Learning, Data Mining y Procesamiento de Lenguaje Natural, utilizando Python y Google para visualización y gestión, permitiendo clasificar y priorizar episodios para su agendamiento, transformando la información extraída de epicrisis en formato estructurado y utilizable para análisis y toma de decisiones, considerando variables demográficas, diagnósticas y tiempos de espera. La finalidad fue reducir errores de lectura e ingreso de datos desde la emisión de la epicrisis hasta la programación del control, disminuyendo desplazamientos innecesarios del personal para la coordinación y facilitando una transmisión rápida y precisa de información clave. Los resultados demostraron que el sistema procesa datos de manera eficiente y coherente en un entorno hospitalario, proporcionando información relevante sobre episodios y su flujo para el agendamiento en diferentes policlínicos, con buena recepción y evaluación final, destacando el potencial de las tecnologías de la información en la gestión médica actual.