



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
ESCUELA DE POSTGRADO Y EDUCACIÓN CONTINUA
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

**APLICACIÓN DE TÉCNICAS DE CIENCIA DE DATOS PARA APOYAR LA
GESTIÓN DE INASISTENCIAS DE PACIENTES A CITAS MÉDICAS**

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIA DE DATOS

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

BENJAMÍN IGNACIO TEJEDA BARRERA

PROFESOR GUÍA:
SEBASTIÁN RÍOS PÉREZ

MIEMBROS DE LA COMISIÓN:
ANGEL JIMÉNEZ MOLINA
JUAN NEME GAVIOLA
MANUEL GRAÑA ROMAY

SANTIAGO DE CHILE

2023

RESUMEN DE LA TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIA DE DATOS Y MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL
POR: BENJAMÍN IGNACIO TEJEDA BARRERA
FECHA: 2023
PROF. GUÍA: SEBASTIÁN RÍOS PÉREZ

APLICACIÓN DE TÉCNICAS DE CIENCIA DE DATOS PARA APOYAR LA GESTIÓN DE INASISTENCIAS DE PACIENTES A CITAS MÉDICAS

La inasistencia de pacientes a citas médicas ambulatorias es un problema transversal en los servicios de atención de salud, el cual genera consecuencias negativas tanto para los pacientes como para los centros médicos. Dado esto, las inasistencias son una problemática comúnmente abordada en el campo de la ingeniería, donde, en particular, resulta natural enfrentar este problema desde la óptica de la ciencia de datos, dado el amplio repertorio de técnicas pertenecientes a esta disciplina que permiten construir modelos y herramientas para predecir la ocurrencia de inasistencias.

De esta forma, el siguiente trabajo de tesis presenta el diseño e implementación de una herramienta que posibilita definir segmentos horarios en los cuales ofertar sobrecupos para citas médicas, en base a una predicción previa sobre la ocurrencia de inasistencias, con el propósito de otorgar un mecanismo que permite mitigar estas ausencias al reemplazarlas con la atención de sobrecupos, consiguiendo un desempeño superior al obtenido por el mecanismo utilizado previamente al desarrollo de este estudio.

La herramienta se desarrolla en base a un conjunto de datos de citas médicas de medicina general disponibilizado por una institución chilena prestadora de servicios de salud. Esta herramienta se diseña considerando un enfoque de modelamiento de clasificación binaria y también un modelamiento en base a un pronóstico de series de tiempo, robusteciendo así el trabajo al sopesar un enfoque alternativo a la clásica predicción binaria comúnmente utilizada.

En base a las herramientas desarrolladas se evalúan diferentes estrategias para definir la oferta de sobrecupos, donde la estrategia finalmente seleccionada cubre un 8% de los bloques horarios en los que se presentan inasistencias (Recall) y reemplaza correctamente una inasistencia en el 34% de los bloques horarios donde se atienden sobrecupos (Precision), mientras que la estrategia utilizada previamente por la institución obtiene una Precision de 34% y un Recall de 4%, incrementando así la cobertura en 4 p.p. Además, la estrategia seleccionada reemplaza un 23% de las inasistencias promedio por semana mediante citas de sobrecupo, mientras que la estrategia utilizada previamente reemplaza un 11%, aumentando así el reemplazo de inasistencias en 12 p.p. Por último, la estrategia seleccionada obtiene una diferencia promedio por semana entre una tasa de inasistencia base y una tasa de inasistencia atenuada de un 3.26 p.p., mientras que la estrategia anterior alcanza una diferencia de 1.94 p.p., aumentando así esta diferencia en 1.32 p.p.

Para mi familia, ustedes lo son todo para mí

Agradecimientos

En primer lugar quiero agradecer a mi mamá Pilar, mi papá Carlos, mi hermana Javiera, mis abuelos Carlos, Mirtha, Victor y Angelica, mi tío Javier quien me cuida desde el cielo, mis tíos Ximena, Gonzalo y Hugo, y a todos los demás miembros de mi familia. Todos ustedes han formado parte de este proceso de alguna forma u otra y para mi este logro es tan mío como suyo, jamás habría llegado hasta este punto sin todo su amor y apoyo incondicional. En particular mi inmensa gratitud a mi mamá y papá, sepan que no pude haber recibido una mejor crianza, es gracias a ustedes que me he forjado como un hombre capaz de afrontar el mundo y valerse por sí mismo, completar este proceso es solo una pequeña forma de compensar todo el amor que me han entregado durante todos estos años, los amo inmensamente.

Agradezco también a mis tíos Alejandro y Betina, quienes me abrieron las puertas de su hogar cuando dejaba mi casa con apenas 18 años y un gran dolor en el corazón. En su hogar encontré justo lo que necesitaba, un espacio lleno de cariño y apoyo que me permitió superar mi primer año de universidad.

Agradezco a mi profesor guía Sebastián, quien fue una parte fundamental de este trabajo, estando siempre disponible para ayudarme cuando lo necesite. Su guía fue clave durante todo este proceso, donde en un inicio estaba lleno de dudas, abrumado por la magnitud de lo que significa un trabajo de tesis y la dificultad de llevarlo a cabo de la manera más exitosa. Sus consejos e indicaciones fueron siempre una fuente de tranquilidad, claridad y motivación para continuar.

Por último, agradezco a los profesores Manuel, Ángel y Juan, por su disposición a evaluar este trabajo desde sus diferentes perspectivas y con su gran conocimiento. En particular, agradezco a Juan por darme la oportunidad de desarrollar este trabajo junto a su gran equipo, el cual me entregó retroalimentación constantemente y me permitió llevar a cabo una implementación del trabajo desarrollado.

Tabla de Contenido

1. Introducción	1
1.1. Motivación e hipótesis	1
1.2. Objetivos del trabajo	2
2. Marco teórico	4
3. Metodología y materiales	7
3.1. Metodología	7
3.2. Materiales	8
3.2.1. Sparse Group Lasso	8
3.2.2. Regresión Logística	9
3.2.3. Extreme Gradient Boosting	9
3.2.4. Procesamiento de variables	11
3.2.5. Balance de datos	12
3.2.6. SARIMA	12
3.2.7. ETS	14
3.2.8. RNN	15
3.2.9. Medidas de evaluación	17
4. Resultados	21
4.1. Comprensión del negocio	21
4.2. Comprensión de los datos	21
4.3. Análisis y selección de datos	23
4.3.1. Análisis del conjunto de datos sobre citas médicas	24
4.3.2. Análisis del conjunto de datos sobre cantidad de inasistencias	30
4.4. Modelamiento	32
4.4.1. Predicción de inasistencia por cita	33
4.4.2. Pronóstico de cantidad de inasistencias	36
4.5. Evaluación	42
4.5.1. Evaluación de modelos de clasificación	42
4.5.2. Evaluación de modelos de pronóstico	46
4.5.3. Evaluación en la definición de sobrecupos	51
4.6. Despliegue	59
5. Conclusiones	69
Bibliografía	73

Anexos	76
A. Resultados selección de atributos vía Sparse Group Lasso.	76
B. Importancia de atributos.	78

Índice de Tablas

1.1.	Presentación de objetivos.	3
3.1.	Opciones para tendencia y estacionalidad.	14
3.2.	Configuraciones de modelos ETS.	15
4.1.	Variables del primer conjunto de datos.	22
4.2.	Estructura del segundo conjunto de datos.	23
4.3.	Estadísticas descriptivas de las variables categóricas (1/2).	24
4.4.	Estadísticas descriptivas de las variables categóricas (2/2).	25
4.5.	Estadísticas descriptivas de las variables numéricas.	25
4.6.	Estadísticas descriptivas sobre cantidad de inasistencias por bloque horario. . .	32
4.7.	Cantidad de bloques en cada semana.	40
4.8.	ROC-AUC y Precision-Recall-AUC (P-R-AUC) obtenidos en predicción de inasistencias sobre pacientes nuevos.	43
4.9.	ROC-AUC y Precision-Recall-AUC (P-R-AUC) obtenidos en predicción de inasistencias sobre pacientes recurrentes.	44
4.10.	Precision, Recall, Accuracy y F1 obtenidos en predicción sobre pacientes nuevos, modelo Logit.	45
4.11.	Precision, Recall, Accuracy y F1 obtenidos en predicción sobre pacientes recurrentes, modelo XGBoost.	45
4.12.	MASE, MAE y RMSE por semana para los pronósticos obtenidos mediante modelo SARIMA.	47
4.13.	MASE, MAE y RMSE por semana para los pronósticos obtenidos mediante modelo ETS.	48
4.14.	MASE, MAE y RMSE por semana para los pronósticos obtenidos mediante modelo RNN.	49
4.15.	Resultados en la predicción de inasistencias en bloques horarios, en base a modelos de clasificación binaria (Logit + XGBoost).	54
4.16.	Resultados en la predicción de inasistencias en bloques horarios, en base a modelo de pronóstico (ETS) y repartición usando tasa de inasistencia de médicos.	54
4.17.	Resultados en la predicción de inasistencias en bloques horarios, en base a modelo de pronóstico (ETS) y repartición usando probabilidad de inasistencia de citas (Logit + XGBoost).	54
4.18.	Resultados en la predicción de inasistencias sobre bloques horarios, en base a modelo de clasificación binaria (Logit + XGBoost). Bloques completos con citas de 15 minutos.	55
4.19.	Resultados en la predicción de inasistencias sobre bloques horarios, en base a modelo de pronóstico (ETS) y repartición usando tasa de inasistencia médicos. Bloques completos con citas de 15 minutos.	56

4.20.	Resultados en la predicción de inasistencias sobre bloques horarios, en base a modelo de pronóstico (ETS) y repartición usando probabilidad de inasistencia de citas (Logit + XGBoost). Bloques completos con citas de 15 minutos. . . .	56
4.21.	Resultados en la predicción de inasistencias sobre bloques horarios, en base a modelo de clasificación binaria (Logit + XGBoost). Bloques completos con citas de 20 minutos.	56
4.22.	Resultados en la predicción de inasistencias sobre bloques horarios, en base a modelo de pronóstico (ETS) y repartición usando tasa de inasistencia médicos. Bloques completos con citas de 20 minutos.	56
4.23.	Resultados en la predicción de inasistencias sobre bloques horarios, en base a modelo de pronóstico (ETS) y repartición usando probabilidad de inasistencia de citas (Logit + XGBoost). Bloques completos con citas de 20 minutos. . . .	57
4.24.	Resultados sobre la atención de sobrecupos (Pre piloto).	63
4.25.	Resultados sobre la atención de sobrecupos (Fase 1).	63
4.26.	Resultados sobre la atención de sobrecupos (Fase 2).	63
4.27.	Resultados sobre la atención de sobrecupos (Fase 3).	63
4.28.	Resultados sobre la atención de sobrecupos (Piloto completo).	63
B.1.	Coefficientes estimados para cada variable por modelo Logit para predicción de inasistencias de pacientes nuevos.	78
B.2.	Promedio de coeficientes estimados para variables categóricas por modelo Logit para predicción de inasistencias de pacientes nuevos.	79
B.3.	Ganancia promedio de información por variable en base a modelo XGBoost para predicción de inasistencias de pacientes recurrentes.	80

Índice de Ilustraciones

3.1.	Fases CRISP-DM.	7
3.2.	Celda long short-term memory.	16
3.3.	Curva ROC.	19
4.1.	Comportamiento variable Estado a través de variables categóricas (1/3).	26
4.2.	Comportamiento variable Estado a través de variables categóricas (2/3).	27
4.3.	Comportamiento variable Estado a través de variables categóricas (3/3).	28
4.4.	Comportamiento variable de interés a través de variables numéricas (1/3).	28
4.5.	Comportamiento variable de interés a través de variables numéricas (2/3).	29
4.6.	Comportamiento variable de interés a través de variables numéricas (3/3).	30
4.7.	Cantidad de inasistencias por fechas y bloques horarios.	30
4.8.	Distribución cantidad de inasistencias en bloques horarios por semana.	31
4.9.	Variables disponibles para cada grupo de pacientes.	33
4.10.	División del conjunto de datos para modelos de clasificación.	35
4.11.	Arquitectura RNN utilizada.	37
4.12.	División del conjunto de datos, modelos de pronóstico.	38
4.13.	Estrategia para repartición de inasistencias pronosticadas por modelo ETS.	53
4.14.	Matrices de confusión para escenarios en bloques horarios completos con citas de 15 y 20 min.	58
4.15.	Cantidad total de citas regulares y citas de sobrecupo ofertadas por semana (barras superpuestas).	61
4.16.	Cantidad total de citas de sobrecupo ofertadas y atendidas por semana (barras superpuestas).	62
4.17.	Matrices de confusión sobre la distribución de escenarios producidos por cada estrategia implementada.	65
4.18.	Proporción entre la cantidad de sobrecupos atendidos y la cantidad de citas no asistidas por semana.	66
4.19.	Tasa de inasistencia base y atenuada por semana.	67
4.20.	Afectación Bruta y Afectación Relativa por semana.	68
A.1.	Coefficientes estimados por Sparse Group Lasso para cada variable en el conjunto de pacientes nuevos.	76
A.2.	Coefficientes estimados por Sparse Group Lasso para cada variable en el conjunto de pacientes recurrentes.	77

Capítulo 1

Introducción

1.1. Motivación e hipótesis

La inasistencia de pacientes a citas médicas ambulatorias es un problema transversal en los servicios de atención de salud, el cual genera diversas consecuencias negativas tanto para los pacientes como para los centros médicos, como la reducción de la oportunidad de que otros pacientes obtengan la atención de salud que necesitan [1], la reducción de la rentabilidad de la prestación y la certeza sobre la condición de salud del paciente [2] y la producción de una subutilización de recursos médicos [3]. Además, las inasistencias no solo tienen un impacto negativo en la administración de la agenda, si no que, también tienen un impacto negativo en los resultados financieros [4], donde, en general, en el ámbito de la prestación de servicios de salud el cobro por servicio se realiza solo cuando el paciente se presenta a la cita.

Conforme a lo anterior, el problema de las inasistencias produce consecuencias negativas sociales y económicas y, dado esto, es un problema comúnmente abordado en la investigación, en especial en el ámbito de la ciencia de datos, donde existen técnicas y herramientas que permiten construir modelos para predecir la ocurrencia de estas inasistencias.

Con esto, para una institución prestadora de servicios médicos, resulta relevante contar con una herramienta de gestión que la habilite a tomar decisiones que permitan abordar esta problemática mediante la generación de acciones de mitigación, para así mejorar la eficiencia operacional, los beneficios económicos de la institución y los beneficios sociales de los pacientes. En particular, tomar acciones para aminorar los efectos de las inasistencias puede incrementar la eficiencia clínica y utilización, reducir pérdidas, mejorar la satisfacción de los médicos y pacientes y obtener mejores resultados en los tratamientos de salud [5, 6].

Este trabajo considera dos alternativas para el estado final de una cita médica: la asistencia o inasistencia (por parte del paciente). No obstante, existen otros dos posibles estados: la anulación previa de la cita por parte del paciente y la anulación previa de la cita por parte del médico tratante o la institución misma (este caso se conoce como "bloqueo"). Estos dos últimos estados no son considerados, ya que, en general, la anulación previa por parte del paciente no representa un costo de oportunidad, debido a que la cita reservada es liberada, existiendo la posibilidad de que esta sea reagendada, a excepción de aquellos casos en los que la anulación se realiza muy cercanamente a la fecha y hora programada para la cita. Por otra parte, los bloqueos de citas médicas son bastante arbitrarios, lo que genera que las opciones

de mitigación sean reducidas ya que, en general, cuando se presentan bloqueos se cancelan segmentos completos de oferta donde la posibilidad de reagendar la cita en el mismo día es prácticamente nula. En esta línea, Verdugo [7] presenta un rediseño del proceso de bloqueo de citas médicas, con el fin de mejorar la tasa de reagendamiento de los pacientes. El marco metodológico de su estudio se basa en Lean Management y en Ingeniería de Negocios, presentando un rediseño que tiene como idea base pasar de una notificación de bloqueo a una notificación con solución oportuna, lo cual implica un cambio de estándares de gestión, pero también de paradigma, al entender que tras una cita bloqueada hay un paciente afectado. La autora diseña cambios en las actividades y responsabilidades de los actores involucrados en el proceso, e incluye mejoras en las tecnologías habilitantes, optimizando un proceso bajo la lógica de aumentar el reagendamiento por un mejor servicio de oportunidad al paciente afectado, y consiguiendo así transformar una gestión que tomaba en promedio seis días hacia una que toma solo un día. El estudio produce beneficios en el ámbito organizacional, ya que genera un aumento de confianza en el personal y en las tareas que realizan, beneficios en el ámbito operacional y beneficios en el ámbito financiero, ya que aumenta la utilización del activo y mejora el valor entregado a los pacientes.

Una forma de mitigar las consecuencias de las inasistencias de pacientes es la programación de una oferta de sobrecupos para citas médicas, aparte de la oferta programada de citas regulares [8–10], esto es, programar más de un paciente en un mismo intervalo de tiempo y con un mismo especialista. Esta medida permite reducir el tiempo de inactividad del médico y aumentar la oferta de citas médicas hacia los pacientes. Así, surge la oportunidad de generar valor para la institución mediante el diseño de una herramienta que permita determinar en qué segmentos horarios existe la posibilidad de ofertar citas de sobrecupo, en base a una estimación de riesgo para la ocurrencia de inasistencias. En particular, este trabajo plantea que la herramienta final desarrollada permitirá a la institución mitigar la ocurrencia de inasistencias mediante el reemplazo de estas con la atención de citas de sobrecupo. Además, se espera conseguir un desempeño superior que el obtenido por una estrategia para la definición de citas de sobrecupos utilizada por la institución antes del desarrollo de este estudio, cuantificando la diferencia en el desempeño mediante indicadores claves de rendimiento que midan tanto el efecto positivo como negativo generado por la atención de citas de sobrecupo.

1.2. Objetivos del trabajo

En la siguiente tabla se presenta el objetivo general definido en este trabajo, junto con sus objetivos específicos correspondientes.

Tabla 1.1: Presentación de objetivos.

Objetivo General		
Diseñar e implementar una herramienta para definir la oferta de sobrecupos para citas médicas, en base a una predicción de inasistencias generada mediante métodos de ciencia de datos.		
Objetivos Específicos		
1. Diagnosticar la problemática abordada y definir ruta de trabajo.	2. Diseñar modelos de predicción de inasistencias.	3. Diseñar herramientas para definir la oferta de sobrecupos.
Resultados Esperados		
<ul style="list-style-type: none"> • Estudiar al cliente y entender su problemática. • Investigar las soluciones en problemáticas relacionadas presentadas en el estado del arte. • Definir la metodología de trabajo y materiales requeridos. • Analizar y procesar la fuente de datos dispuesta. 	<ul style="list-style-type: none"> • Proponer, implementar y evaluar modelos de predicción de inasistencias. • Seleccionar los modelos óptimos para construir la herramienta. • Analizar la relevancia de las variables en la predicción. 	<ul style="list-style-type: none"> • Concebir una herramienta para definir la oferta de sobrecupos. • Evaluar y validar el desempeño teórico de la herramienta. • Desplegar la herramienta. • Evaluar y validar el desempeño práctico de la herramienta. • Contrastar el desempeño de la herramienta desarrollada y la herramienta precedente.

Capítulo 2

Marco teórico

En la actualidad existe una amplia investigación con respecto al problema de la inasistencia de pacientes a citas médicas. En particular, muchos estudios se enfocan en predecir la inasistencia mediante técnicas de ciencia de datos, con una tendencia cada vez más creciente en la utilización de algoritmos de aprendizaje de máquinas.

Elgueta [11] aborda la inasistencia de pacientes a citas médicas y las cancelaciones de citas por parte de los médicos tratantes (bloqueos). Este trabajo utiliza modelos de aprendizaje automático como Random Forest, Redes Neuronales y Redes Neuronales Recurrentes para desarrollar una herramienta de gestión que define sobrecupos para citas médicas de acuerdo a grupos de riesgo, los que se definen en base a probabilidades previstas para la inasistencia y bloqueos. El conjunto de datos utilizado en este estudio comprende registros de citas médicas de un centro médico ubicado en la ciudad de Santiago, Chile. El resultado del trabajo permite detectar alrededor de un 42 % de las inasistencias al considerar un grupo del 30 % de citas con mayor probabilidad de inasistencia, y permite detectar alrededor de un 70 % de cancelaciones considerando un grupo del 10 % de citas con mayor probabilidad de cancelación. Finalmente, el estudio concluye que los modelos desarrollados permiten apoyar la gestión de la agenda médica a través de la segmentación de citas en grupos de riesgo, para así focalizar los esfuerzos en reducir las inasistencias y los bloqueos.

En esta línea, Fan y compañía [12] diseñan un modelo de predicción para la inasistencia de pacientes utilizando un conjunto de datos de citas médicas ambulatorias reservadas en línea en un hospital de China. Los investigadores utilizan algoritmos de aprendizaje de máquinas como Regresión Logística, K-Vecinos Más Cercanos (KNN), Boosting, Árboles de Decisión, Random Forest y Bagging. Específicamente, mediante el uso de Bagging, Random Forest y Boosting obtienen un desempeño en la clasificación de un AUC-ROC de 0.990, 0.987 y 0.976, respectivamente, evaluando el desempeño sobre un conjunto de validación. Este estudio propone la utilización de una variable predictora que mide la calificación recibida por los médicos en línea y, concluye que, variables descriptoras como la distancia de la residencia del paciente al hospital, el tipo de especialidad del médico tratante y la calificación del doctor, son variables críticas para predecir el comportamiento de inasistencia.

Kurasawa y compañía [13] diseñan un modelo de predicción de inasistencias para pacientes que sufren de diabetes y que llevan a cabo sus citas médicas en el hospital de la Universidad de Tokio, Japón. Este estudio plantea el uso de un algoritmo de Regresión Logística de norma

L2 y un método de 10-fold Cross-Validation para evaluar el desempeño de las predicciones. Los investigadores dividen el conjunto de datos en uno relacionado con la condición clínica del paciente y otro relacionado con hallazgos previos. En particular, obtienen un desempeño en la clasificación de un AUC-ROC de 0.958, al utilizar ambos conjuntos de datos como un solo grupo. Asimismo, obtienen una Precisión de 0.757 y un Recall de 0.659. Finalmente, los autores plantean que el descriptor más relevante para predecir la inasistencia corresponde al día de semana en el que la cita fue atendida.

Mohammadi y compañía [14] utilizan técnicas de modelamiento predictivo para desarrollar y comparar modelos de predicción para inasistencias, con objetivo de comprender la adherencia a las citas por parte de poblaciones que reciben una menor atención médica. El estudio utiliza datos provenientes de Centros de Salud Comunitarios (CHCs) de Estados Unidos, que incluyen información financiera y operativa de los centros y características demográficas y clínicas de los pacientes. En este estudio se utilizan los algoritmos de Regresión Logística, Redes Neuronales Artificiales (ANNs) y Naive Bayes, donde también aplican un método de 10-fold Cross-Validation para entrenar y evaluar el desempeño de los modelos. En particular, los autores obtienen un desempeño de un AUC-ROC de 0.860 mediante el uso del algoritmo Naive Bayes. Así mismo, estos plantean que la aplicación de técnicas de predicción ayudan a priorizar el diseño y la implementación de intervenciones que pueden mejorar la eficiencia en los CHCs para así generar una atención más oportuna.

Simsek y compañía [15] presentan una metodología híbrida basada en minería de datos para predecir la inasistencia de pacientes, emplear un procedimiento de selección de variables exhaustivo, desarrollar un modelo robusto frente al desequilibrio de datos y proporcionar a los prestadores de servicios médicos una herramienta que indica un nivel de riesgo específico de inasistencia para los pacientes. El conjunto de datos utilizado en este estudio proviene de la plataforma Kaggle. Los investigadores determinan la utilización de un algoritmo de clasificación de Redes Neuronales Artificiales (ANNs) junto con los métodos de Algoritmo Genético (Genetic Algorithm GS) para la selección de atributos y de Random Under Sampling para mejorar el rendimiento del modelo al predecir la clase minoritaria, donde esta combinación alcanza un AUC-ROC de 0.844. Los investigadores definen las puntuaciones de riesgo específicas de cada paciente aplicando un análisis de sensibilidad sobre el umbral de clasificación. El estudio concluye con la presentación de una aplicación en línea para ser utilizada como herramienta de apoyo a la decisión para los prestadores de servicios médicos.

Salazar y compañía [16] presentan una revisión sistemática del estado del arte respecto al uso de técnicas de aprendizaje de máquinas para la predicción de inasistencias de pacientes a citas médicas, analizando un total de 24 artículos. Este estudio indica que el tamaño del conjunto de datos no afecta directamente los resultados de los modelos, donde el desempeño de modelos desarrollados con grandes volúmenes de datos fue inferior o equivalente a modelos desarrollados con cantidades de datos inferiores. Por otro lado, los autores concluyen que la calidad de los datos es un factor crítico a la hora de construir los modelos de aprendizaje de máquinas, donde la mayoría de los artículos considerados construyen nuevos atributos a partir de las características existentes en el conjunto de datos original, o complementan la base con nuevos conjuntos de datos para mejorar el ajuste de los modelos. En esta misma línea, en la mayoría de los conjuntos de datos que se utilizan en los estudios se aplican técnicas de balance de clases en búsqueda de no sesgar el entrenamiento de los modelos. En lo que

respecta a las características más importantes, el estudio concluye qué variables como la edad del paciente, el historial de ausencias previas y el tiempo de espera desde el agendamiento de la cita resultan ser las características que más contribuyen a la predicción de inasistencias. Por otra parte, los autores resaltan que los algoritmos más adecuados para predecir inasistencias son los algoritmos de árboles de decisión, ya sea con un árbol de decisión binario o con métodos de ensamblaje, como Random Forest, donde en los diferentes artículos la elección de los algoritmos de mejor desempeño también se encuentra guiada por el nivel de complejidad de comprensión del modelo construido, donde se prioriza la elección de los árboles de decisión frente a algoritmos más complejos como redes neuronales, los cuales son considerados como “cajas negras” debido a su escasa interpretabilidad. Por último, los hallazgos de los autores indican que las soluciones desarrolladas aún no están ampliamente integradas en los sistemas de información de los centros médicos, donde la mayoría de los artículos estudiados discuten soluciones para mitigar las ausencias con la ayuda del aprendizaje automático, pero en sólo uno de estos se valida la solución en un experimento llevado a cabo en un entorno real.

Capítulo 3

Metodología y materiales

3.1. Metodología

Dado el objetivo general del estudio y su índole de proyecto en ciencia de datos, se utiliza una metodología Cross Industry Standard Process for Data Mining (CRISP-DM) [17]. Esta metodología entrega un marco de referencia que cubre las fases del trabajo, sus tareas respectivas y las relaciones entre estas tareas.

CRISP-DM contempla el proceso de análisis de datos como un proyecto profesional, estableciendo así un contexto que influye en la elaboración de los modelos. Este contexto tiene en cuenta la existencia de una contraparte que no es parte del equipo de desarrollo, así como el hecho de que el proyecto no acaba una vez que se define el modelo idóneo, ya que posteriormente se requiere de un despliegue.

En base a esta metodología, el ciclo de vida de un proyecto consiste en seis fases que son presentadas en la siguiente figura.

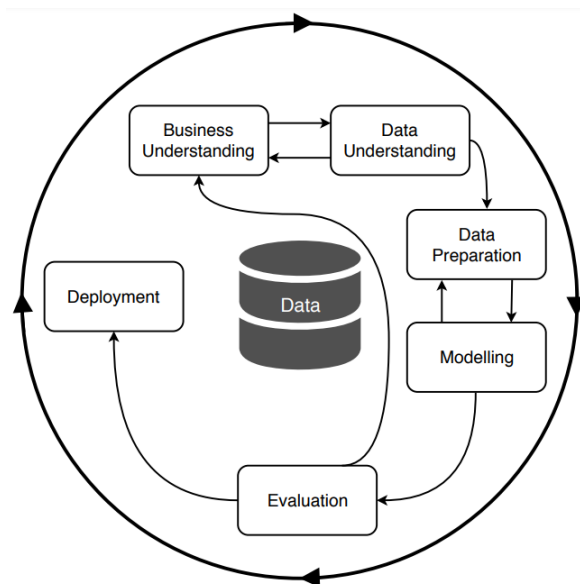


Figura 3.1: Fases CRISP-DM.

De esta forma, CRISP-DM permite estructurar el desarrollo del trabajo en base a las seis etapas presentes en la Figura 3.1: comprensión del negocio, comprensión de los datos, análisis y selección de datos, desarrollo de modelos, evaluación de modelos y, finalmente, el despliegue del modelo final. Específicamente, cada etapa contempla lo siguiente:

- **Comprensión del negocio (Business Understanding):** En esta etapa se estudia la contraparte con la que se trabaja, con objetivo de comprender su visión respecto a la problemática abordada y sus necesidades específicas.
- **Comprensión de los datos (Data Understanding):** En esta etapa se analizan y estudian los datos disponibilizados por la contraparte, con objetivo de definir el nivel de procesamiento y depuración requerido.
- **Análisis y selección de datos (Data Preparation):** En esta etapa se lleva a cabo un análisis exploratorio de datos, el cual permite seleccionar la información relevante a considerar en el desarrollo de modelos.
- **Modelamiento (Modelling):** En esta etapa se seleccionan y aplican las técnicas y modelos de ciencia de datos que son pertinentes para el problema abordado.
- **Evaluación (Evaluation):** En esta etapa se evalúan y comparan los diferentes modelos construidos, donde se califica su desempeño en las tareas encomendadas y su utilidad como herramienta para la contraparte.
- **Despliegue (Deployment):** Esta etapa considera la implementación del modelo final seleccionado. El despliegue también contempla la presentación del trabajo y conocimiento desarrollado hacia la contraparte.

3.2. Materiales

3.2.1. Sparse Group Lasso

Cuando se dispone de una gran cantidad de variables descriptoras para alimentar los algoritmos, es común buscar disminuir el tamaño de este conjunto, ya sea por la necesidad de disminuir la complejidad de los modelos o porque pueden existir variables redundantes. Un método para conseguir esta reducción en el número de variables es el uso de Sparse Group Lasso [18].

Sparse Group Lasso es una extensión del método de regularización Group Lasso [19] que también promueve la selección de características. En particular, esta extensión es la combinación de la penalización de Group Lasso y la penalización normal de Lasso.

Originalmente, el algoritmo Group Lasso [19] se define como una regresión lineal regularizada con la siguiente función de pérdida:

$$\arg \min_{\beta_g \in \mathbb{R}^{d_g}} \frac{1}{n} \left\| \sum_{g \in \mathcal{G}} [\mathbf{X}_g \beta_g] - \mathbf{y} \right\|_2^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \sum_{g \in \mathcal{G}} \sqrt{d_g} \|\beta_g\|_2 \quad (3.1)$$

donde $\mathbf{X}_g \in \mathbb{R}^{n \times d_g}$ es la matriz de datos correspondiente a las variables de un mismo grupo g , β_g son los coeficientes de regresión correspondientes al grupo g , $\mathbf{y} \in \mathbb{R}^n$ es el objetivo de

la regresión, n es el número de observaciones, d_g es la dimensionalidad del grupo g , λ_1 es la penalización de regularización a nivel de variables, λ_2 es la penalización de regularización a nivel de grupos y G es el conjunto de todos los grupos.

Este método se ha popularizado para regularizar algoritmos de aprendizaje automático. Para el caso de la regresión logística, es necesario intercambiar el término de norma cuadrada $\left\| \sum_{g \in \mathcal{G}} [\mathbf{X}_g \beta_g] - \mathbf{y} \right\|_2^2$ con un término de pérdida general, $L(\beta; \mathbf{X}, \mathbf{y})$, donde β y X son la concatenación de todos los coeficientes de grupo y las matrices de datos de grupo, respectivamente.

3.2.2. Regresión Logística

La Regresión Logística, comúnmente denominada Logit, es un algoritmo de clasificación estadístico probabilístico perteneciente a la familia de modelos lineales generalizados, en el cual se supone que la distribución de la variable de respuesta sigue una distribución exponencial.

El algoritmo Logit es utilizado principalmente para predecir variables dependientes dicotómicas, donde, a diferencia de la regresión lineal, el resultado de Logit sobre una observación es la probabilidad de que esta sea positiva o negativa (lo que se interpreta como la probabilidad de que una muestra pertenezca o no a una clase de interés), donde esta probabilidad proviene de un cálculo lineal entre los coeficientes del algoritmo y las variables de la observación.

Más formalmente, para una observación $x_i \in \mathbb{R}^p$ cuya etiqueta es y_i , se predice que la probabilidad de que y_i pertenezca a la clase positiva es la siguiente:

$$\mathbb{P}\{y_i = +1\} = \frac{1}{1 + e^{-\beta^\top x_i}} \quad (3.2)$$

donde β es el parámetro del modelo Logit.

Para definir el parámetro β que consigue el mejor ajuste, se utiliza el conjunto de observaciones de entrenamiento etiquetadas $(x_1, y_1), \dots, (x_n, y_n)$, con las cuales se obtiene el parámetro β que maximiza la función de verosimilitud sobre la muestra de entrenamiento.

3.2.3. Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) [20] es una implementación del algoritmo gradient boosting utilizando árboles de decisión (tree gradient boosting), el cual pertenece a la familia de algoritmos de Boosting.

Boosting es una técnica de ensamblaje en la que se utilizan varios modelos en conjunto, en la cual los modelos se añaden secuencialmente para corregir errores cometidos por modelos previos, hasta que no se pueden realizar más mejoras.

El algoritmo gradient boosting es un enfoque en el que se crean nuevos modelos que predicen los residuos o errores de los modelos anteriores y que luego se suman para hacer la predicción final. Se le denomina gradient boosting debido a que utiliza un algoritmo de descenso de gradiente para minimizar la función de pérdida al añadir nuevos modelos.

Formalmente, dado un conjunto de datos $D = \{(x_i, y_i)\}$ ($|D| = n, x_i \in \mathbb{R}^m, y_i \in \mathbb{R}^n$) con n observaciones y m variables descriptoras, la salida \hat{y}_i de un modelo de tree boosting con K árboles de decisión es la siguiente:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (3.3)$$

donde $F = \{f(x) = \omega_q(x)\}$ ($q: \mathbb{R}^m \rightarrow T, \omega \in \mathbb{R}^T$) es el espacio de árboles de decisión para la clasificación. Cada f_k divide un árbol en una parte de estructura q y una parte de pesos para cada hoja ω , mientras que T es el número de hojas en el árbol.

El conjunto de funciones f_k de los árboles de decisión se define minimizando la siguiente función objetivo:

$$O = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (3.4)$$

El primer término l en la Ec. 3.4 es una función de pérdida para el entrenamiento que mide la distancia entre la predicción \hat{y}_i y el valor real y_i . El segundo término Ω en la Ec. 3.4 representa el término de penalización sobre la complejidad del modelo del árbol.

El algoritmo de Tree Boosting, cuya función objetivo es la Ec. 3.4, no puede ser optimizado mediante métodos de optimización tradicionales en el espacio euclidiano. Gradient Tree Boosting es una versión mejorada de Tree Boosting en base al entrenamiento de los modelos de árbol en forma aditiva, lo que implica una predicción para una t -ésima iteración, la cual está dada por $\hat{y}^{(t)} = \hat{y}^{(t-1)} + f_t(x)$. Así, la función objetivo en la t -ésima iteración se cambia a:

$$O^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (3.5)$$

XGBoost aproxima la Ec. 3.5 utilizando una expansión de Taylor de segundo orden, con lo que la función objetivo en el paso t puede reescribirse como:

$$O^{(t)} \simeq \tilde{O}^{(t)} = \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \quad (3.6)$$

donde $g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$ y $h_i = \partial_{\hat{y}^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)})$ son el gradiente de primer y segundo orden en la función de pérdida, respectivamente, mientras que se define $\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2$ para el caso de XGBoost.

Denotando $I_j = \{i \mid q(x_i) = j\}$ como el conjunto de instancias de la hoja j y después de eliminar los términos constantes y expandir Ω , la Ec. 3.6 puede simplificarse como:

$$\tilde{O}^{(t)} = \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) \omega_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) \omega_j^2 \right] + \gamma T \quad (3.7)$$

Así, la solución para el peso óptimo ω_j^* de una hoja j para una estructura de árbol fija $q(x)$ se puede obtener aplicando la siguiente fórmula:

$$\omega_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (3.8)$$

Donde, sustituyendo ω_j^* en Ec. 3.7 se obtiene:

$$\tilde{O}(q) = -\frac{1}{2} \sum_{j=1}^T \frac{\left(\sum_{i \in I_j} g_i\right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \quad (3.9)$$

La Ec. 3.9 se define como una función de scoring para evaluar la estructura de árbol $q(x)$ y encontrar las estructuras de árbol óptimas para la clasificación. Sin embargo, en la práctica es imposible buscar todas las posibles estructuras de árbol q . En [20] se describe un algoritmo “greedy” que parte de una sola hoja y añade iterativamente ramas para hacer crecer la estructura del árbol. El hecho de añadir una división a la estructura de árbol existente puede decidirse mediante la siguiente función:

$$O_{split} = \frac{1}{2} \left[\frac{\left(\sum_{i \in I_L} g_i\right)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{\left(\sum_{i \in I_R} g_i\right)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{\left(\sum_{i \in I} g_i\right)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma \quad (3.10)$$

donde I_L e I_R son los conjuntos de instancias de los nodos izquierdo y derecho después de la división e $I = I_L \cup I_R$.

3.2.4. Procesamiento de variables

Para la construcción de modelos de clasificación se cuenta con variables numéricas y categóricas. Usualmente, las variables numéricas poseen diferentes escalas, por lo que las distancias entre sus diferentes valores tienen magnitudes diferentes. Dado esto, la estandarización de un conjunto de datos es un requisito común para muchos algoritmos de aprendizaje de máquinas.

Normalmente, la estandarización se hace eliminando la media y escalando mediante la varianza. Sin embargo, los valores atípicos (outliers) pueden influir en la media y la varianza de la muestra de forma negativa, por lo que existe un método de estandarización que es robusto frente a valores atípicos, denominado Robust Scaler.

Este escalador elimina la mediana y estandariza los datos de acuerdo con el rango intercuartílico (IQR), donde el IQR es el rango entre el primer cuartil (cuartil 25) y el tercer cuartil (cuartil 75).

$$X_{robust\ scale} = \frac{X - X_{mediana}}{X_{75} - X_{25}} = \frac{X - X_{mediana}}{IQR} \quad (3.11)$$

Otra forma de estandarizar las variables numéricas es mediante un método Min-Max Scaler. Este escalador transforma todas las variables en un rango donde el valor mínimo y máximo de una variable se encuentra entre 0 y 1, respectivamente.

$$X_{\text{minmax scale}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \quad (3.12)$$

Sin embargo, es importante tener en cuenta que la transformación mediante Min-Max Scaler es influenciada por los valores máximos y mínimos de los datos, por lo que también es sensible frente a valores atípicos.

Por otro lado, las variables categóricas son aquellas que no contemplan valores continuos, sino que representan una categoría para una observación correspondiente. Por ejemplo, una persona puede ser del sexo “Femenino” o “Masculino”, o puede realizar la creación de una reserva a través del “Portal Web”, “Chatbot”, “Call Center” o “Recepción”. Dado esto, para poder ingresar esta información a los algoritmos, es necesario codificar los valores de las variables. Para este caso, se utiliza una codificación mediante un método de One Hot Encoder, el cual transforma cada variable categórica con n valores posibles en $n - 1$ variables de categorías binarias (comúnmente denominadas dummies), donde para cada valor se crea una variable que tiene valor 1 cuando la observación pertenece a esa categoría y valor 0 cuando no.

3.2.5. Balance de datos

Trabajar con conjuntos de datos que poseen clases no equilibradas es bastante frecuente, en los cuales la clase de interés muchas veces es minoritaria, como ocurre en este caso, donde los pacientes que no asisten a sus citas son menos frecuentes que aquellos que sí asisten. Dado esto, la implementación de algoritmos de clasificación sobre conjuntos desbalanceados puede resultar en desempeños predictivos deficientes, debido a la baja cantidad de información de la clase de interés y minoritaria, con respecto a la clase mayoritaria.

Para abordar el desbalance de clases existen diferentes técnicas, donde, en este caso, se consideran dos métodos. El primer método es uno de los más directos, llamado Random Under Sampling (RUS), que consiste en eliminar registros de la clase mayoritaria seleccionados aleatoriamente en el conjunto de datos de entrenamiento, equilibrando así la cantidad de registros de cada clase. El segundo método es denominado Synthetic Minority Oversampling Technique (SMOTE) y consiste en sintetizar nuevos registros a partir de los registros existentes. En esta técnica primero se selecciona aleatoriamente un registro a de la clase minoritaria y se seleccionan los k vecinos de clase minoritaria más cercanos. Luego, el registro sintético se crea eligiendo b aleatoriamente, el cual es uno de los k vecinos más cercanos, conectando a y b para formar un segmento de línea en el espacio de características, donde las instancias sintéticas se generan como una combinación convexa de las dos instancias seleccionadas a y b .

3.2.6. SARIMA

El modelo ARIMA(p, d, q) fue introducido por primera vez por Box y Jenkin en 1976 [21] y es frecuentemente utilizado para predecir datos de series temporales estacionarias no estacionarias. Un modelo ARIMA se caracteriza por 3 términos: p , d y q , donde p es el orden del término de Auto-Regresión (AR), q es el orden del término de Media Móvil (MA) y d es el orden de diferenciación requerido para producir que la serie temporal sea estacionaria.

La Auto-Regresión (AR) no es más que la predicción de la variable mediante una regresión contra sí misma, donde se correlaciona el patrón del periodo de tiempo del paso actual con sus periodos de tiempo anteriores.

La Media Móvil (MA) es un modelo similar a la regresión que utiliza los errores asociados con el pronóstico en un paso de tiempo anterior para predecir la variable en un paso de tiempo posterior.

A continuación, se presentan las ecuaciones generalizadas del modelo AR de orden p (Ec. 3.13) y del modelo MA de orden q (Ec. 3.14).

$$y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \mathcal{E}_t \quad (3.13)$$

$$y_t = C + \mathcal{E}_t + \theta_1 \mathcal{E}_{t-1} + \theta_2 \mathcal{E}_{t-2} + \cdots + \theta_q \mathcal{E}_{t-q} \quad (3.14)$$

Los modelos ARIMA se construyen a partir de la incorporación del modelo AR, la integración (I) y el modelo MA. La integración (I) es el proceso inverso a la diferenciación para generar el pronóstico. De esta manera, el modelo ARIMA general se representa matemáticamente en la siguiente ecuación:

$$y_t = C + \phi_1 y_t + \phi_p y_{t-p} + \cdots + \phi_n y_{t-n} + \theta_1 \mathcal{E}_{t-1} + \theta_q \mathcal{E}_{t-q} + \mathcal{E}_t \quad (3.15)$$

donde C es un intercepto, $\phi_i (i = 1, 2, \dots, p)$ son los parámetros del modelo autorregresivo, $\theta_i (i = 1, 2, \dots, q)$ son los parámetros del modelo de media móvil, y_t es el valor actual de la serie temporal, $y_{t-1}, y_{t-2} \dots y_{t-p}$ son los valores pasados y \mathcal{E}_t es el error aleatorio o término residual para el paso t que viene dado por $\mathcal{E}_t = y_t - y_{t-1}$.

Por otra parte, el modelo SARIMA (Seasonal ARIMA) incluye los términos ARIMA(p, d, q) no estacionales y añade términos estacionales adicionales $(P, D, Q)_m$, para dar cuenta de la estacionalidad de los datos de la serie temporal para un número m de pasos de tiempo correspondientes a un único período estacional.

Los términos P, Q y D son el orden del término de auto-regresión estacional, el término de media móvil estacional y el término de diferenciación estacional, respectivamente. El modelo general de SARIMA se representa matemáticamente como sigue:

$$\Phi_P(B^m) \phi_p(B) (1 - B^m)^D (1 - B)^d y_t = \Theta_Q(B^m) \theta_q(B) w_t \quad (3.16)$$

donde y_t es la serie temporal, w_t es un proceso de ruido blanco gaussiano, $\phi(B)$ es el polinomio autorregresivo no estacional, $\theta(B)$ es el polinomio de media móvil no estacional y D es el término de diferenciación estacional. Generalmente, un valor de $D = 1$ es suficiente para reforzar la estacionariedad en los datos. Continuando, $\Phi_P(B^m)$ es el polinomio autorregresivo estacional y $\Theta_Q(B^m)$ es el polinomio de media móvil estacional. Por otro lado, B se define como el operador de desplazamiento hacia atrás que se expresa como $B^k y_t = y_{t-k}$.

Las expresiones para el modelo AR no estacional (Ec. 3.17), el modelo MA no estacional (Ec. 3.18), el modelo AR estacional (Ec. 3.19) y el modelo MA estacional (Ec. 3.20) se presentan a continuación:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \tag{3.17}$$

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q \tag{3.18}$$

$$\Phi_P(B^m) = 1 - \Phi_1 B^m - \Phi_2 B^{2m} + \dots + \Phi_P B^{Pm} \tag{3.19}$$

$$\Theta_Q(B^m) = 1 + \Theta_1 B^m + \Theta_2 B^{2m} + \dots + \Theta_Q B^{Qm} \tag{3.20}$$

3.2.7. ETS

Exponential Smoothing (suavización exponencial) es un algoritmo utilizado para realizar pronósticos de series temporales univariadas, donde las predicciones se realizan considerando medias ponderadas de las observaciones pasadas. Las observaciones más recientes tienen un peso exponencialmente mayor que las observaciones más antiguas.

Desde su creación en 1950, se han desarrollado diversos métodos en base a suavización exponencial. Considerando todas las modificaciones y extensiones se han desarrollado un total de quince métodos exponenciales diferentes. Estos quince métodos se discriminan en función de la naturaleza observada de una componente de tendencia (Trend) y una componente de estacionalidad (Seasonality). La componente de tendencia puede estar ausente o ser aditiva, aditiva amortiguada, multiplicativa o multiplicativa amortiguada. Por otra parte, la componente de estacionalidad puede estar ausente, ser aditiva o multiplicativa.

En [22] se introducen componentes de error aditivo y multiplicativo junto con la tendencia y la estacionalidad, y se introducen dos posibles modelos de espacio de estado para cada uno de los quince modelos, lo que da como resultado treinta modelos diferentes (Tabla 3.2). Para distinguir estos modelos, se utiliza una notación (E,T,S), la que hace referencia a las componentes de Error (E), Trend (T) y Seasonality (S). Con esto, ETS engloba las diferentes configuraciones posibles de algoritmos para el pronóstico de series temporales univariadas.

La flexibilidad del modelo ETS radica en su capacidad para modelar las componentes de tendencia y estacionalidad de diferentes rasgos, donde las diferentes configuraciones para estas componentes son las siguientes.

Tabla 3.1: Opciones para tendencia y estacionalidad.

Componente de Tendencia	Componente de Estacionalidad		
	N (Ninguna)	A (Aditiva)	M (Multiplicativa)
N (Ninguna)	NN	NA	NM
A (Aditiva)	AN	AA	AM
Ad (Aditiva Amortiguada)	AdN	AdA	AdM
M (Multiplicativa)	MN	MA	MM
Md (Multiplicativa Amortiguada)	MdN	MdA	MdM

Dada las diferentes opciones para las componentes de ETS presentes en la Tabla 3.1, se pueden obtener las siguientes combinaciones de modelos ETS.

Tabla 3.2: Configuraciones de modelos ETS.

Configuraciones		
ETS (M, M, N)	ETS (A, M, A)	ETS (M, N, M)
ETS (M, A, N)	ETS (A, Md, N)	ETS (M, N, A)
ETS (M, A, M)	ETS (A, Md, M)	ETS (M, N, N)
ETS (A, M, N)	ETS (A, N, A)	ETS (M, A, A)
ETS (A, N, N)	ETS (M, Ad, M)	ETS (A, Ad, M)
ETS (A, A, M)	ETS (M, Ad, N)	ETS (M, M, A)
ETS (M, M, M)	ETS (M, Md, M)	ETS (A, A, A)
ETS (A, N, M)	ETS (A, Ad, N)	ETS (A, Ad, A)
ETS (A, A, N)	ETS (M, Md, A)	ETS (M, Ad, A)
ETS (A, M, M)	ETS (M, Md, N)	ETS (A, Md, A)

El modelo general para todas las configuraciones presentes en la Tabla 3.2 implica un vector de estado $x_t = (l_t, b_t, s_t, s_{t-1}, \dots, s_{t-m+1})'$ y ecuaciones del espacio de estado que tienen la siguiente forma:

$$y_t = w(x_{t-1}) + r(x_{t-1}) \varepsilon_t \tag{3.21}$$

$$x_t = f(x_{t-1}) + g(x_{t-1}) \varepsilon_t \tag{3.22}$$

donde ε_t es un proceso de ruido blanco gaussiano con media cero y varianza σ^2 , y $\mu_t = w(x_{t-1})$.

El modelo con error aditivo tiene $r(x_{t-1}) = 1$, por lo que $y_t = \mu_t + \varepsilon_t$. El modelo con error multiplicativo tiene $r(x_{t-1}) = \mu_t$, lo que implica $y_t = \mu_t(1 + \varepsilon_t)$. Así, $\varepsilon_t = (y_t - \mu_t) / \mu_t$ es un error relativo para el modelo multiplicativo. El detalle específico de los modelos de espacio de estado para cada configuración de ETS se pueden revisar en [22, 23].

3.2.8. RNN

Actualmente, la posibilidad de contar con grandes cantidades de datos es cada vez mayor, y es por esto que las redes neuronales se han convertido en un técnica frecuente y dominante en tareas de aprendizaje automático. Existen diversos tipos de modelos de redes neuronales que son utilizadas para diferentes tareas, donde la red del tipo Multi-Layer Perceptron (MLP), también conocida como Fully Connected Neural Network (FFNN) o Artificial Neural Network (ANN), es el tipo más básico de red, la cual solo considera conexiones hacia adelante entre las neuronas. En cambio, las Redes Neuronales Recurrentes (RNNs) incorporan la posibilidad de utilizar bucles de retroalimentación.

La forma más habitual de introducir datos de series temporales en una red MLP consiste en dividir toda la secuencia en ventanas de entrada consecutivas, para que luego la red entregue una predicción para una ventana o un único punto de datos inmediatamente siguiente a la ventana de entrada. Dado esto, la red MLP ignora el orden temporal dentro de las ventanas y considera las nuevas entradas de manera aislada, ya que ninguna información de estado de las entradas anteriores es trasladada a los pasos de tiempo futuros. Aquí es donde entran en juego las RNNs, donde este tipo de redes son creadas especialmente para modelar datos que contemplan una dimensión temporal.

Las RNNs son la arquitectura de red neuronal más utilizada para problemas de predicción de secuencias, la cual ha ganado especial popularidad en el ámbito del procesamiento del lenguaje natural. Al igual que las redes MLP, las RNNs son aproximadores universales, sin embargo, a diferencia de las neuronas (o celdas) utilizadas en las redes MLP, las neuronas recurrentes se ocupan del orden temporal y de las dependencias temporales de las secuencias, gracias a los ciclos de retroalimentación que contemplan.

Cada RNN es una combinación de varias celdas recurrentes, donde estas unidades pueden constituir una RNN en varios tipos de arquitecturas. En particular, la arquitectura que considera capas de neuronas recurrentes apiladas una tras de otra es la arquitectura más comúnmente utilizada en tareas de pronóstico mediante el uso de RNNs.

En este trabajo se utiliza la unidad recurrente Long short-term memory, la cual es una unidad presentada en [24] y que corresponde a una variante de la unidad recurrente básica introducida en [25]. La celda LSTM es quizás la celda más popular para tareas de procesamiento del lenguaje natural, ya que puede capturar dependencias a largo plazo en la secuencia. La estructura de una celda LSTM se presenta en la siguiente figura.

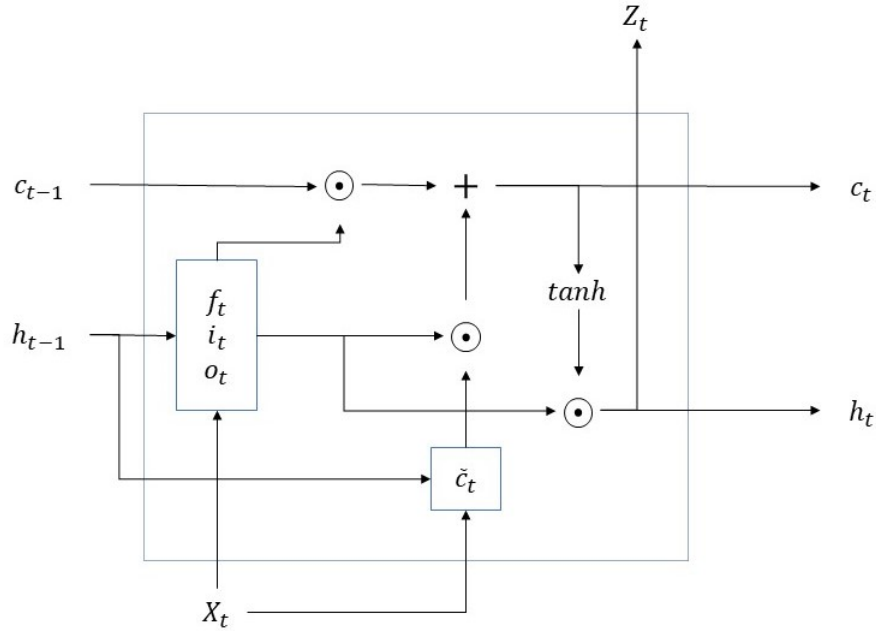


Figura 3.2: Celda long short-term memory.

Los elementos presentes en la Figura 3.2 representan los siguientes valores:

$$i_t = \sigma(W_i \cdot h_{t-1} + V_i \cdot x_t + b_i) \quad (3.23)$$

$$o_t = \sigma(W_o \cdot h_{t-1} + V_o \cdot x_t + b_o) \quad (3.24)$$

$$f_t = \sigma(W_f \cdot h_{t-1} + V_f \cdot x_t + b_f) \quad (3.25)$$

$$\tilde{C}_t = \tanh(W_c \cdot h_{t-1} + V_c \cdot x_t + b_c) \quad (3.26)$$

$$C_t = i_t \odot \tilde{C}_t + f_t \odot C_{t-1} \quad (3.27)$$

$$h_t = o_t \odot \tanh(C_t) \tag{3.28}$$

$$Z_t = h_t \tag{3.29}$$

Cada celda LSTM posee dos componentes de estado: el estado oculto y el estado interno. El estado oculto corresponde a una componente de memoria a corto plazo, mientras que el estado interno corresponde a una componente de memoria a largo plazo. Con un carrusel de errores constante basado en el estado interno de las unidades, las celdas LSTM evitan los problemas de desaparición y estallido del gradiente. Además, la celda LSTM introduce un mecanismo de compuerta que comprende puertas de entrada, olvido y salida.

En las Ecuaciones (3.23) a (3.29), $h_t \in \mathbb{R}^d$ es un vector que denota el estado oculto de la celda, donde d es la dimensión de la celda. Semejantemente, $C_t \in \mathbb{R}^d$ es la estado de la celda y $\tilde{C}_t \in \mathbb{R}^d$ es el estado candidato de la celda en el paso t , el cual captura información relevante que persiste hacia el futuro. $x_t \in \mathbb{R}^m$ (donde m es el tamaño del vector de entrada) y $Z_t \in \mathbb{R}^d$ denotan la entrada y salida de la celda en el paso t . W_i, W_o, W_f y $W_c \in \mathbb{R}^{m \times m}$ denotan la matriz de pesos en la puerta de entrada, salida y olvido, y en el estado de la celda, respectivamente. Así mismo, V_i, V_o, V_f y $V_c \in \mathbb{R}^{d \times m}$ y b_i, b_o, b_f y $b_c \in \mathbb{R}^d$ denotan la matriz de pesos del vector de entrada (en el paso t) y los vectores de sesgo, respectivamente. Por último, i_t, o_t y $f_t \in \mathbb{R}^d$ son los vectores correspondientes a las puertas de entrada, salida y olvido, respectivamente. σ corresponde a la función de activación sigmoide.

Como se aprecia en la ecuación (3.27), las puertas de entrada y olvido en conjunto determinan que tanta información correspondiente al pasado debe ser retenida en el estado actual de la celda, y que tanta información del estado actual se debe propagar hacia el futuro en los próximos pasos. El símbolo \odot denota la operación de multiplicación elemento por elemento (producto de Hadamard). Un valor de cero en la puerta de olvido f_t indica que ninguna información proveniente del estado de celda previo debe ser propagada hacia delante. Así mismo, un valor de 1 implica que el estado de celda previo debe ser completamente retenido. La noción anterior también aplica para la puerta de entrada i_t y salida o_t , donde un valor entre 0 y 1 en ambas puertas puede controlar cuidadosamente el valor del estado de celda actual utilizando solo la información que es importante, dados el estado de celda previo y el estado de celda candidato actual. Finalmente, para el estado de celda candidato se utiliza una función de activación de tangente hiperbólica (\tanh).

3.2.9. Medidas de evaluación

Rendimiento en la clasificación

El rendimiento en modelos de clasificación suele evaluarse mediante las medidas de *Precision*, *Recall*, *Accuracy* y *F1*. Estas medidas se calculan utilizando las salidas de una matriz de confusión de acuerdo a la predicción de un modelo para cada clase, donde esta matriz considera los escenarios de: verdaderos positivos (TP), verdaderos negativos (TN), falsos positivos (FP) y falsos negativos (FN). La clase positiva es la clase de interés que se desea predecir, la cual en este caso son las citas en las que el paciente no asiste (inasistencias). Con esto, los escenarios anteriores se componen de la siguiente manera.

- **Verdadero Positivo:** Cita donde el paciente no asiste y donde el modelo predice que no asiste.

- **Verdadero Negativo:** Cita donde el paciente asiste y donde el modelo predice que asiste.
- **Falso Positivo:** Cita donde el paciente asiste y donde el modelo predice que no asiste.
- **Falso Negativo:** Cita donde el paciente no asiste y donde el modelo predice que asiste.

De acuerdo con los escenarios presentados anteriormente, las medidas de evaluación se definen como sigue.

- **Accuracy:** Mide la proporción de casos en los que el modelo acierta correctamente.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.30)$$

- **Precision:** Mide la proporción de casos de la clase positiva predichos correctamente, como una proporción sobre el total de predicciones de la clase positiva realizadas.

$$Precision = \frac{TP}{TP + FP} \quad (3.31)$$

- **Recall:** Mide la proporción de casos de la clase positiva que fueron identificados correctamente, como una proporción sobre el total de casos reales observados de la clase positiva.

$$Recall = \frac{TP}{TP + FN} \quad (3.32)$$

- **F1:** Media armónica entre *Precision* y *Recall*.

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (3.33)$$

Las medidas anteriores dependen de un umbral de clasificación, es decir, un valor de corte para determinar si una predicción es clasificada como clase positiva o negativa (en este caso, inasistencia o asistencia, respectivamente), en base a la probabilidad estimada de pertenecer a la clase positiva por parte de un modelo. Dado esto, también se consideran las medidas de evaluación ROC-AUC y Precision-Recall-AUC, las cuales son independientes del umbral de clasificación.

Una curva ROC (Receiver Operating Characteristic Curve) es un gráfico que muestra el rendimiento de un modelo de clasificación para todos los umbrales de clasificación. Esta curva considera dos parámetros: una tasa de verdaderos positivos (TPR) y una tasa de falsos positivos (FPR). La tasa de verdaderos positivos es un sinónimo de *Recall*, mientras que la tasa de falsos positivos se define como sigue.

$$FPR = \frac{FP}{FP + TN} \quad (3.34)$$

La curva ROC representa la TPR frente a la FPR para diferentes umbrales de clasificación. Si se reduce el umbral de clasificación, se clasifican más elementos como positivos, lo que aumenta los falsos positivos y los verdaderos positivos. La siguiente figura muestra una curva ROC estándar.

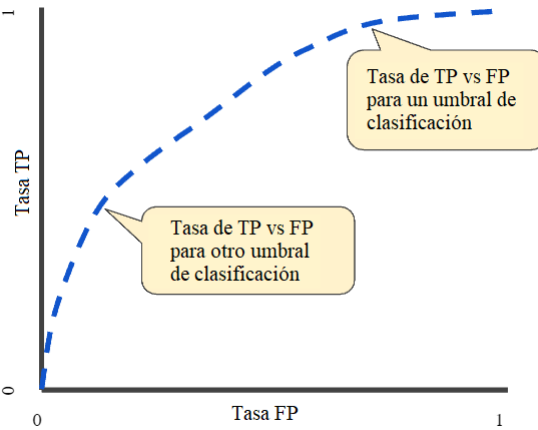


Figura 3.3: Curva ROC.

En base a la curva ROC es posible calcular el área bajo la curva (AUC). AUC mide toda el área bidimensional bajo la curva ROC completa desde (0,0) hasta (1,1), proporcionando así una medida agregada del rendimiento en todos los umbrales de clasificación posibles. Esta medida oscila entre 0 y 1, donde un modelo cuyas predicciones son 100 % erróneas tiene un AUC de 0, mientras que uno cuyas predicciones son 100 % correctas tiene un AUC de 1.

Por otra parte, una curva Precision-Recall (P-R) es un gráfico que representa la Precision y el Recall para diferentes umbrales de clasificación. En este caso, el foco de la curva P-R está en la clase de interés la cual muchas veces es minoritaria, lo que convierte a esta curva en un diagnóstico eficaz para escenarios de modelos de clasificación binaria con clases desequilibradas.

El AUC sobre la curva Precision-Recall es igual que el AUC sobre la curva ROC, ya que resume una curva con diferentes valores sobre el umbral de clasificación con una única puntuación. Esta puntuación es utilizada como punto de evaluación en un problema de clasificación binaria, donde una puntuación de 1 representa un modelo con una habilidad perfecta, mientras que una puntuación de 0 representa un modelo sin habilidad alguna.

Rendimiento en el pronóstico

Las medidas utilizadas para evaluar y comparar el rendimiento de los modelos para pronosticar una serie temporal univariada son Mean Absolute Error (MAE), Mean Absolute Scaled Error (MASE) y Root Mean Square Error (RMSE).

La definición para el cálculo de cada una de estas medidas se presenta a continuación.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \tag{3.35}$$

$$MASE = \frac{1}{n} \sum_{i=1}^n \left(\frac{|y_i - \hat{y}_i|}{\frac{1}{n-1} \sum_{i=2}^n |y_i - y_{i-1}|} \right) \tag{3.36}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \tag{3.37}$$

donde n representa el número de observaciones, y_i denota los valores reales de la serie e \hat{y}_i indica los valores pronosticados.

MAE y RMSE son medidas que dependen de la escala de los datos, aunque se basan en errores diferentes, donde la primera se basa en errores absolutos y la segunda se basa en los errores al cuadrado. Por otro lado, MASE, que se introduce en [26], es una medida de error que escala el error absoluto en base a un error absoluto promedio (MAE) calculado dentro de la muestra de entrenamiento en base a un método de pronóstico ingenuo (random walk), el cual pronostica como valor próximo el valor actual observado de la serie.

Capítulo 4

Resultados

4.1. Comprensión del negocio

El trabajo de tesis se desarrolla en colaboración con una institución chilena prestadora de servicios de salud. Esta institución cuenta con diversos centros médicos de atención ambulatoria, donde ofrece una amplia gama de prestaciones médicas.

La institución en cuestión pone a disposición un conjunto de datos que contiene información sobre citas médicas de especialidad en Medicina General llevadas a cabo entre enero de 2019 y junio de 2022, en un centro médico ubicado en la ciudad de Santiago, Chile.

En específico, se trabaja en comunicación continua con el área de planificación de la oferta médica de la institución, que tiene como principal función gestionar la oferta para toda la variedad de especialidades, exámenes de diagnóstico y tratamientos de salud, con el objetivo de coordinar la oferta con la demanda. De esta forma, el desafío principal de esta área es disponibilizar los recursos humanos competentes, para cada especialidad y durante todo el año, con el fin de alcanzar metas definidas a nivel institucional. Para esto, el área lleva a cabo un análisis de diferentes indicadores, a partir de los cuales obtiene información sobre el comportamiento de la demanda según la oferta dispuesta.

Una problemática abordada en esta área corresponde a las inasistencias de pacientes a sus citas médicas, debido a los impactos negativos que este comportamiento produce, donde es relevante para el área contar con herramientas que permitan aminorar este impacto. Dado esto, en acuerdo con el área, se propone desarrollar una herramienta que identifique bloques horarios de atención médica en los cuales existe mayor posibilidad de ocurrencia de inasistencias, para ofertar sobrecupos que permitan cubrir las citas finalmente no asistidas en dichos bloques y así contribuir a mitigar los efectos causados por las estas.

4.2. Comprensión de los datos

La información presente en los datos disponibles contempla las siguientes variables sobre la cita, el paciente y el médico tratante.

- **Variables sobre la cita:** fecha y hora de la atención, fecha y hora de la creación de la reserva, duración predeterminada, canal de creación de la reserva y estado final.

- **Variables sobre el paciente:** identificador anónimo, fecha de nacimiento, plan de salud y sexo.
- **Variables sobre el médico:** identificador anónimo.

Utilizando estos datos se construyen dos nuevos conjuntos que se usan posteriormente en la construcción de modelos de predicción. El primer conjunto contiene información sobre las citas médicas, manteniendo la misma estructura del conjunto original y conservando los registros a nivel de cita, añadiendo nuevas variables con información de la cita, del paciente y del médico. El segundo conjunto contiene información sobre la cantidad de inasistencias en función del tiempo, comprendiendo una estructura de serie temporal.

De esta forma, el primer conjunto posee 15 variables categóricas y 10 variables numéricas. Estas variables se detallan a continuación.

Tabla 4.1: Variables del primer conjunto de datos.

Nombre	Descripción
<i>Variables categóricas</i>	
Estado	Estado final de la cita: Asistencia o Inasistencia.
Tiempo visita	Segmento del día (AM-PM) con respecto a la hora de la cita.
Canal creación	Canal de creación en la reserva de la cita.
Día cita	Número de día (1-31) de la fecha de la cita.
Día creación cita	Número de día (1-31) de la fecha de creación de la reserva de la cita.
Día semana cita	Nombre del día de la fecha de la cita.
Día semana creación cita	Nombre del día de la fecha de creación de la reserva de la cita.
Duración cita	Duración predeterminada de la cita (minutos).
Hora cita	Hora de la cita (sin minutos).
Hora creación cita	Hora de la creación de la reserva de la cita (sin minutos).
Mes cita	Número de mes (1-12) de la fecha de la cita.
Mes creación cita	Número de mes (1-12) de la fecha de creación de la reserva de la cita.
Plan de salud	Plan de salud del paciente.
Sexo	Sexo del paciente.
<i>Variables numéricas</i>	
Antelación cita	Cantidad de días entre la fecha de creación de la reserva y la fecha de la cita.
Demanda médico	Cantidad de citas agendadas ^a (asistencias + inasistencias) que ha tenido el médico en los últimos 30 días de atención previos a la cita.
Demanda paciente	Cantidad de citas agendadas ^a (asistencias + inasistencias) previas a la cita por parte del paciente.
Edad paciente	Edad del paciente a la fecha de la cita.
Tasa inasistencia médico	Proporción entre la cantidad de inasistencias y el total de citas que ha tenido el médico en los últimos 30 días de atención previos a la cita.
Tasa inasistencia paciente	Proporción entre la cantidad de inasistencias previas y el total de citas previas del paciente.
Inasistencias médico	Cantidad de inasistencias que ha sufrido el médico en los últimos 30 días de atención previos a la cita.
Inasistencias paciente	Cantidad de inasistencias por parte del paciente en sus citas previas.
Atenciones médico	Cantidad de atenciones (citas asistidas) que ha tenido el médico en los últimos 30 días de atención previos a la cita.
Atenciones paciente	Cantidad de atenciones (citas asistidas) por parte del paciente en sus citas previas.

^a No incluye citas anuladas.

Las variables presentes en la Tabla 4.1 se crean en base a los datos originales disponibilizados, a los cuales previamente se les realiza una limpieza y transformación. En particular, las variables de fechas y horas son transformadas a formato horario chileno, diferenciando entre horarios de verano e invierno. Asimismo, la variable sobre el plan de salud de los pacientes es agrupada, disminuyendo su cardinalidad. Por otra parte, se eliminan valores erróneos o atípicos presentes en las variables de canal de creación, edad de los pacientes y días y horarios de citas. También se eliminan los registros duplicados y los registros donde la fecha de creación de la reserva de la cita es mayor a la fecha de atención.

Por otra parte, el segundo conjunto de datos tiene la siguiente estructura.

Tabla 4.2: Estructura del segundo conjunto de datos.

Fecha y bloque horario	Cantidad de inasistencias
2/11/2021 08:00	1
2/11/2021 09:00	0
2/11/2021 10:00	2
2/11/2021 11:00	0
2/11/2021 12:00	1
...	...
18/6/2022 09:00	1
18/6/2022 10:00	1
18/6/2022 11:00	2
18/6/2022 12:00	2
18/6/2022 13:00	0

El conjunto presentado en la Tabla 4.2 comprende una serie univariada que contiene la cantidad de inasistencias ocurridas en bloques horarios de una hora, los cuales comienzan a las 8:00 y terminan a las 20:00 hrs los días de semana, y comienzan a las 8:00 y terminan a las 14:00 hrs los días sábados. Los bloques horarios incluyen los horarios de inicio y cierre de atención del centro médico.

Esta serie de tiempo se considera sólo desde noviembre del año 2021 hasta junio de 2022, ya que, como se menciona más adelante, estos datos son utilizados para pronosticar inasistencias para las 24 primeras semanas del año 2022.

4.3. Análisis y selección de datos

En esta sección se presenta un análisis de datos de los conjuntos sobre citas médicas y cantidad de inasistencias.

En el primer conjunto se analizan las variables categóricas y numéricas con respecto al estado final de las citas, con objetivo de examinar el comportamiento de las asistencias e inasistencias en los diferentes valores de las variables y, de esta forma, formular hipótesis sobre la importancia de las variables como descriptores para la fase de modelamiento.

En el segundo conjunto se estudia la serie de tiempo, analizando su comportamiento a lo largo del tiempo y calculando estadísticas para la cantidad de inasistencias a través de bloques horarios y semanas, para así cuantificar el nivel de variabilidad de los datos.

4.3.1. Análisis del conjunto de datos sobre citas médicas

En las tablas siguientes se presenta un resumen de estadísticas descriptivas para las variables del conjunto. Se reporta la frecuencia de distribución para cada variable categórica y el promedio y desviación estándar para cada variable numérica.

Tabla 4.3: Estadísticas descriptivas de las variables categóricas (1/2).

Variables categóricas	Valores	Frecuencia	Porcentaje (%)
Estado	Asistencia	38,718	75.6
	Inasistencia	12,505	24.4
Tiempo visita	AM	18,456	36.0
	PM	32,767	64.0
Canal creación	Portal Web	37,618	73.5
	Call Center	8,102	15.8
	Recepción	5,404	10.5
	Chatbot	99	0.2
Día cita	[1, 10]	16,891	33.0
	(10, 20]	17,167	33.5
	(20, 31]	17,165	33.5
Día creación cita	[1, 10]	17,115	33.4
	(10, 20]	16,865	32.9
	(20, 31]	17,243	33.7
Día semana cita	Lunes	11,459	22.4
	Martes	10,098	19.7
	Miércoles	9,038	17.6
	Jueves	7,993	15.6
	Viernes	7,574	14.8
	Sábado	5,061	9.9
Día semana creación cita	Lunes	11,307	22.1
	Martes	9,975	19.5
	Miércoles	8,879	17.3
	Jueves	8,192	16.0
	Viernes	7,189	14.0
	Sábado	3,066	6.0
	Domingo	2,615	5.1

Tabla 4.4: Estadísticas descriptivas de las variables categóricas (2/2).

Variables categóricas	Valores	Frecuencia	Porcentaje (%)
Duración cita	15	47,180	92.1
	20	4,043	7.9
Hora cita	[8, 10]	12,800	25.0
	(10, 13]	16,416	32.0
	(13, 16]	13,335	26.0
	(16, 19]	8,672	17.0
Hora creación cita	(0, 6]	2,470	4.8
	(6, 12]	24,866	48.5
	(12, 18]	16,973	33.2
	(18, 0]	6,914	13.5
Mes cita	[1, 3]	14,480	28.3
	(3, 6]	15,062	29.4
	(6, 9]	11,829	23.1
	(9, 12]	9,852	19.2
Mes creación cita	[1, 3]	14,488	28.3
	(3, 6]	15,099	29.5
	(6, 9]	11,811	23.1
	(9, 12]	9,825	19.1
Plan de salud	Fonasa	17,204	33.6
	Isapre	15,399	30.1
	Isapre Institución ^a	11,945	23.3
	Particular	6,675	13.0
Sexo	Femenino	29,826	58.2
	Masculino	21,397	41.8

^a Isapre ligada a la institución.

Tabla 4.5: Estadísticas descriptivas de las variables numéricas.

Variables numéricas	Promedio	Desviación estándar
Antelación cita	1.0	2.1
Demanda médico	387.1	187.5
Demanda paciente	2.2	2.1
Edad paciente	38.4	14.8
Tasa inasistencia médico (%)	24.8	10.6
Tasa inasistencia paciente (%)	17.7	33.4
Inasistencias médico	93.8	54.4
Inasistencias pacientes	0.4	0.9
Atenciones médico	293.3	151.9
Atenciones paciente	1.8	1.8

En las Tablas 4.3 y 4.4 se observa que aproximadamente un 24 % de las citas son inasistencias. Por otro lado, la mayoría de las citas son programadas en las tardes, a través del portal web de la institución y con médicos que atienden con una duración predeterminada de 15 minutos. Asimismo, las citas son programadas mayormente por personas de sexo femenino y con plan de salud Fonasa o Isapre.

En la Tabla 4.5 se observa que las citas en promedio son agendadas con un día de anticipación, donde los pacientes tienen una edad promedio de 38 años y un promedio de dos citas agendadas históricamente.

En los siguientes gráficos se presenta la distribución de la variable de interés (Estado) a través de los diferentes valores de las variables categóricas.

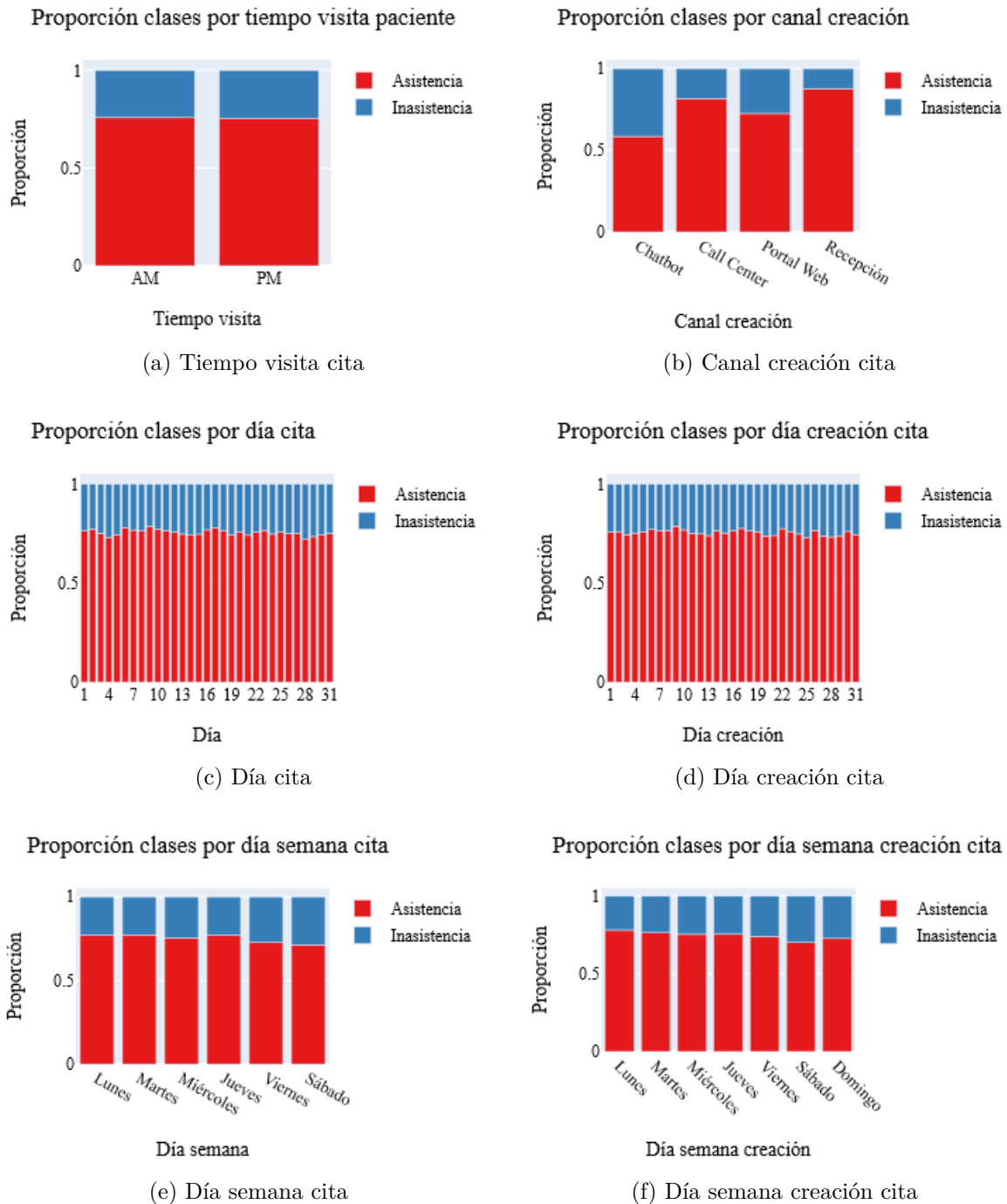


Figura 4.1: Comportamiento variable Estado a través de variables categóricas (1/3).

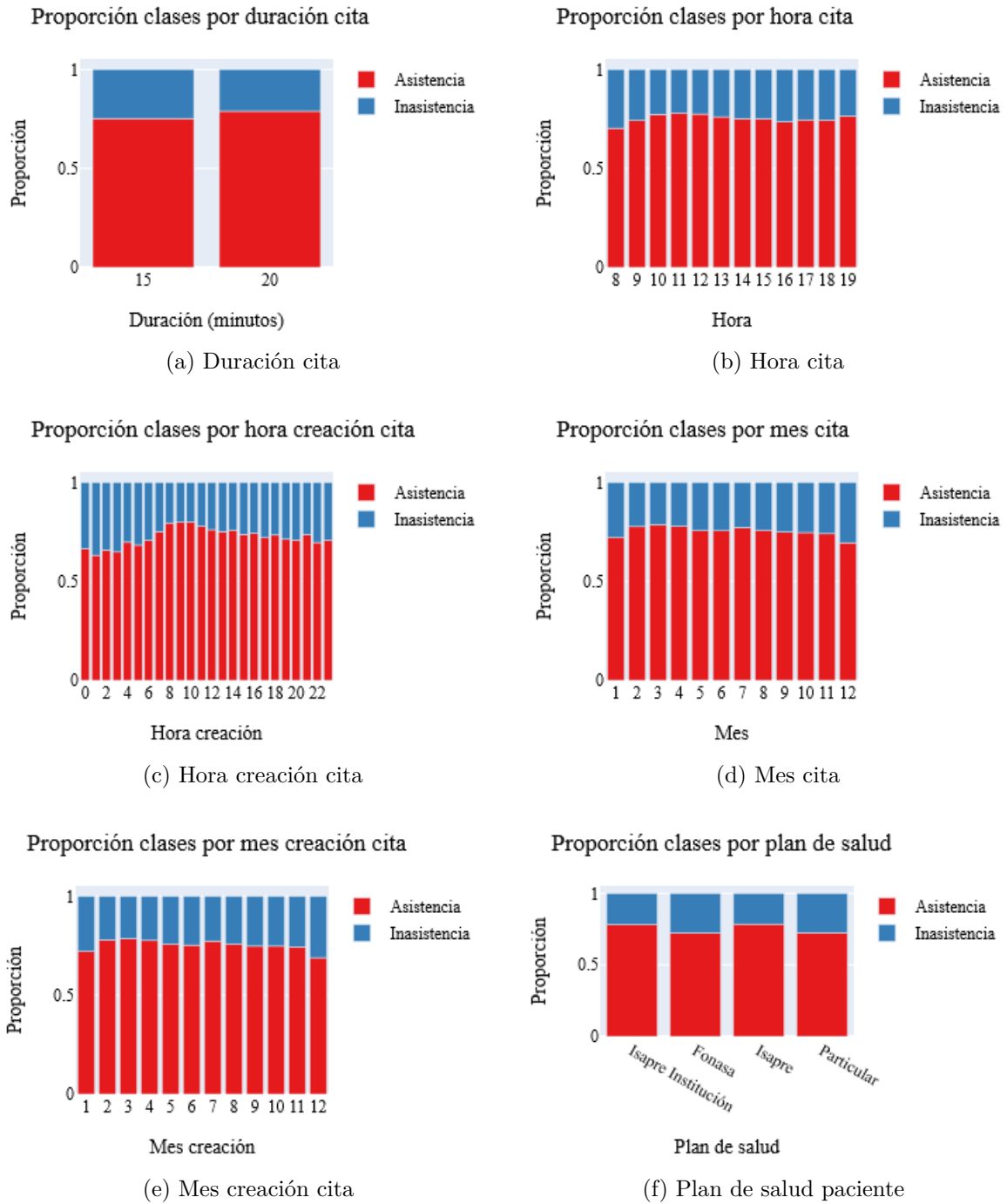
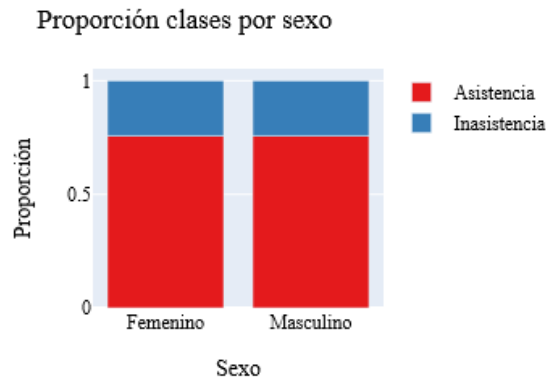


Figura 4.2: Comportamiento variable Estado a través de variables categóricas (2/3).

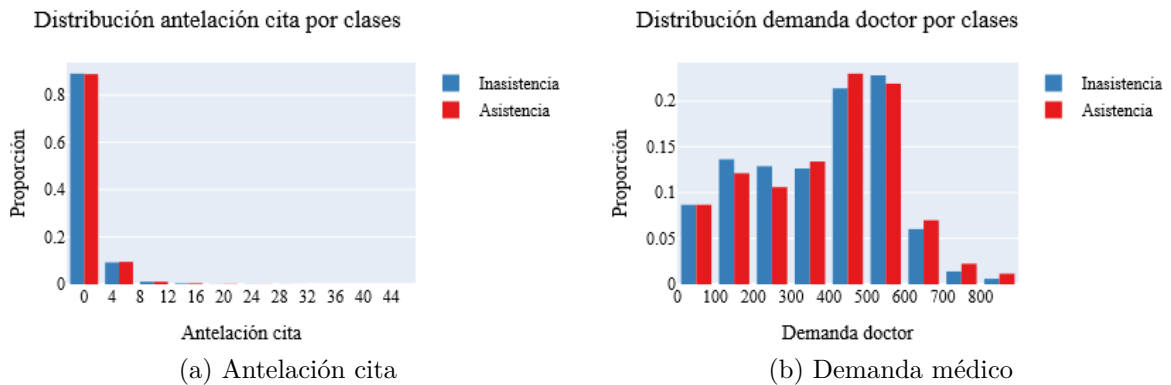


(a) Sexo paciente

Figura 4.3: Comportamiento variable Estado a través de variables categóricas (3/3).

En las Figuras 4.1, 4.2 y 4.3 se observa que la mayoría de las variables no poseen diferencias considerables en la proporción de asistencias e inasistencias sobre sus diferentes valores, a excepción de canal de creación y hora de creación, en las cuales si se observa una distribución heterogénea entre clases. Por otra parte, hay variables en las cuales prácticamente no existe diferencia en la proporción entre clases, como en sexo del paciente y tiempo de visita. Con esto, se espera que canal de creación y hora de creación sean variables descriptoras importantes, mientras que sexo del paciente y tiempo de visita sean variables descriptoras irrelevantes.

Siguiendo con el análisis, a continuación se presenta la distribución de la variable de interés a través de los diferentes valores de las variables numéricas. Para esto, se construyen histogramas con la proporción de datos para cada clase, donde la proporción correspondiente a cada intervalo es calculada con respecto al total de datos de cada clase, siendo así ambas distribuciones comparables.



(a) Antelación cita

(b) Demanda médico

Figura 4.4: Comportamiento variable de interés a través de variables numéricas (1/3).

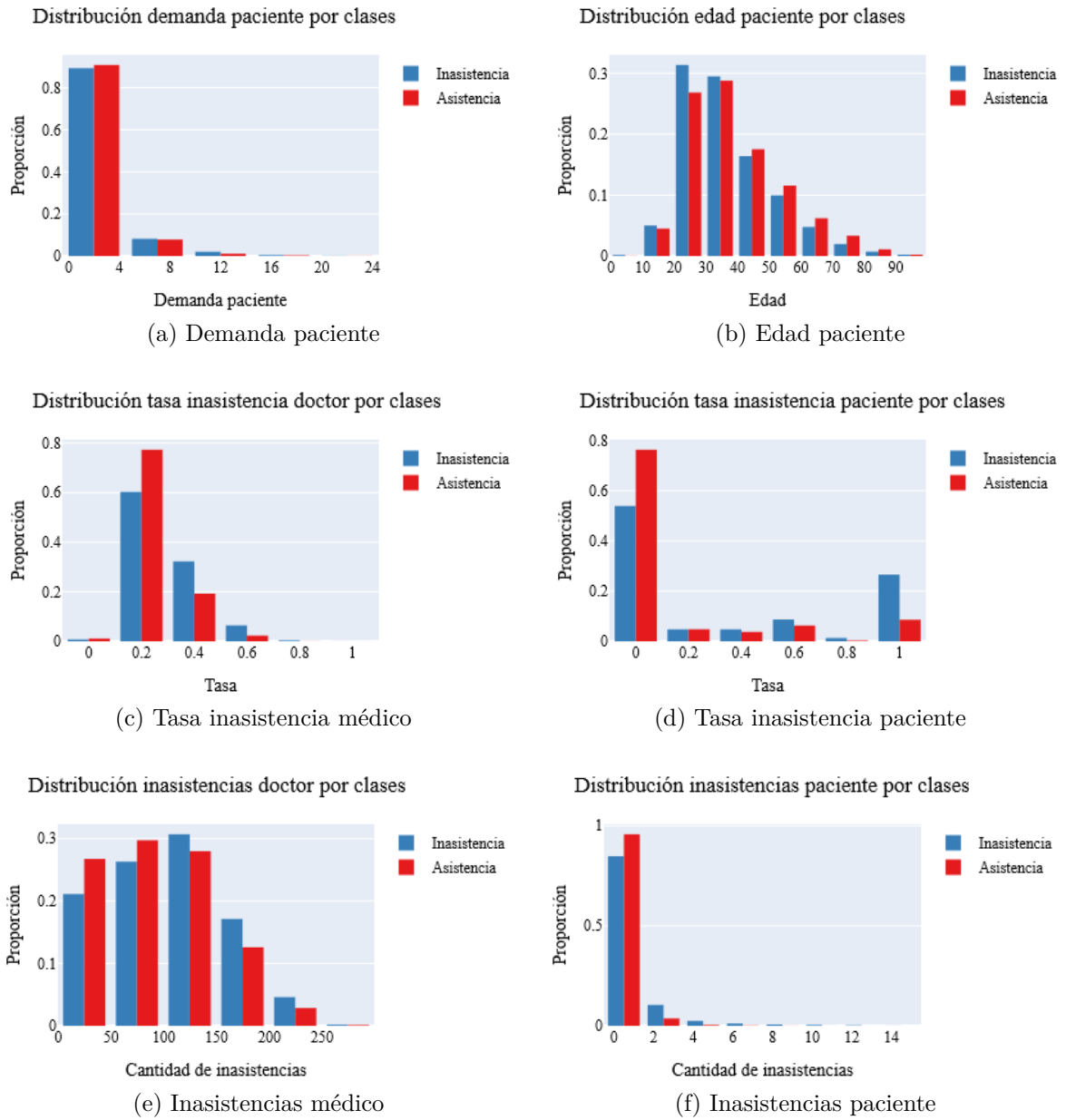


Figura 4.5: Comportamiento variable de interés a través de variables numéricas (2/3).

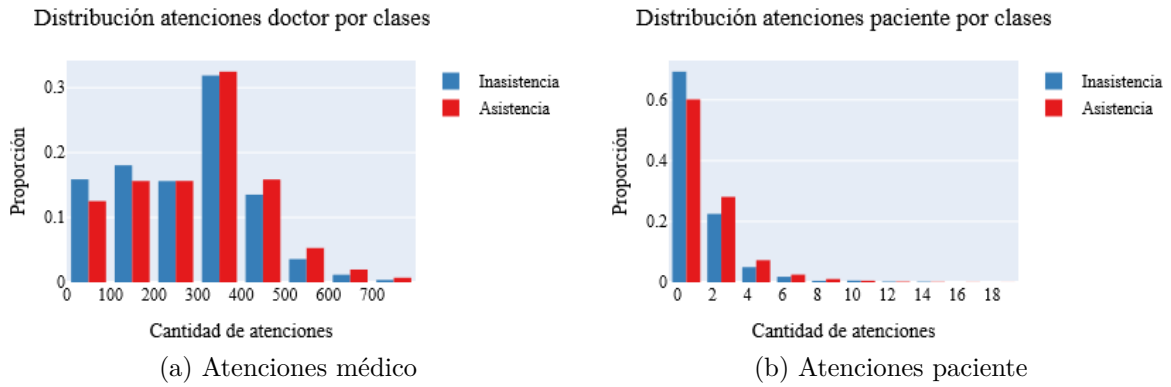


Figura 4.6: Comportamiento variable de interés a través de variables numéricas (3/3).

En las Figuras 4.4, 4.5 y 4.6 se observa que en ciertas variables la diferencia en la distribución entre las clases de asistencia e inasistencia es bastante pequeña, como en el caso de antelación cita y demanda paciente, con distribuciones por clase prácticamente idénticas. Por otra parte, variables como tasa inasistencia paciente y tasa inasistencia médico si tienen diferencias en sus distribuciones, donde, en particular, la variable tasa inasistencia paciente tiene una mayor cantidad de valores cercanos a 1 en la clase inasistencia, mientras que en la variable tasa inasistencia médico los valores más bajos son más frecuentes en la clase asistencia.

A priori, no se descartan variables en base al análisis de datos, ya que en la fase de modelamiento se diseñan modelos con y sin la utilización de técnicas de selección de variables.

4.3.2. Análisis del conjunto de datos sobre cantidad de inasistencias

En el siguiente gráfico se presenta la serie de tiempo construida a lo largo de las diferentes fechas y bloques horarios considerados.

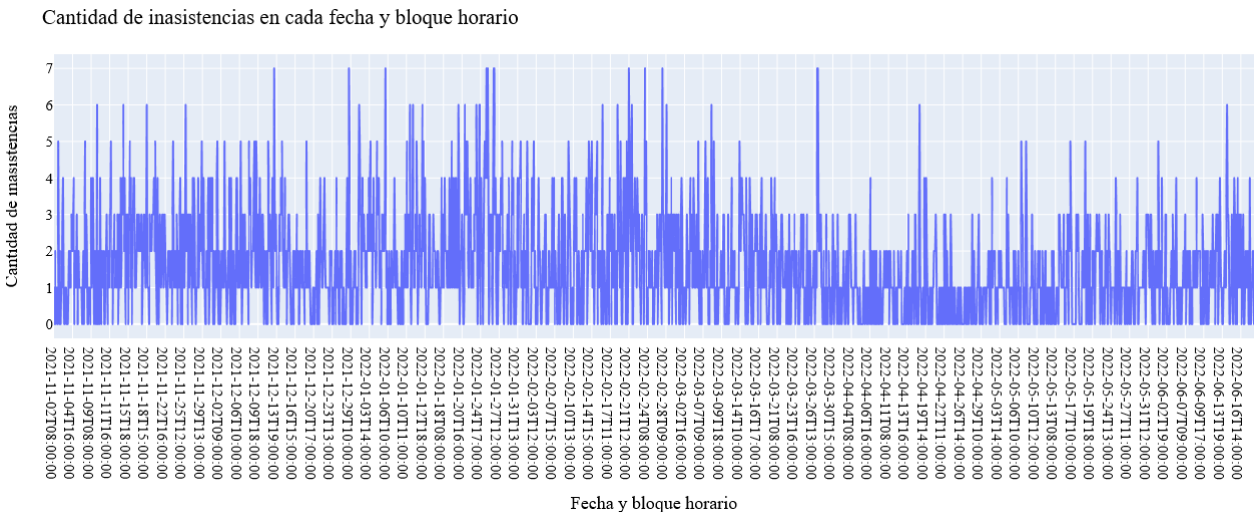


Figura 4.7: Cantidad de inasistencias por fechas y bloques horarios.

En la Figura 4.7 se observa una componente considerable de variabilidad en la cantidad de inasistencias. En particular, la totalidad de esta serie tiene un promedio de 1.57 y una desviación estándar de 1.43, lo que implica un coeficiente de variación de 0.91, contando así con una serie con un nivel de variación medio-alto.

Con respecto al comportamiento de inasistencias en las semanas, en la siguiente figura se presentan diagramas de caja (o box plots) para representar gráficamente la serie de datos. Cada diagrama presenta una serie de largo igual a la cantidad total de bloques horarios de atención de cada semana, mientras que los valores de la serie representan la cantidad de inasistencias observadas en cada bloque horario. Se presentan las series correspondientes a las semanas 44 a 52 de 2021, y a las semanas 1 a 24 de 2022.

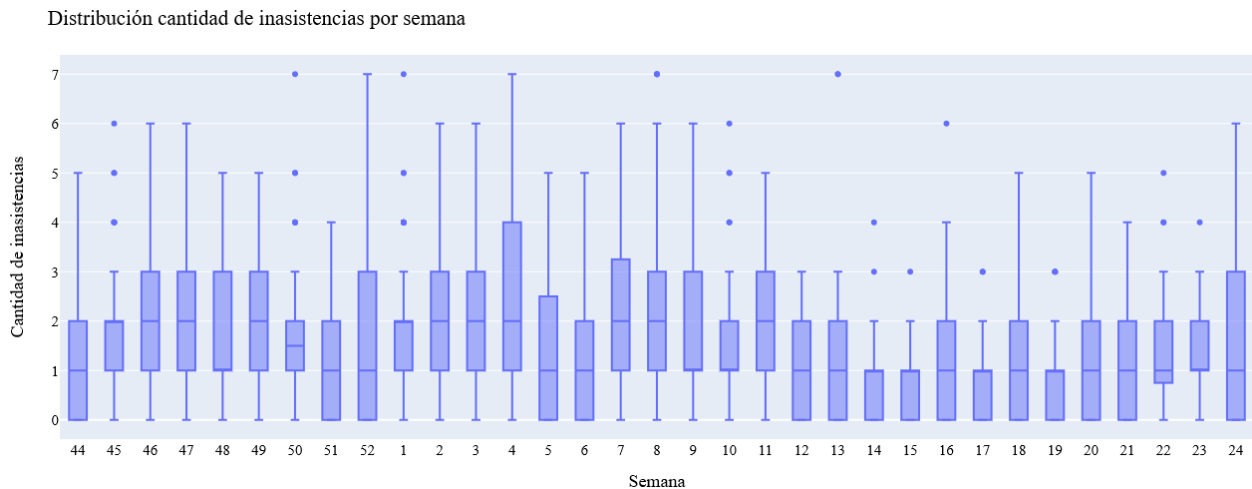


Figura 4.8: Distribución cantidad de inasistencias en bloques horarios por semana.

En la Figura 4.8 se observa que la distribución no es homogénea entre semanas, donde la cantidad de inasistencias ocurre con mayor frecuencia entre 0 y 3 en la mayoría de las semanas. También se aprecia que existen semanas con una distribución de inasistencias concentrada, como en el caso de las semanas 14, 15, 17 y 19 de 2022, mientras que en otras semanas la distribución es dispersa, como en el caso de la semana 4 de 2022. Por otro lado, existen valores que se pueden considerar como atípicos para todas las semanas, donde una cantidad de 6 o 7 inasistencias presentes en un bloque horario es poco frecuente.

Con respecto al comportamiento de inasistencias en los bloques horarios, en la siguiente tabla se presentan estadísticas descriptivas para la cantidad de inasistencias por bloque.

Tabla 4.6: Estadísticas descriptivas sobre cantidad de inasistencias por bloque horario.

Inicio bloque horario	Cantidad de bloques	Promedio	Desviación Estándar	Mínimo	25 %	50 %	75 %	Máximo
8	174	1.04	1.10	0	0	1	2	5
9	183	1.54	1.37	0	1	1	2	7
10	189	1.55	1.37	0	1	1	2	6
11	189	1.68	1.39	0	1	2	3	6
12	189	1.76	1.53	0	1	2	3	7
13	186	1.10	1.19	0	0	1	2	7
14	159	1.67	1.28	0	1	2	2.5	6
15	159	2.04	1.79	0	1	2	3	7
16	159	2.09	1.41	0	1	2	3	7
17	159	2.08	1.53	0	1	2	3	7
18	155	1.63	1.39	0	1	1	3	6
19	106	0.36	0.54	0	0	0	1	2

En la Tabla 4.6 se observa que en prácticamente todos los bloques se tienen promedios de inasistencias mayores a 1, con excepción del bloque de 19:00 a 20:00 hrs. Por otra parte, nuevamente se aprecia un nivel de variación considerable, existiendo horarios en los que la desviación estándar es mayor al promedio. A su vez, se nota que la cantidad de observaciones en los diferentes bloques horarios no es constante, donde existen bloques con una mayor cantidad de registros, como los bloques de 10:00 a 11:00, 11:00 a 12:00 y 13:00 a 14:00 hrs, y bloques con una cantidad menor de registros, como el bloque de 19:00 a 20:00 hrs. Otro aspecto a destacar es que en todos los bloques el mínimo de cantidad de inasistencias es 0, donde la serie temporal contiene una cantidad considerable de valores nulos, lo que también se puede observar en la Figura 4.8.

De acuerdo con el análisis anterior, es necesario procesar la serie de tiempo en la fase de modelamiento, con objetivo de estabilizar el nivel de varianza y tratar valores faltantes en los bloques horarios.

4.4. Modelamiento

En la fase de modelamiento, el objetivo es diseñar un modelo de predicción que permita definir segmentos en los cuales ofertar sobrecupos para reservas de citas médicas. Para esto, se diseñan modelos en base a dos enfoques:

- **Predicción de inasistencia por cita:** predecir una probabilidad de que el paciente no asista la cita, para así categorizar como inasistencias aquellas citas con una probabilidad que supera un umbral determinado.
- **Pronóstico de cantidad de inasistencias:** pronosticar la cantidad de inasistencias que se producirán en bloques horarios de atención de una hora de largo.

Ambos enfoques tienen como finalidad entregar información base que luego permita definir fechas y bloques horarios en los cuales ofertar sobrecupos.

A continuación, se detalla el proceso de modelamiento seguido en cada enfoque.

4.4.1. Predicción de inasistencia por cita

Este enfoque considera un problema de clasificación binaria, donde la variable de predicción es el estado final de una cita, el cual contempla dos resultados: Asistencia o Inasistencia.

Estructura de datos

Los datos que se utilizan comprenden dos grupos de pacientes, aquellos que asisten por primera vez al centro (pacientes nuevos) y aquellos que ya han asistido anteriormente dentro de la ventana de datos considerada (pacientes recurrentes).

La diferencia entre estos grupos es la cantidad de variables predictoras con la que cada uno dispone, ya que los pacientes nuevos no tienen información histórica sobre su comportamiento respecto a asistencias e inasistencias, mientras que los pacientes recurrentes sí cuentan con esta información. Dado esto, se diseñan modelos de predicción diferenciando entre conjuntos de pacientes nuevos y recurrentes.

En la siguiente figura se presenta la cantidad de observaciones y variables descriptoras disponibles para cada grupo.

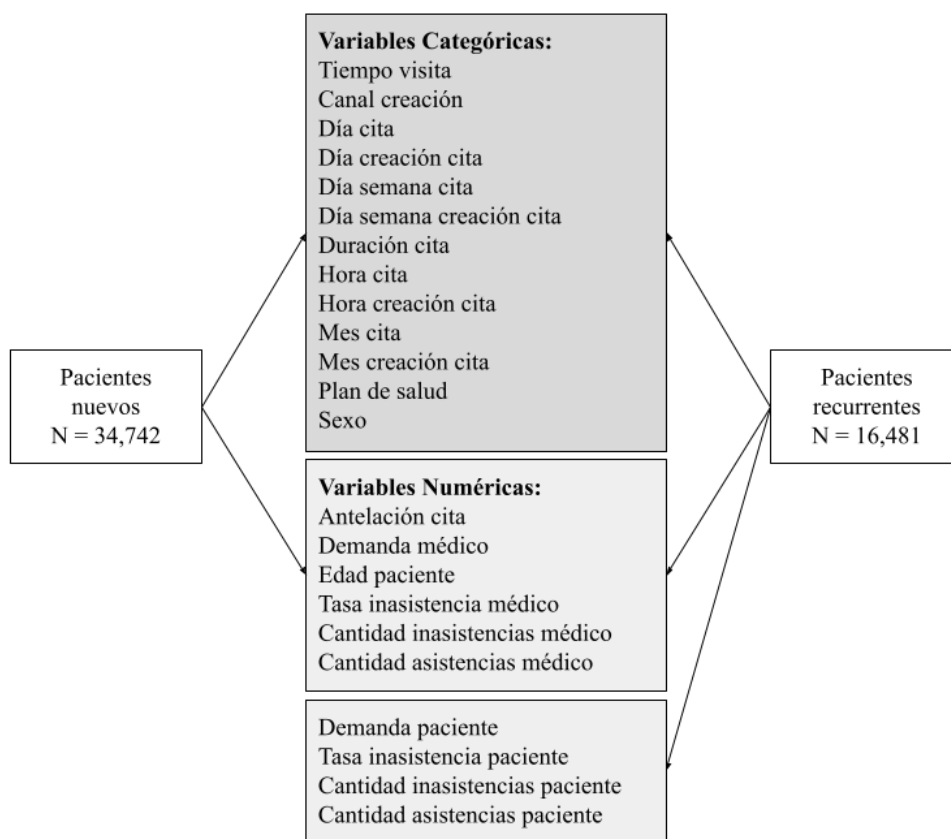


Figura 4.9: Variables disponibles para cada grupo de pacientes.

Como se observa en la Figura 4.9, el conjunto de pacientes nuevos comprende 34,742 registros y dispone de 19 variables, mientras que el conjunto de pacientes recurrentes comprende 16,481 registros y dispone de 23 variables.

Selección de variables

En búsqueda de eliminar variables predictoras colineales redundantes, se implementa una selección de variables reconociendo las variables más relevantes para la predicción mediante un método Sparse Group Lasso. Los resultados de esta selección de variables se presentan en las Figuras A1 y A2 del Anexo A.

Sin perjuicio de lo anterior, en vista de que ciertos algoritmos de aprendizaje de máquinas utilizados en este trabajo tienen una capacidad intrínseca de selección de variables como parte de su modelo de aprendizaje [27] (por ejemplo, los modelos de ensamblaje como Random Forest, Bagging y Boosting, utilizan criterios de ganancia de información en su estructura de aprendizaje y, por ende, son capaces de seleccionar qué variables entregan mayor información para la clasificación), también se considera la inclusión de todas las variables disponibles en el entrenamiento de los modelos, para así comparar los desempeños de la predicción con y sin la utilización de selección de variables.

División del conjunto de datos

Los datos son divididos en tres conjuntos, un conjunto de entrenamiento, un conjunto de testing y un conjunto de proyección.

Para definir los primeros dos conjuntos se utilizan los datos de los años 2019, 2020 y 2021 y se divide aleatoriamente un 75 % como conjunto de entrenamiento y un 25 % como conjunto de testing, manteniendo la proporción de las clases (asistencia e inasistencia) en la división.

Por otra parte, los datos del año 2022 conforman el conjunto de proyección, el cual se utiliza para evaluar el rendimiento de los modelos en datos fuera de la muestra de entrenamiento y testing, donde, el conjunto de proyección, a diferencia del conjunto de testing, considera datos continuos en el tiempo, que son más recientes y que no contemplan una selección aleatoria.

Comparar el desempeño de los modelos en el conjunto de testing y de proyección permite dimensionar la capacidad de generalización de los modelos, es decir, la capacidad de estos de equiparar su desempeño en nuevos datos.

En la siguiente figura se esquematiza la división de datos y la cantidad de observaciones que contiene cada conjunto.

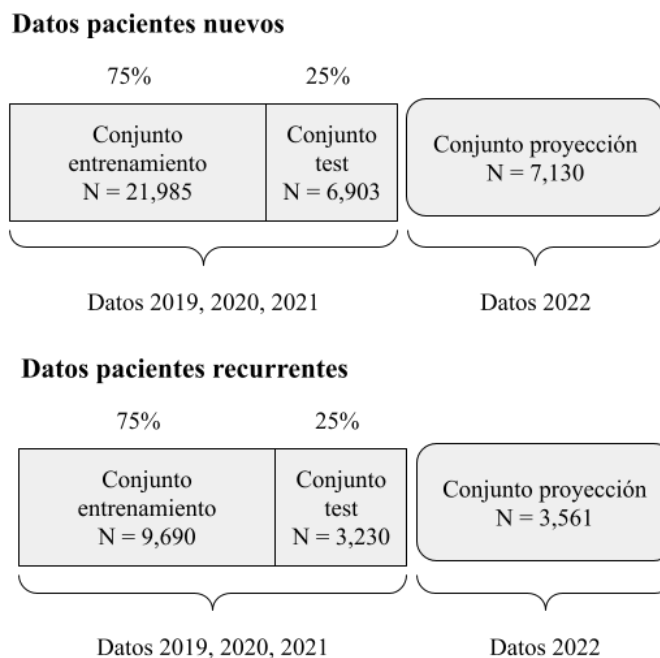


Figura 4.10: División del conjunto de datos para modelos de clasificación.

Modelos de clasificación

En esta etapa se consideran los siguientes algoritmos de aprendizajes de máquinas para el diseño de un modelo de predicción de inasistencias: Regresión Logística (Logit), Random Forest (RF), Bagging, AdaBoost, XGBoost, K-Vecinos Más Cercanos (KNN), Naive Bayes y una Red Neuronal Multi-Layer Perceptron (MLP).

Para cada uno de los algoritmos se realiza una grilla de búsqueda sobre los valores para sus hiperparámetros que obtienen el mejor resultado, utilizando un método de Cross Validation (CV). CV divide el conjunto de entrenamiento en k conjuntos más pequeños, donde se entrena el modelo utilizando $k - 1$ subconjuntos como datos de entrenamiento y se evalúa en el subconjunto restante, repitiendo este procedimiento k veces. Para este caso, se selecciona $k = 5$, determinando la mejor configuración de hiperparámetros como aquella que obtiene el valor de ROC-AUC más alto.

Procesamiento de variables y balance de datos

Previo al entrenamiento de los algoritmos, las variables numéricas son escaladas utilizando un método de Robust Scaler, mientras que las variables categóricas son codificadas utilizando un método de One-Hot-Encoding.

Por otra parte, la distribución de clases entre citas asistidas y no asistidas no se encuentra balanceada, existiendo una relación de aproximadamente 2:5 entre citas no asistidas y asistidas en el caso de pacientes nuevos, y una relación de aproximadamente 1:5 entre citas no asistidas y asistidas en el caso de pacientes recurrentes. Dado esto, se entrenan los algoritmos considerando la distribución de clases natural y considerando el uso de las técnicas de balance de datos Random Under Sampling (RUS) y Synthetic Minority Oversampling Technique (SMOTE). Específicamente, se busca obtener una relación de 1:2 entre citas no asistidas y

asistidas, por lo que en el conjunto de datos de pacientes nuevos se utiliza únicamente SMOTE, mientras que, en el conjunto de datos de pacientes recurrentes se utiliza en primer lugar RUS para obtener una relación de 1:4 y luego se utiliza SMOTE para obtener la relación final buscada.

4.4.2. Pronóstico de cantidad de inasistencias

Para este enfoque se considera el pronóstico de una serie de tiempo univariada. De esta forma, la variable dependiente es la cantidad de inasistencias que ocurren en un intervalo de una hora a lo largo de los días, donde los valores futuros para esta variable son pronosticados en base a los valores pasados de esta misma.

Estructura de datos

La estructura de datos del conjunto que se utiliza en este enfoque se presenta en la Sección 4.2, donde la base de datos contiene la cantidad de inasistencias que se presentan en cada fecha y bloque horario de atención en base a la información de todas las citas médicas. En particular, las características y comportamiento de esta estructura se presentan en la Sección 4.3.2.

Modelos de pronóstico

Para predecir la serie temporal en cuestión se utilizan los algoritmos SARIMA, ETS y una Red Neuronal Recurrente (RNN). Los modelos SARIMA y ETS son métodos tradicionalmente utilizados en el modelamiento univariado que han obtenido resultados dominantes en competencias de pronóstico, mientras que las Redes Neuronales Recurrentes son métodos de aprendizaje profundo que recientemente han ganado popularidad como modelos para el pronóstico de series de tiempo [28].

Los modelos considerados poseen diferentes hiperparámetros que deben ser definidos previo a su implementación. En el caso de SARIMA, se genera una búsqueda automática de la mejor configuración de hiperparámetros en base al criterio de información de Akaike (AIC), haciendo uso de la funcionalidad incluida en la librería utilizada para su implementación, donde se seleccionan los hiperparámetros que obtienen el menor AIC. En el caso de ETS, se consideran diversas combinaciones de hiperparámetros implementando cada una de ellas y seleccionando la mejor configuración en base al Mean Absolute Scaled Error (MASE) obtenido, seleccionando la combinación que obtiene el menor MASE. Con respecto a la Red Neuronal Recurrente, existen múltiples arquitecturas que pueden ser utilizadas para generar un pronóstico de una serie temporal. En este caso, en base al estudio realizado en [28], en el cual se evalúa el desempeño obtenido en diferentes tareas de pronóstico por diversas arquitecturas de RNN, se utiliza la arquitectura que obtiene el mejor desempeño en el estudio junto con el uso de celdas recurrentes del tipo LSTM. Específicamente, la arquitectura seleccionada contempla capas LSTM apiladas una tras otra, donde la última capa LSTM se conecta con una capa densa que luego conecta con la capa de salida, la cual finalmente entrega un vector con los valores pronosticados para la serie correspondiente. La arquitectura descrita anteriormente se presenta en el siguiente esquema.

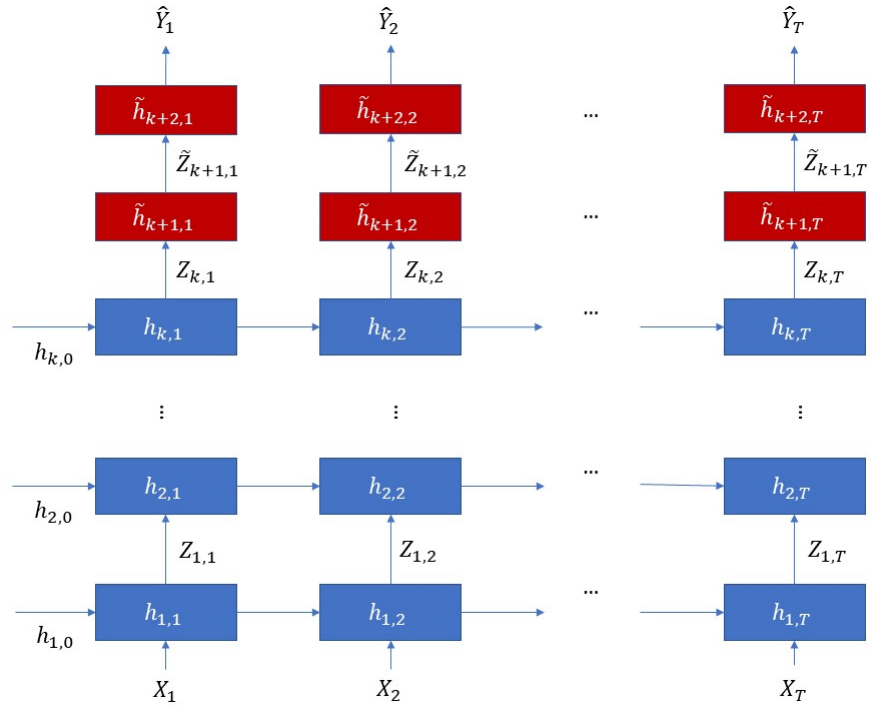


Figura 4.11: Arquitectura RNN utilizada.

En la Figura 4.11, los cuadros azules representan capas que utilizan celdas (o neuronas) de tipo LSTM, mientras que los cuadros rojos representan las capas densas que utilizan celdas convencionales de una Red Neuronal Multi-Layer Perceptron. La capa $k + 1$ es una capa completamente conectada con la última capa LSTM y la capa $k + 2$ es la última capa que no posee función de activación y que entrega el vector de la serie temporal pronosticada. El valor que representa cada elemento de la figura se encuentra en la Sección 3.2.8.

Por otra parte, también es necesario definir hiperparámetros para la red RNN, los cuales comprenden el número de celdas en cada capa, el número k de capas LSTM y la función de activación en cada capa, mientras que, para el entrenamiento de la red, es necesario definir el número de épocas, el tamaño del batch de datos de entrada y la tasa de aprendizaje. Para esto, se evalúan diferentes combinaciones de hiperparámetros y se selecciona la combinación que obtiene el menor valor de Mean Absolute Error (MAE). Con esto, los hiperparámetros que obtienen el mejor desempeño corresponden a 10 celdas por capa, una capa LSTM ($k = 1$), tangente hiperbólica como función de activación en cada capa, 50 épocas para el entrenamiento, un tamaño de batch de entrada de 64 observaciones y una tasa de aprendizaje de 0.1 utilizando Adam como optimizador.

División del conjunto de datos

Generalmente, los modelos de pronóstico consideran una ventana de entrenamiento y predicción móvil en el tiempo, ya que estos utilizan los valores pasados de la serie para pronosticar los valores futuros. En este caso, se busca pronosticar la cantidad de inasistencias en cada fecha y bloque horario de una semana para las primeras 24 semanas del año 2022, generando un vector de predicción de largo igual a la cantidad de bloques horarios de atención presentes en cada semana.

Dado esto, es necesario definir la cantidad de semanas que contendrá la serie de entrenamiento, para lo cual se evalúa el desempeño de la predicción utilizando 4, 5 y 6 semanas de entrenamiento, donde se selecciona la cantidad de semanas que obtiene el menor MASE promedio calculado entre las 24 semanas de predicción. De esta forma, se seleccionan 5 semanas para entrenar los modelos SARIMA en días de semana y sábado, 5 semanas para entrenar los modelos ETS en días de semana y 6 semanas para entrenar los modelos ETS en días sábados.

Por otra parte, en el entrenamiento del modelo RNN se consideran las 4 semanas previas a la semana de pronóstico, donde la serie de entrenamiento se divide en múltiples ventanas las cuales corresponden a un segmento de la serie desplazada hacia adelante en un paso, ya que este mecanismo de división resulta eficaz para el aumento de datos, más aún en el entrenamiento de un modelo de red neuronal en el cual es importante contar con una amplia cantidad de registros.

De esta manera, para el entrenamiento de cada uno de los modelos se utiliza la cantidad de semana definidas anteriormente y se pronostica la semana siguiente a la última semana de entrenamiento. Posteriormente, los resultados obtenidos en la predicción para cada semana son evaluados en base a los valores reales observados.

A continuación, se presenta un esquema de la división móvil entre el conjunto de entrenamiento y predicción para la serie de tiempo.

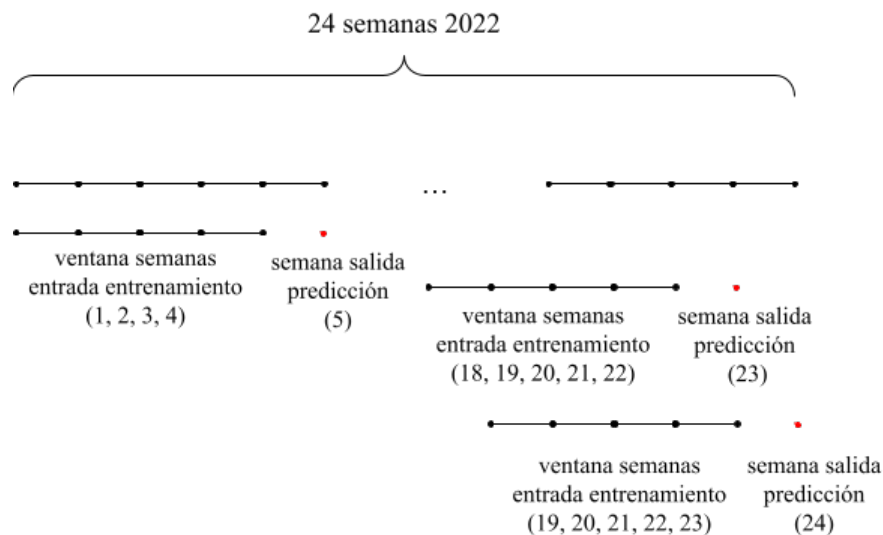


Figura 4.12: División del conjunto de datos, modelos de pronóstico.

Para pronosticar las 24 primeras semanas del año 2022 se considera la serie de tiempo desde noviembre de 2021 en adelante, ya que para generar la predicción de las primeras semanas de 2022 es necesario contar con los datos de las últimas semanas de 2021.

Tratamiento de valores faltantes

El entrenamiento de los algoritmos se realiza utilizando los datos de la serie temporal dividida en semanas, donde cada semana considera 6 días (lunes, martes, miércoles, jueves, viernes y sábado), 12 bloques horarios en cada día de semana y 5 bloques horarios en cada

día sábado.

Los bloques de inicio y fin de atención (bloques de 8:00 a 9:00 hrs y 19:00 a 20:00 hrs en días de semana, y, 8:00 a 9:00 hrs y 13:00 a 14:00 hrs en días sábados) son removidos de la serie temporal, ya que no se contempla la oferta de sobrecupos en los bloques de inicio y fin de atención del centro. De esta forma, la cantidad de bloques considerada por día es de 10 para los días de semana y de 4 para los días sábados.

Como se menciona en la Sección 4.3.2, no todos los bloques horarios tienen la misma cantidad de observaciones en el conjunto, ya que existen bloques en los que no ocurren atenciones por parte de los doctores y que no poseen información sobre la cantidad de inasistencias en el bloque. Dado esto, en aquellos bloques donde no hay información, se reemplaza el valor faltante de inasistencias por un promedio de inasistencias calculado a lo largo del mismo bloque en días pasados, considerando las últimas 4 semanas anteriores a la fecha del día del bloque imputado. De esta forma, todos los días en los datos consideran 10 bloques horarios en los días de semana y 4 bloques horarios en los días sábados. Cabe mencionar que la imputación de información se realiza sólo en aquellos días donde la cantidad de bloques faltantes es máximo uno, dejando fuera de los datos aquellos días donde existe más de un bloque con datos faltantes.

Relacionado a lo anterior, también existen días en los que no hay atención en el centro y por lo tanto no se cuenta con información de la cantidad de inasistencias en ninguno de los bloques horarios. En estos casos no se imputan datos, ya que esto involucra crear artificialmente muchos valores (10 en el caso de los días de semana y 4 en el caso de días sábados), produciendo una gran cantidad de datos y distorsionando el comportamiento natural de la serie.

De acuerdo a lo señalado anteriormente, la cantidad de datos en cada día es estándar, de 10 datos en días de semana, correspondientes a la información de la cantidad de inasistencias en los bloques horarios desde las 9:00 hasta las 19:00hrs, y de 4 datos en días sábados, correspondientes a la información de los bloques horarios desde las 9:00 hasta las 13:00hrs.

En la siguiente tabla se presenta la cantidad de datos presentes en cada semana del año 2021 y 2022.

Tabla 4.7: Cantidad de bloques en cada semana.

Año	Semana	Cantidad de bloques (días de semana)	Cantidad de bloques (día sábado)
2021	44	40	4
	45	50	4
	46	50	4
	47	50	4
	48	50	4
	49	40	4
	50	50	4
	51	40	-
	52	40	-
	2022	1	50
2		50	4
3		40	4
4		50	4
5		50	4
6		50	4
7		50	4
8		50	4
9		50	4
10		50	4
11		50	4
12		50	4
13		50	4
14		50	4
15	40	-	
16	50	4	
17	50	4	
18	50	4	
19	50	4	
20	50	-	
21	50	4	
22	50	4	
23	50	4	
24	50	4	

Como se observa en la Tabla 4.7, la cantidad de bloques no es la misma en todas las semanas, lo cual ocurre porque hay semanas con días en los que el centro médico no atiende, lo que produce una cantidad de bloques menor a 50 para el caso de días de semana, y la no existencia de información para el caso de los días sábados.

Estacionalidad

La estacionalidad es un hiperparámetro importante de determinar tanto para el modelo ETS como para SARIMA. En este caso, la serie temporal tiene una componente de estacionalidad natural de largo 10, ya que al transcurrir 10 bloques horarios comienza un nuevo día, en el caso de los días de semana. En el caso de los días sábados la estacionalidad es de largo 4. Con esto, los modelos ETS y SARIMA consideran una serie de tiempo que contempla estacionalidad.

Dado que los días de semana y los días sábados poseen largos diferentes en su componente

de estacionalidad, se implementan modelos diferenciados, donde un modelo pronostica para los días de semana, mientras que otro pronostica para los días sábados, los que en su conjunto generan la predicción para la totalidad de la semana. Con esto, los modelos ETS y SARIMA reciben un hiperparámetro de valor 10 o 4 indicando el largo de la estacionalidad, según sea la predicción para días de semana o para un día sábado.

Así mismo, se implementan dos modelos RNN diferenciados, uno para pronosticar los días de semana y otro para pronosticar los días sábados, donde, en los días de semana, se define un largo de 10 para el vector de entrada, y, en los días sábados, un largo de 4, ya que, de esta forma, cada registro de entrada contiene la información de la cantidad de inasistencias presentadas en los 10 o 4 bloques horarios de atención, según sea un día de semana o un día sábado, representado así la estacionalidad de la serie.

Estabilización de varianza

Como se mencionó en la Sección 4.3.2 la serie de tiempo posee una alta variabilidad por lo que con objetivo de reducir la varianza se aplica una transformación a los datos de entrenamiento.

En el pronóstico de series de tiempo, estabilizar la varianza se puede realizar de diversas maneras, donde es común aplicar un escalamiento o una transformación logarítmica, donde este último mecanismo es bastante directo y funcional. Sin embargo, la transformación logarítmica requiere que todos los valores de la serie sean estrictamente positivos, dada la indefinición del logaritmo para valores negativos y ceros. En este caso, este requisito es abordado mediante una corrección simple, ya que los valores de la serie temporal son todos positivos o iguales a 0, por lo que sumando un valor de 1 a todos los valores de la serie es factible aplicar la transformación logarítmica.

Para el caso del entrenamiento de SARIMA se aplica la corrección sumando un valor de 1 a los valores de la serie temporal, mientras que para el entrenamiento de ETS se suma un valor de 1.5, ya que para el entrenamiento de ciertas configuraciones de hiperparámetros de ETS es necesario contar con valores estrictamente positivos, donde, al sumar un valor de 1.5, no se obtienen valores iguales a 0 al aplicar el logaritmo. Por otro lado, para el caso del entrenamiento de RNN se utiliza un método de escalamiento Min-Max Scaler, ya que este modelo no posee restricciones con respecto a sus valores de entrada.

Post procesamiento

Para obtener el pronóstico de cantidad de inasistencias para cada fecha y bloque horario es necesario procesar la salida entregada por los modelos ETS y SARIMA, debido a las transformaciones aplicadas a los datos de entrada en los algoritmos.

De esta forma, la salida pronosticada por los modelos es post procesada de la siguiente manera:

1. Se invierte la transformación logarítmica aplicando una transformación exponencial en el caso de los modelos SARIMA y ETS. Se invierte el escalamiento en el caso del modelo RNN.
2. Se resta un valor de 1 a la serie pronosticada por SARIMA, y un valor de 1.5 a la serie

pronosticada por ETS.

3. Se redondea el valor pronosticado al entero más cercano.
4. Se fija un valor de 0 en los valores pronosticados negativos.

Luego de aplicar el post procesamiento anterior, se evalúa el desempeño de los modelos utilizando la serie de tiempo pronosticada y la serie de tiempo real observada.

4.5. Evaluación

La evaluación de los diferentes modelos construidos se lleva a cabo en dos etapas. En primer lugar, se evalúa el desempeño en la tarea de clasificación para los modelos diseñados bajo el primer enfoque y el desempeño en la tarea de pronóstico para los modelos diseñados bajo el segundo enfoque. En segundo lugar, se seleccionan los modelos con mejor desempeño de ambos enfoques y sobre estos se evalúa su desempeño como herramientas para definir bloques horarios en los cuales ofertar sobrecupos para reservas de citas médicas.

4.5.1. Evaluación de modelos de clasificación

Como se menciona en la Sección 3.2.9, el desempeño de modelos en tareas de clasificación es típicamente evaluado utilizando medidas como *Precision*, *Recall*, *Accuracy* y *F1*. Sin embargo, la evaluación en base a las primeras cuatro medidas tiene ciertas desventajas, ya que estas dependen de un umbral de clasificación para definir la predicción de una observación en una clase correspondiente, en base a la probabilidad estimada por los modelos, donde el umbral de clasificación usualmente corresponde a un valor por defecto de 0.5, el cual puede llevar a una interpretación errónea de los resultados de acuerdo a las probabilidades predichas.

Por otro lado, los valores de ROC-AUC y Precision-Recall-AUC son medidas que resumen ambas curvas y que permiten evaluar y comparar el desempeño de diferentes modelos de una manera más robusta. Aun así, cabe mencionar que la curva ROC y su valor AUC pueden entregar una visión excesivamente optimista del desempeño cuando existe un desbalance considerable de clases, y es por esto que también se considera la curva de Precision-Recall sobre la clase de interés (en este caso inasistencia) y su AUC respectivo para la evaluación. Esto último se debe a que el uso de esta medida es recomendada cuando existen escenarios de clasificación desbalanceados [29], puesto que los diferentes valores de *Precision* y *Recall* (de acuerdo al umbral de clasificación) sobre la clase de interés permiten evaluar el desempeño de un clasificador en la predicción de la clase minoritaria [30].

Conforme a lo anterior, se utilizan las medidas de ROC-AUC y Precision-Recall-AUC para evaluar el desempeño de los clasificadores y seleccionar aquellos que tienen el mejor desempeño en predecir inasistencias sobre pacientes nuevos y recurrentes en el conjunto de proyección.

A continuación, se presenta una tabla con los valores de ROC-AUC y Precision-Recall-AUC obtenidos para la predicción en los conjuntos de testing y proyección, para cada configuración de modelos implementada.

Tabla 4.8: ROC-AUC y Precision-Recall-AUC (P-R-AUC) obtenidos en predicción de inasistencias sobre pacientes nuevos.

Modelo	Variables seleccionadas	Método de Balance	ROC-AUC Test	ROC-AUC Proyección	P-R-AUC Test	P-R-AUC Proyección
Logit	Todas	SMOTE	0.695	0.577	0.475	0.294
AdaBoost	Todas	SMOTE	0.698	0.576	0.476	0.293
Logit	G-LASSO	SMOTE	0.700	0.574	0.479	0.293
XGBoost	Todas	SMOTE	0.707	0.574	0.498	0.292
Logit	G-LASSO	Ninguno	0.699	0.573	0.479	0.291
Logit	Todas	Ninguno	0.696	0.572	0.476	0.289
AdaBoost	Todas	Ninguno	0.698	0.566	0.479	0.287
AdaBoost	G-LASSO	SMOTE	0.700	0.567	0.478	0.287
XGBoost	G-LASSO	Ninguno	0.710	0.562	0.505	0.286
XGBoost	G-LASSO	SMOTE	0.710	0.566	0.505	0.286
AdaBoost	G-LASSO	Ninguno	0.700	0.564	0.481	0.285
XGBoost	Todas	Ninguno	0.708	0.563	0.502	0.284
RF	G-LASSO	Ninguno	0.708	0.567	0.504	0.283
Bagging	G-LASSO	SMOTE	0.710	0.566	0.508	0.282
RF	G-LASSO	SMOTE	0.708	0.568	0.505	0.281
Bagging	G-LASSO	Ninguno	0.706	0.564	0.504	0.281
RF	Todas	SMOTE	0.707	0.563	0.500	0.280
RF	Todas	Ninguno	0.706	0.561	0.503	0.280
MLP	Todas	Ninguno	0.699	0.563	0.467	0.279
MLP	G-LASSO	SMOTE	0.683	0.560	0.466	0.278
Bagging	Todas	Ninguno	0.703	0.560	0.502	0.278
Bagging	Todas	SMOTE	0.705	0.561	0.502	0.277
Naive Bayes	G-LASSO	Ninguno	0.655	0.531	0.403	0.274
MLP	G-LASSO	Ninguno	0.709	0.558	0.489	0.274
Naive Bayes	G-LASSO	SMOTE	0.656	0.532	0.404	0.274
Naive Bayes	Todas	SMOTE	0.636	0.531	0.391	0.273
Naive Bayes	Todas	Ninguno	0.635	0.525	0.392	0.271
MLP	Todas	SMOTE	0.682	0.538	0.463	0.262
KNN	G-LASSO	Ninguno	0.676	0.535	0.457	0.260
KNN	Todas	SMOTE	0.668	0.536	0.458	0.260
KNN	G-LASSO	SMOTE	0.666	0.532	0.446	0.259
KNN	Todas	Ninguno	0.677	0.521	0.468	0.252

Tabla 4.9: ROC-AUC y Precision-Recall-AUC (P-R-AUC) obtenidos en predicción de inasistencias sobre pacientes recurrentes.

Modelo	Variables se-leccionadas	Método de Balance	ROC-AUC Test	ROC-AUC Proyección	P-R-AUC Test	P-R-AUC Proyección
XGBoost	Todas	Ninguno	0.701	0.651	0.399	0.288
AdaBoost	Todas	Ninguno	0.679	0.648	0.374	0.284
RF	Todas	Ninguno	0.670	0.661	0.390	0.282
AdaBoost	G-LASSO	Ninguno	0.672	0.637	0.377	0.277
MLP	G-LASSO	RUS + SMOTE	0.676	0.627	0.377	0.276
Logit	Todas	Ninguno	0.658	0.640	0.367	0.275
XGBoost	Todas	RUS + SMOTE	0.686	0.651	0.386	0.273
XGBoost	G-LASSO	Ninguno	0.693	0.641	0.393	0.273
Logit	G-LASSO	Ninguno	0.658	0.642	0.365	0.271
Logit	G-LASSO	RUS + SMOTE	0.659	0.641	0.367	0.271
Logit	Todas	RUS + SMOTE	0.655	0.632	0.359	0.269
Bagging	Todas	Ninguno	0.664	0.647	0.376	0.268
Naive Bayes	G-LASSO	Ninguno	0.661	0.621	0.342	0.266
Naive Bayes	G-LASSO	RUS + SMOTE	0.661	0.621	0.349	0.265
RF	Todas	RUS + SMOTE	0.676	0.648	0.392	0.265
AdaBoost	G-LASSO	RUS + SMOTE	0.658	0.628	0.362	0.265
Bagging	Todas	RUS + SMOTE	0.674	0.653	0.376	0.260
KNN	G-LASSO	Ninguno	0.636	0.595	0.336	0.260
RF	G-LASSO	Ninguno	0.677	0.632	0.377	0.258
Bagging	G-LASSO	Ninguno	0.679	0.624	0.383	0.253
KNN	Todas	Ninguno	0.627	0.609	0.335	0.252
AdaBoost	Todas	RUS + SMOTE	0.657	0.622	0.344	0.251
Bagging	G-LASSO	RUS + SMOTE	0.677	0.629	0.378	0.248
XGBoost	G-LASSO	RUS + SMOTE	0.657	0.603	0.346	0.247
MLP	Todas	Ninguno	0.631	0.604	0.331	0.243
RF	G-LASSO	RUS + SMOTE	0.682	0.624	0.378	0.242
KNN	Todas	RUS + SMOTE	0.606	0.577	0.305	0.234
KNN	G-LASSO	RUS + SMOTE	0.615	0.577	0.302	0.227
MLP	Todas	RUS + SMOTE	0.648	0.608	0.347	0.224
MLP	G-LASSO	Ninguno	0.652	0.629	0.261	0.220
Naive Bayes	Todas	Ninguno	0.604	0.573	0.273	0.214
Naive Bayes	Todas	RUS + SMOTE	0.592	0.564	0.270	0.209

En las Tablas 4.8 y 4.9 se observa que, para el caso de la predicción sobre pacientes nuevos, el mejor desempeño lo obtiene el modelo Logit utilizando todas las variables y con un método de balance de clases, donde esta estrategia alcanza el mayor valor de P-R-AUC en el conjunto de proyección. Por otra parte, para el caso de predicción sobre pacientes recurrentes, el mejor desempeño lo obtiene el modelo XGBoost utilizando todas las variables y sin ningún método de balance de clases.

Además, se observa que en todos los modelos el valor de ROC-AUC es considerablemente más alto que el de P-R-AUC, tanto en el conjunto de testing como en el conjunto de proyección, lo que muestra que el desempeño en la predicción de los clasificadores no es el mismo para ambas clases. En particular, el desempeño es bastante deficiente para la predicción de la clase de interés y minoritaria (inasistencia). Por otro lado, también se observa que para todos los modelos el valor de ROC-AUC y P-R-AUC es más grande en el conjunto de testing que en el de proyección, por lo que los modelos no tienen una buena capacidad de generalización, disminuyendo su desempeño de predicción en un conjunto de datos nuevos y más recientes.

A continuación, se presentan las medidas de *Precision*, *Recall*, *Accuracy* y *F1* obtenidas

por los mejores modelos en la predicción de inasistencias para pacientes nuevos y recurrentes, bajo un umbral de clasificación por defecto, es decir, prediciendo como inasistencias aquellas citas con una probabilidad de inasistencia estimada mayor o igual a 0.5.

Tabla 4.10: Precision, Recall, Accuracy y F1 obtenidos en predicción sobre pacientes nuevos, modelo Logit.

		Conjunto Testing				
Clase		Precision	Recall	F1	Accuracy	N
	Asistencia	0.76	0.89	0.82		4,886
	Inasistencia	0.54	0.31	0.39		2,017
					0.72	6,903
		Conjunto Proyección				
Clase		Precision	Recall	F1	Accuracy	N
	Asistencia	0.76	0.96	0.85		5,409
	Inasistencia	0.35	0.07	0.12		1,721
					0.74	7,130

Tabla 4.11: Precision, Recall, Accuracy y F1 obtenidos en predicción sobre pacientes recurrentes, modelo XGBoost.

		Conjunto Testing				
Clase		Precision	Recall	F1	Accuracy	N
	Asistencia	0.85	0.98	0.91		2,679
	Inasistencia	0.58	0.14	0.23		551
					0.84	3,230
		Conjunto Proyección				
Clase		Precision	Recall	F1	Accuracy	N
	Asistencia	0.87	0.99	0.92		3,049
	Inasistencia	0.53	0.08	0.15		512
					0.86	3,561

En las Tablas 4.10 y 4.11 se observa que el desempeño en la clasificación de la clase de asistencia es considerablemente mejor que en la clasificación de la clase de inasistencia, donde se obtienen valores mayores en la *Precision*, *Recall* y *F1* para la clasificación de asistencia, tanto en los conjuntos de testing y proyección. Por otra parte, se obtienen mejores resultados en la predicción de la inasistencia para pacientes recurrentes, donde se obtiene un *F1* mayor que en los pacientes nuevos, dentro del conjunto de proyección. Como se mencionó anteriormente, la predicción para la clase de interés y minoritaria no tiene un buen desempeño en el conjunto de proyección, donde, bajo un umbral de clasificación por defecto, solo se logra identificar un 7% de las inasistencias de pacientes nuevos y un 8% de pacientes recurrentes, mientras que, al predecir inasistencias, se acierta en solo un 35% de los casos en pacientes nuevos y en un 53% de los casos en pacientes recurrentes.

Con respecto a la relevancia de las variables en la predicción, en la Sección Anexos B se presenta la puntuación de importancia asignada por los modelos Logit y XGBoost a

las variables de entrada, en función de su utilidad para predecir la variable objetivo. En el modelo Logit, esta puntuación se mide en base a los coeficientes estimados para cada variable y sus respectivas odds, las que corresponden al factor de incremento en la probabilidad de pertenecer a la clase inasistencia en caso de tener un valor específico en una variable categórica, o en caso de aumentar en uno el valor de una variable numérica. Para el caso del modelo XGBoost, la puntuación se mide en base a la ganancia de información promedio generada por cada variable. Específicamente, en base a la importancia de atributos entregada por el modelo Logit, se observa que las variables tasa de inasistencia médico y canal de creación son las que afectan con mayor magnitud la probabilidad de pertenecer a la clase inasistencia, donde las citas agendadas con médicos con una mayor tasa de inasistencia previa tienen mayor probabilidad de resultar en ausencias, mientras que, para el caso del canal de creación, las citas agendadas mediante recepción tienen la menor probabilidad de resultar en inasistencias, seguidas por las citas agendadas por call center y finalmente las citas agendadas por el portal web. Por otro lado, en base a la importancia de atributos entregada por el modelo XGBoost, se obtiene que las cinco variables que contribuyen de mayor manera en la predicción son: tasa de inasistencia paciente, canal de creación, demanda paciente, tasa de inasistencia médico y edad paciente, donde las dos primeras características son aquellas que contribuyen de mayor manera. De esta manera, se observa que ambos modelos consideran la variable canal de creación como un atributo importante para la predicción, mientras que, la variable tasa de inasistencia paciente contribuye notablemente en la predicción cuando es posible contar con información histórica sobre pacientes que han agendado citas previamente en el centro médico.

4.5.2. Evaluación de modelos de pronóstico

La evaluación del desempeño de los modelos de pronóstico se realiza en base a las medidas de Mean Absolute Scaled Error (MASE), Mean Absolute Error (MAE) y Root Mean Square Error (RMSE). Estas medidas se utilizan debido a que la escala de los datos es constante, donde no es necesario utilizar medidas de error porcentuales para comparar pronósticos sobre datos con magnitudes diferentes.

Específicamente, MASE se utiliza debido a su directa interpretación, donde aquellos valores de MASE superiores a uno indican que los pronósticos son peores, en promedio, que los pronósticos de un solo paso adelante obtenidos utilizando un método ingenuo. Asimismo, MASE es altamente recomendada cuando existe una cantidad considerable de valores nulos en los datos [26]. En esta misma línea, MAE también tiene una interpretación directa, ya que esta cuantifica qué tan distante se encuentran, en promedio, los pronósticos de los valores reales, midiendo esta distancia en la misma escala de los datos. Por otra parte, RMSE se utiliza para cuantificar el nivel de variabilidad en los errores cometidos en el pronóstico de la serie temporal, para así poder medir la calidad del ajuste del pronóstico respecto a los valores reales.

Las medidas anteriores se calculan para cada semana, utilizando los valores pronosticados por ETS, SARIMA y RNN para los días de semana y días sábados correspondientes. Luego, se calcula un promedio entre las medidas obtenidas a lo largo de todas las semanas.

A continuación, se presentan los resultados obtenidos en el pronóstico de la cantidad de inasistencias para cada fecha y bloque horario pertenecientes a las 24 primeras semanas de

2022.

Tabla 4.12: MASE, MAE y RMSE por semana para los pronósticos obtenidos mediante modelo SARIMA.

Semana	Días de semana			Sábado		
	MASE	MAE	RMSE	MASE	MAE	RMSE
2022-01	0.872	1.200	1.649	-	-	-
2022-02	0.973	1.280	1.778	-	-	-
2022-03	1.449	1.800	2.449	0.950	1.750	1.936
2022-04	1.451	1.920	2.441	1.086	2.000	2.550
2022-05	1.414	2.021	2.471	0.576	1.000	1.225
2022-06	0.757	1.240	1.600	0.475	0.750	1.118
2022-07	1.149	1.880	2.392	1.644	2.250	2.291
2022-08	0.825	1.360	1.855	1.140	1.500	2.550
2022-09	0.779	1.300	1.619	1.446	1.750	1.803
2022-10	0.669	1.080	1.428	1.462	2.000	2.121
2022-11	0.744	1.160	1.456	1.859	2.250	2.291
2022-12	0.586	0.920	1.166	0.432	0.500	1.000
2022-13	0.792	1.208	1.720	0.704	1.000	1.225
2022-14	0.590	0.837	1.187	0.620	0.750	0.866
2022-15	0.467	0.615	0.816	-	-	-
2022-16	0.946	1.102	1.552	1.900	2.000	2.236
2022-17	0.602	0.633	0.845	-	-	-
2022-18	0.773	0.857	1.212	0.905	1.000	1.414
2022-19	0.715	0.700	0.970	1.118	1.000	1.414
2022-20	1.085	1.080	1.371	-	-	-
2022-21	0.809	0.796	1.097	0.760	1.000	1.225
2022-22	0.909	0.960	1.217	-	-	-
2022-23	0.720	0.700	0.927	0.880	1.250	1.500
2022-24	1.130	1.180	1.581	0.880	1.250	1.323
Promedio	0.884	1.160	1.533	1.047	1.389	1.672
Desv. Estándar	0.273	0.405	0.505	0.453	0.557	0.553

Tabla 4.13: MASE, MAE y RMSE por semana para los pronósticos obtenidos mediante modelo ETS.

Semana	Días de semana			Sábado		
	MASE	MAE	RMSE	MASE	MAE	RMSE
2022-01	0.989	1.360	1.811	-	-	-
2022-02	0.912	1.200	1.649	-	-	-
2022-03	1.188	1.475	2.019	0.548	1.000	1.414
2022-04	1.451	1.920	2.425	0.958	1.750	2.291
2022-05	0.863	1.234	1.502	0.295	0.500	0.707
2022-06	0.660	1.080	1.371	0.575	1.000	1.225
2022-07	0.978	1.600	2.049	1.278	2.000	2.121
2022-08	0.934	1.540	2.093	1.643	2.000	2.646
2022-09	0.659	1.100	1.349	1.586	2.000	2.345
2022-10	0.681	1.100	1.463	1.855	2.500	2.646
2022-11	0.732	1.140	1.435	1.977	2.750	2.872
2022-12	0.599	0.940	1.225	1.150	1.250	1.323
2022-13	0.805	1.229	1.652	0.595	0.750	1.500
2022-14	0.561	0.796	1.040	0.192	0.250	0.500
2022-15	0.467	0.615	0.816	-	-	-
2022-16	0.894	1.041	1.436	1.380	1.500	1.732
2022-17	0.602	0.633	0.845	-	-	-
2022-18	0.773	0.857	1.212	1.438	1.500	1.871
2022-19	0.776	0.760	1.039	0.958	1.000	1.414
2022-20	1.145	1.140	1.449	-	-	-
2022-21	0.706	0.694	1.030	0.767	1.000	1.225
2022-22	0.871	0.920	1.281	-	-	-
2022-23	0.720	0.700	0.949	0.397	0.500	1.000
2022-24	1.034	1.080	1.510	0.348	0.500	0.707
Promedio	0.833	1.090	1.444	0.997	1.319	1.641
Desv. Estándar	0.225	0.328	0.414	0.565	0.727	0.718

Tabla 4.14: MASE, MAE y RMSE por semana para los pronósticos obtenidos mediante modelo RNN.

Semana	Días de semana			Sábado		
	MASE	MAE	RMSE	MASE	MAE	RMSE
2022-01	0.903	1.148	1.571	-	-	-
2022-02	1.011	1.240	1.690	-	-	-
2022-03	1.002	1.355	1.789	0.388	0.750	0.866
2022-04	1.069	1.568	2.124	0.877	1.500	1.871
2022-05	0.759	1.296	1.584	0.439	0.750	1.118
2022-06	0.777	1.300	1.526	0.298	0.500	1.000
2022-07	0.887	1.400	1.854	0.969	1.500	1.732
2022-08	0.833	1.344	1.812	1.329	1.800	2.608
2022-09	0.760	1.168	1.408	1.938	2.500	2.646
2022-10	0.776	1.256	1.512	0.901	1.250	1.323
2022-11	0.664	1.068	1.322	0.596	0.750	1.118
2022-12	0.668	0.980	1.238	0.567	0.750	0.866
2022-13	0.916	1.252	1.601	1.107	1.250	1.323
2022-14	0.802	0.992	1.245	0.419	0.500	0.707
2022-15	0.656	0.705	0.943	-	-	-
2022-16	0.927	0.960	1.342	1.192	1.500	1.574
2022-17	0.561	0.612	0.832	-	-	-
2022-18	0.908	0.840	1.200	1.356	1.400	2.017
2022-19	0.665	0.672	0.885	1.140	1.250	1.461
2022-20	1.089	1.084	1.415	-	-	-
2022-21	0.796	0.784	1.011	1.124	1.450	1.717
2022-22	0.845	0.900	1.265	-	-	-
2022-23	0.732	0.736	0.992	0.581	0.750	0.866
2022-24	1.086	1.184	1.538	0.397	0.500	0.707
Promedio	0.837	1.077	1.404	0.868	1.147	1.418
Desv. Estándar	0.146	0.259	0.331	0.442	0.535	0.594

Como se observa en las Tablas 4.12, 4.13 y 4.14, no se realizan predicciones para los días sábados en las semanas 1, 2, 15, 20 y 22. En el caso de las semanas 15 y 20, esto se debe a que en esas semanas no hubo atención del centro médico. Además, existen semanas en que no se cuenta con la cantidad de semanas previas necesarias para el entrenamiento, donde, en estos casos, se desplaza la ventana de entrenamiento una semana hacia atrás, para contar con las semanas de entrenamiento necesarias, sin embargo, en las semanas 1, 2 y 22, aún desplazando la ventana de entrenamiento una semana hacia atrás, no se consigue la cantidad necesaria de semanas de entrenamiento, por lo que para estas semanas no se realizan predicciones.

Respecto al desempeño de los modelos implementados utilizando SARIMA, en la Tabla 4.12 se observa que los pronósticos tienen un MASE, MAE y RMSE promedio de 0.884, 1.160 y 1.533 respectivamente, para el caso del pronóstico en días de semana. Por otra parte, los promedios son de 1.047, 1.389 y 1.672 para el pronóstico de días sábados. Con esto, el pronóstico para los días de semana es mejor que el pronóstico para los días sábados, donde las medidas de error promediadas entre semanas son más altas en el caso de los días sábados.

En general, para los días de semana, se observa que el pronóstico de SARIMA tiene un MASE menor a 1, mostrando un mejor desempeño que un método de pronóstico ingenuo de un paso adelante, mientras que, para los días sábados, esto no ocurre, con un promedio de MASE mayor a 1.

Por otro lado, el valor de MAE promedio entre semanas para los días de semana y los días sábados es cercano a 1, indicando que, en promedio, el error del pronóstico es de aproximadamente 1 inasistencia.

Finalmente, el valor de RMSE promedio entre semanas es mayor en el caso de los días sábados, lo cual indica que el ajuste del modelo SARIMA es mejor en los días de semana que en el día sábado, teniendo una menor variabilidad en los errores de pronóstico en los días de semana.

Respecto al desempeño de los modelos implementados utilizando ETS, en la Tabla 4.13 se observa que los pronósticos tienen un MASE, MAE y RMSE promedio de 0.833, 1.090 y 1.444 respectivamente, para el caso del pronóstico en días de semana. Por otra parte, los promedios son de 0.997, 1.319 y 1.641 en el pronóstico de días sábados. Al igual que en el modelo SARIMA, el pronóstico para los días de semana es mejor que el pronóstico para los días sábados, donde las medidas de error promediadas entre semanas son más grandes en el caso de los días sábados. Sin embargo, en este caso la brecha entre ambos pronósticos es ligeramente menor que la obtenida con el modelo SARIMA.

En contraste al caso de SARIMA, para el modelo ETS se observa que el pronóstico tiene un MASE menor a 1, tanto para los días de semana como para los días sábados, teniendo así un desempeño mejor que un método de pronóstico ingenuo de un paso adelante.

Por otro lado, el valor de MAE promedio entre semanas para los días de semana y los días sábados es cercano a 1, indicando que, en promedio, el error del pronóstico es de aproximadamente 1 inasistencia, al igual que en el caso de SARIMA.

Finalmente, el valor de RMSE promedio entre semanas es mayor en el caso de los días sábados, obteniendo un mejor ajuste del modelo ETS en los días de semana.

Por último, respecto al desempeño de los modelos implementados utilizando RNN, en la Tabla 4.14 se observa que los pronósticos tienen un MASE, MAE y RMSE promedio de 0.837, 1.077 y 1.404 respectivamente, para el caso del pronóstico en días de semana. Por otra parte, los promedios son de 0.868, 1.147 y 1.418 en el pronóstico de días sábados. Se mantiene la misma tendencia que en los modelos anteriores, donde el pronóstico para los días de semana es mejor que el pronóstico para los días sábados. Sin embargo, la brecha entre ambos pronósticos es menor que en el caso de SARIMA y ETS.

El pronóstico de RNN obtiene un MASE menor a 1, tanto para los días de semana como para los días sábados, al igual que el pronóstico de ETS. Así, el desempeño es mejor que el obtenido mediante un método de pronóstico ingenuo.

Por otro lado, el valor de MAE promedio entre semanas para los días de semana y los días sábados también es cercano a 1, fallando en aproximadamente 1 inasistencia, al igual que en el caso de ETS y SARIMA.

Finalmente, en este caso, el valor de RMSE promedio es ligeramente menor en los días de semanas que en los días sábados, obteniendo un ajuste similar en ambos casos.

Comparando los resultados de los tres modelos, se observa que RNN obtiene un desempeño ligeramente mejor que ETS y SARIMA, donde las medidas de error promedio entre semanas son menores para en el modelo RNN tanto en días de semana como en días sábados, donde la mejora en el rendimiento es más notoria en el pronóstico de los días sábados. Sin embargo, el desempeño en la predicción en los días de semana es muy similar en el modelo ETS y en el modelo RNN.

Tomando en consideración el hecho de que el rendimiento en la predicción para los bloques horarios de días de semana es muy similar en los modelos ETS y RNN, se selecciona el modelo ETS para la definición de sobrecupos en base al pronóstico de la cantidad de inasistencias por fecha y bloque horario, ya que, si bien el desempeño del modelo RNN es ligeramente mejor en los días de semana y días sábado, este es un modelo mucho más complejo de implementar y que requiere una mayor calibración de hiperparámetros, mientras que, el modelo ETS es un modelo más sencillo y directo de implementar, el cual obtiene un desempeño en el pronóstico similar al del modelo RNN en los días de semana, donde los días de semana representan la mayor cantidad de datos.

4.5.3. Evaluación en la definición de sobrecupos

Para determinar en qué fechas y bloques horarios se presenta la oportunidad de ofertar sobrecupos resulta necesario contar con una directriz sobre la ocurrencia de inasistencias. De esta forma, en base a las inasistencias predichas por los modelos seleccionados en la Sección 4.5.1 y 4.5.2, se definen tramos para ofertar sobrecupos en cada fecha y bloque horario. Esto se realiza para todos los bloques horarios y días de atención de las primeras 24 semanas de 2022, lo que corresponde a un total de 3,518 bloques horarios.

Los modelos de clasificación binaria seleccionados (Logit y XGBoost) entregan predicciones sobre la probabilidad de inasistencia para una cita en particular, donde, posteriormente, se define la predicción del estado final de una cita en base a un umbral de clasificación. En este caso, se utiliza el valor de umbral que maximiza la medida $F1$ obtenida para la clase inasistencia, calculando esta medida en el conjunto de testing, tanto en el modelo seleccionado para predecir citas de pacientes nuevos (Logit) como para predecir citas de pacientes recurrentes (XGBoost).

Seleccionado el umbral de corte, se predicen como inasistencias aquellas citas donde la probabilidad estimada de inasistencia es igual o mayor que el umbral y, de esta forma, se calcula la cantidad de inasistencias predichas para un médico en una fecha y bloque horario, reestructurando los datos de manera análoga a los que se utilizan para entrenar los modelos de pronóstico de la serie temporal, con la diferencia en que, en este caso, se calcula la cantidad de inasistencias predichas por fecha, bloque horario y también por médico. Luego, utilizando el número de inasistencias predichas, es posible definir sobrecupos de reserva en

aquellos médicos, bloques horarios y fechas donde la predicción de inasistencias es mayor a cero, ya que, en estos casos, se predice que al menos un paciente no asistirá a su cita en la fecha y bloque horario correspondientes.

Por otro lado, el modelo de pronóstico seleccionado (ETS) entrega directamente una predicción sobre la cantidad de inasistencias en una fecha y bloque horario, sin embargo, esta predicción no conlleva información sobre la distribución de esta cantidad de inasistencias sobre los diferentes médicos que atienden en la fecha y bloque horario correspondientes. Lo anterior se debe a que la serie de tiempo utilizada para entrenar los modelos tiene la información de la cantidad de inasistencias por fecha y bloque horario para todas las citas ocurridas en el centro, donde no se entrena diferenciando entre series temporales particulares para cada médico, ya que, es necesario contar con una serie temporal con una cantidad constante de información a lo largo del tiempo, disponiendo de la información de todos los bloques horarios de cada día, siendo esto imposible de conseguir al generar una serie temporal específica para cada médico, puesto que, la mayoría de ellos no atienden durante todos los bloques horarios de un día, ni durante todos los días de una semana.

Dado lo anterior, es necesario distribuir la cantidad de inasistencias pronosticadas por el modelo ETS entre los diferentes médicos. Para esto, se utiliza información disponibilizada por la institución sobre la tasa de inasistencia específica para cada bloque y médico, calculada en base a las citas agendadas en los pasados 20 días de atención previos a la fecha del bloque horario en el que se pronostica inasistencias. Esta tasa está disponible para todos los médicos y se calcula tomando la cantidad de inasistencias sobre el total de citas que tiene el médico en cada bloque horario específico. Con esto, se cuenta con información sobre qué médicos tienen una mayor tasa de inasistencia en cada bloque horario.

Entonces, utilizando la tasa de inasistencia descrita anteriormente, se distribuye la cantidad total de inasistencias predichas por el modelo de pronóstico, repartiendo esta cantidad sobre aquellos médicos que tienen la mayor tasa de inasistencia, donde se asigna un máximo de una inasistencia por bloque horario para cada médico.

En la siguiente figura se presenta la lógica utilizada para la repartición.

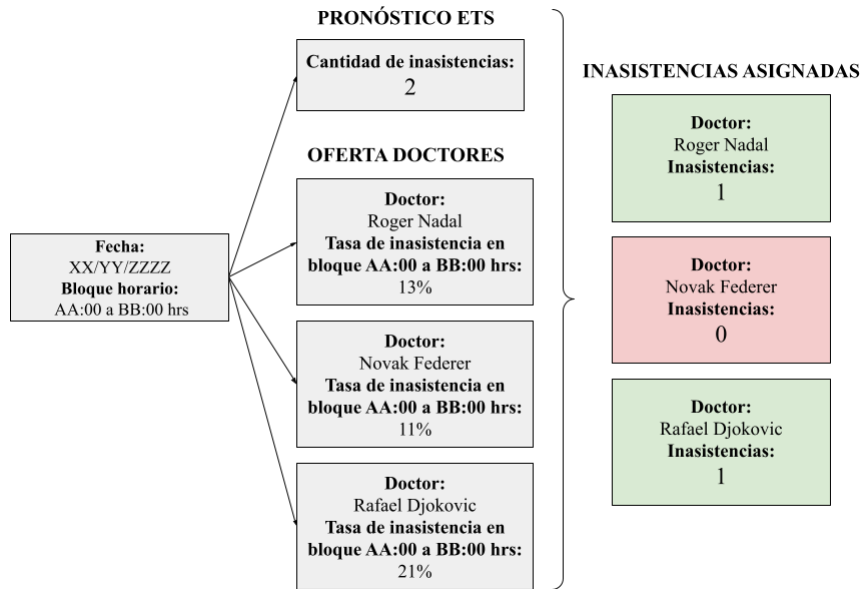


Figura 4.13: Estrategia para repartición de inasistencias pronosticadas por modelo ETS.

Además de la estrategia anterior, se considera también una segunda modalidad para la repartición de inasistencias, la cual consiste en la combinación de los modelos de ambos enfoques (Logit+XGBoost y ETS), distribuyendo la cantidad de inasistencias estimada por los modelos de pronóstico en base a la máxima probabilidad de inasistencia predicha por los modelos de clasificación, con lo cual se selecciona la máxima probabilidad de inasistencia para cada médico entre las citas pertenecientes a la fecha y bloque horario correspondiente. De esta forma, la estrategia de repartición es similar a la presentada en la Figura 4.13, con la diferencia en que se cuenta con la probabilidad máxima de inasistencia entre las citas de cada médico en reemplazo de la tasa de inasistencia.

De este modo, se evalúan tres estrategias para predecir inasistencias en bloques horarios y, por consiguiente, definir aquellos bloques horarios en los cuales ofertar sobrecupos, donde estas estrategias se construyen en base a las predicciones entregadas por los modelos diseñados bajo los enfoques de una clasificación binaria y el pronóstico de una serie temporal.

Para evaluar el desempeño de los modelos en la definición de sobrecupos se utiliza el bloque horario como unidad de análisis. Así, en cada bloque horario se observa si ocurre una inasistencia y si en ese mismo bloque las estrategias mencionadas anteriormente predicen inasistencias. De esta manera, se cuenta con la información observada sobre las inasistencias presentadas en cada bloque horario y con la información predicha sobre la ocurrencia de inasistencias en estos mismos bloques. Los sobrecupos son ofertados en aquellos bloques horarios donde se predicen inasistencias, por lo que es relevante evaluar la capacidad de las estrategias de repartición de identificar aquellos bloques con inasistencias (*Recall*) y de predecir correctamente la ocurrencia de inasistencias en los bloques señalados (*Precision*).

A continuación, se presentan las medidas de *Precision* y *Recall* obtenidas en cada estrategia de repartición.

Tabla 4.15: Resultados en la predicción de inasistencias en bloques horarios, en base a modelos de clasificación binaria (Logit + XGBoost).

		Precision	Recall	N
Inasistencia en bloque	No	0.58	0.89	1,974
	Si	0.56	0.17	1,544
				3,518 Total

Tabla 4.16: Resultados en la predicción de inasistencias en bloques horarios, en base a modelo de pronóstico (ETS) y repartición usando tasa de inasistencia de médicos.

		Precision	Recall	N
Inasistencia en bloque	No	0.58	0.56	1,974
	Si	0.46	0.47	1,544
				3,518 Total

Tabla 4.17: Resultados en la predicción de inasistencias en bloques horarios, en base a modelo de pronóstico (ETS) y repartición usando probabilidad de inasistencia de citas (Logit + XGBoost).

		Precision	Recall	N
Inasistencia en bloque	No	0.63	0.61	1,974
	Si	0.52	0.54	1,544
				3,518 Total

En la Tabla 4.15 se observa que, la predicción de inasistencias en los bloques horarios en base al modelo de clasificación binaria, es capaz de identificar un 17 % de las inasistencias, mientras que acierta en la predicción en un 56 % de los bloques en los que predice inasistencias.

Por otra parte, en la Tabla 4.16 se observa que, la predicción de inasistencias para bloques horarios en base al modelo de pronóstico y a una repartición utilizando la tasa de inasistencia de los médicos, es capaz de identificar un 47 % de las inasistencias, mientras que acierta en la predicción en un 46 % de los bloques en los que predice inasistencias.

Por último, en la Tabla 4.17 se observa que, la predicción de inasistencias para bloques horarios en base al modelo de pronóstico y a una repartición utilizando la probabilidad de inasistencia predicha por el modelo de clasificación, es capaz de identificar un 52 % de las inasistencias, mientras que acierta en la predicción en un 54 % de los bloques en los que predice inasistencias.

En base a los resultados obtenidos, se observa que el valor de *Precision* más alto es de un 56 %, obtenido por el modelo de clasificación binaria. Además, esta estrategia alcanza un

Recall de 17%. Por otra parte, la predicción en base al modelo de pronóstico acoplado al modelo de clasificación binaria alcanza una *Precision* de 52%, donde obtiene 4 puntos porcentuales menos que la *Precision* más alta obtenida, pero con un valor de *Recall* de 54%, el cual es considerablemente mayor que el obtenido por la estrategia que alcanza la *Precision* más alta. Conforme a lo anterior, la estrategia con mejor desempeño es la predicción en base al modelo de pronóstico acoplado con el modelo de clasificación.

Si bien es relevante analizar el desempeño de la predicción sobre todos los bloques horarios, esto no permite evaluar teóricamente el desempeño de las estrategias al definir sobrecupos de forma más cercana a la realidad, ya que, una cita es catalogada como sobrecupo solo en aquellos casos donde el médico se encuentra con la agenda colmada en un bloque horario correspondiente, recordando que la definición de un sobrecupo es programar más de un paciente para un mismo intervalo de tiempo y médico, y, para que esto ocurra, el médico debe contar con todos sus cupos del bloque horario agendados. En este caso, el médico tiene su agenda completa cuando tiene un mínimo de 4 citas agendadas en el bloque, en el caso de médicos que atienden con una duración de 15 min, y cuando tiene un mínimo de 3 citas agendadas en el bloque, en el caso de médicos que atienden con una duración de 20 min.

Dado esto, se repite el análisis anterior considerando sólo aquellos bloques horarios donde los médicos se encuentran con su agenda colmada, ya que, en estos casos, un sobrecupo sería teóricamente agendado y atendido. Para esto, es necesario diferenciar entre bloques horarios que contienen citas de 15 y 20 minutos, ya que, los médicos colman su agenda con una cantidad diferente de citas en cada caso. Del total de bloques, en 1,245 casos los médicos que atienden en 15 minutos poseen su agenda completa, mientras que en los médicos que atienden en 20 minutos esto ocurre en 145 casos.

A continuación, se presentan las medidas de *Precision* y *Recall* obtenidas en cada estrategia, calculando estas medidas en base a los bloques horarios en los cuales los médicos tienen la agenda completa.

Tabla 4.18: Resultados en la predicción de inasistencias sobre bloques horarios, en base a modelo de clasificación binaria (Logit + XGBoost). Bloques completos con citas de 15 minutos.

		Precision	Recall	N
Inasistencia en bloque	No	0.44	0.86	516
	Si	0.68	0.21	729
				1,245 Total

Tabla 4.19: Resultados en la predicción de inasistencias sobre bloques horarios, en base a modelo de pronóstico (ETS) y repartición usando tasa de inasistencia médicos. Bloques completos con citas de 15 minutos.

		Precision	Recall	N
Inasistencia en bloque	No	0.45	0.59	516
	Si	0.63	0.49	729
				1,245 Total

Tabla 4.20: Resultados en la predicción de inasistencias sobre bloques horarios, en base a modelo de pronóstico (ETS) y repartición usando probabilidad de inasistencia de citas (Logit + XGBoost). Bloques completos con citas de 15 minutos.

		Precision	Recall	N
Inasistencia en bloque	No	0.49	0.52	516
	Si	0.65	0.62	729
				1,245 Total

Tabla 4.21: Resultados en la predicción de inasistencias sobre bloques horarios, en base a modelo de clasificación binaria (Logit + XGBoost). Bloques completos con citas de 20 minutos.

		Precision	Recall	N
Inasistencia en bloque	No	0.46	0.83	66
	Si	0.56	0.18	79
				145 Total

Tabla 4.22: Resultados en la predicción de inasistencias sobre bloques horarios, en base a modelo de pronóstico (ETS) y repartición usando tasa de inasistencia médicos. Bloques completos con citas de 20 minutos.

		Precision	Recall	N
Inasistencia en bloque	No	0.45	0.45	66
	Si	0.54	0.54	79
				145 Total

Tabla 4.23: Resultados en la predicción de inasistencias sobre bloques horarios, en base a modelo de pronóstico (ETS) y repartición usando probabilidad de inasistencia de citas (Logit + XGBoost). Bloques completos con citas de 20 minutos.

		Precision	Recall	N
Inasistencia en bloque	No	0.52	0.48	66
	Si	0.59	0.62	79
				145 Total

En la Tablas 4.18, 4.19 y 4.20 se observa que el desempeño en la predicción de inasistencias en bloques horarios completos con citas de 15 minutos mejora con respecto a la evaluación considerando la totalidad de los bloques. En particular, se aprecia que en las tres estrategias aumenta el valor de la *Precision* y el *Recall*, con respecto a los valores obtenidos al evaluar utilizando todos los bloques. Específicamente, en bloques de citas de 15 minutos, el mejor desempeño lo obtiene la predicción mediante el modelo de pronóstico (ETS) con una repartición en base a la probabilidad de inasistencia predicha por el modelo de clasificación binaria (Logit+XGBoost), donde esta estrategia es capaz de identificar un 62 % de las inasistencias con un exactitud del 65 %.

Por otra parte, en las Tablas 4.21, 4.22 y 4.23 se observa que, el desempeño en la evaluación de la predicción de inasistencias en bloques horarios completos con citas de 20 minutos también mejora con respecto a la evaluación realizada considerando todos los bloques. Sin embargo, en este caso, la mejora en el desempeño es levemente menor que la obtenida en los bloques completos con citas de 15 minutos, donde la diferencia entre la *Precision* y el *Recall* entre ambos escenarios de evaluación es menor en el caso de los bloques con citas de 20 minutos.

Específicamente, en los bloques horarios completos con citas de 20 minutos, el mejor desempeño nuevamente lo obtiene la predicción mediante el modelo de pronóstico (ETS) con una repartición en base a la probabilidad de inasistencia predicha por el modelo de clasificación binaria (Logit+XGBoost), donde esta estrategia es capaz de identificar un 59 % de las inasistencias con una exactitud del 62 %.

Resulta razonable que la evaluación de la predicción de inasistencias en un escenario de bloques horarios completos obtenga un mejor resultado que la evaluación en la predicción considerando todos los bloques, ya que, al existir una mayor cantidad de citas en el bloque, la posibilidad de que un paciente no asista es más grande, considerando que la tasa de inasistencia general para las citas es de aproximadamente un 24 %, como se puede observar en la Tabla 4.3.

En base al análisis anterior, el mejor resultado se obtiene en base a la estrategia que utiliza el modelo de pronóstico y clasificación en conjunto. Sin embargo, esta estrategia es la más compleja de implementar, ya que, para poder generar las predicciones sobre las probabilidades de inasistencia, es necesario contar con las variables descriptoras correspondientes a cada cita, donde no es posible contar con la totalidad de esta información con un plazo de holgura

que permita definir y posteriormente ofertar los sobrecupos, debido a que la antelación promedio con la que los pacientes reservan las citas es de un día, como se puede ver en la Tabla 4.5. Los sobrecupos deben ser ofertados con una semana de anticipación, teniendo que cargar su programación de oferta a los sistemas de información de la institución con una antelación superior a un día, y, es por esto que, la estrategia que obtiene los mejores resultados no es factible de implementar.

Sin embargo, la predicción de inasistencias en bloques horarios utilizando el modelo de pronóstico y una repartición en base a la tasa de inasistencia de los médicos si es factible de implementar. Esta estrategia posee el segundo mejor desempeño, con una *Precision* similar (63 % vs 65 % en citas de 15 min y 54 % vs 59 % en citas de 20 min) y un menor *Recall* (49 % vs 62 % en citas de 15 min y 54 % vs 62 % en citas de 20 min). A diferencia de la estrategia que obtiene el mejor desempeño, este método se puede implementar ya que cuenta con la información requerida para generar el pronóstico y distribuir las inasistencias en los diferentes bloques horarios con al menos una semana de antelación, recordando que, la predicción de inasistencias en bloques horarios utilizando el modelo de pronóstico y una repartición en base a la tasa de inasistencia de los médicos, se alimenta de una serie de tiempo y de información histórica de los médicos de semanas anteriores a la semana objetivo de predicción correspondiente.

A continuación, se presenta una matriz de confusión que muestra los escenarios que se producen con la atención (teórica) de citas de sobrecupo, en base a la predicción de inasistencias en bloques horarios en los cuales los médicos tienen su agenda completamente reservada, donde esta predicción viene dada por el modelo de pronóstico (ETS) con una repartición de inasistencias utilizando la tasa de inasistencia de los médicos. La matriz de confusión se presenta para el caso de bloques horarios completos en citas de 15 y 20 minutos.

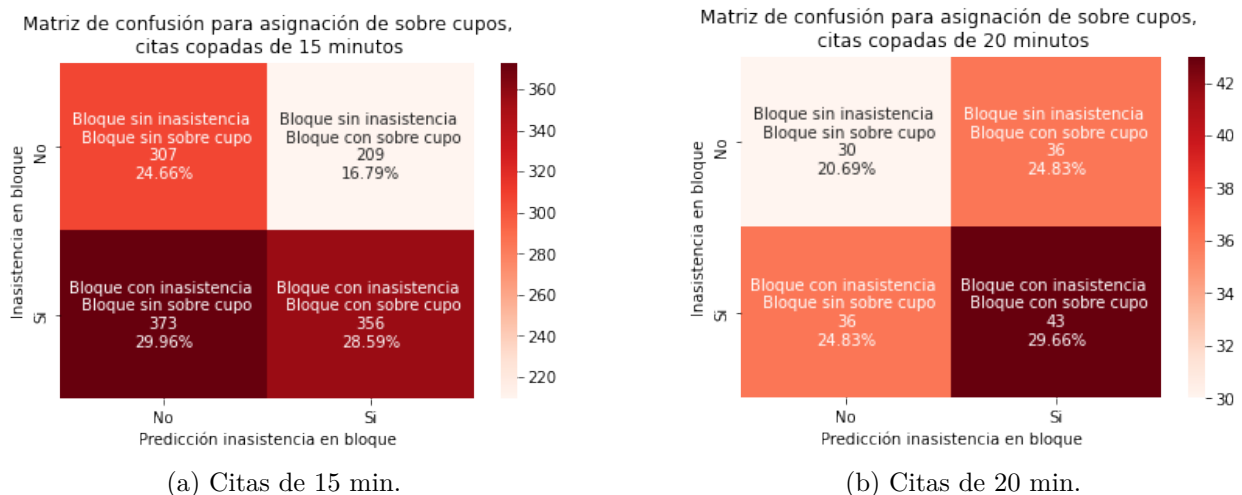


Figura 4.14: Matrices de confusión para escenarios en bloques horarios completos con citas de 15 y 20 min.

En la Figura 4.14, los escenarios contemplados son los siguientes:

1. **Bloque sin inasistencia y sin sobrecupo asignado:** En estos casos, el bloque horario no presenta ninguna inasistencia y no posee ningún sobrecupo asignado, por lo que

este es un escenario con un impacto neutro por parte de la estrategia de definición de sobrecupos. En el caso de las citas de 15 minutos, esto ocurre en un 24.66 % del total de los bloques, mientras que, en las citas de 20 minutos, esto ocurre en un 20.69 % del total de los bloques.

2. **Bloque sin inasistencia y con sobrecupo asignado:** En estos casos, el sobrecupo definido no mitiga la inasistencia de ningún paciente, generando una colisión debido a una atención extra no programada y un mayor tiempo de espera para los pacientes. En el caso de las citas de 15 minutos, esto ocurre en un 16.79 % del total de los bloques, mientras que, en las citas de 20 minutos, esto ocurre en un 24.83 % del total de los bloques.
3. **Bloque con inasistencia y con sobrecupo asignado:** En estos casos, el sobrecupo definido reemplaza una inasistencia, cubriendo una atención programada perdida. En el caso de las citas de 15 minutos, esto ocurre en un 28.59 % del total de los bloques, mientras que, en las citas de 20 minutos, esto ocurre en un 29.66 % del total de los bloques.
4. **Bloque con inasistencia y sin sobrecupo asignado:** En estos casos, se pierde la oportunidad de reemplazar una inasistencia, ya que no se define un sobrecupo en un bloque donde si se presenta una inasistencia. En el caso de las citas de 15 minutos, esto ocurre en un 29.96 % del total de los bloques, mientras que, en las citas de 20 minutos, esto ocurre en un 24.83 % del total de los bloques.

Los porcentajes de ocurrencia de los escenarios señalados anteriormente están directamente ligados a la *Precision* y *Recall* (Tablas 4.19 y 4.22) que obtiene la estrategia utilizada para la definición de sobrecupos dada la predicción de inasistencias en bloques horarios. Así, existe una mayor cantidad de bloques con un impacto positivo que un impacto negativo (*Precision* mayor al 50 %), en ambos bloques con citas de 15 y 20 minutos, y, así mismo, existe una proporción similar de bloques en los cuales si se consigue reemplazar inasistencias y bloques en los cuales no se consigue (*Recall* cercano al 50 %).

4.6. Despliegue

La evaluación anterior indica que, teóricamente, la atención de sobrecupos en base al pronóstico de una cantidad de inasistencias dada por el modelo ETS y la repartición de estas entre los médicos en base a sus tasas de inasistencia, es un método que permite reemplazar la inasistencia de pacientes en bloques horarios, siendo este escenario más frecuente que el escenario donde se producen colisiones. Sin embargo, esta evaluación teórica conlleva un fuerte supuesto, ya que asume que todos aquellos sobrecupos propuestos son agendados y asistidos por parte de los pacientes, y, es por esto que, en búsqueda de relajar este supuesto se lleva a cabo una evaluación en un escenario de bloques de alta demanda, ya que en los bloques horarios de atención completos (bloques con todas las citas disponibles agendadas) existe una mayor posibilidad de que una cita de sobrecupo sea ofertada y reservada, ya que para que una cita de sobrecupo sea visible y reservable por parte los pacientes, debe existir al menos una cita agendada en el bloque horario correspondiente.

De este modo, la institución permite la implementación de un programa piloto utilizando diferentes estrategias para la oferta de sobrecupos derivadas de la estrategia evaluada teóricamente, con objetivo de estudiar el desempeño real de estas. Este piloto considera a un grupo específico de médicos definido por la institución, los cuales previamente forman parte de un programa para la oferta de sobrecupos en base a una lógica desarrollada por la organización.

El piloto contempla las siguientes fases, de acuerdo con el uso de diferentes estrategias.

- **Fase 1 (29 de junio de 2022-17 de julio de 2022):** Inicio del piloto, se reemplaza la estrategia utilizada para ofertar sobrecupos desarrollada por la organización por la estrategia propuesta y evaluada en la Sección 4.5.3, es decir, se definen bloques horarios para ofertar sobrecupos en base a la predicción de inasistencias dada por el modelo ETS y la repartición de la predicción entre los diferentes médicos a través de su tasa individual de inasistencia.
- **Fase 2 (18 de julio de 2022-25 de septiembre de 2022):** Se modifica la estrategia anterior de manera que se definen bloques horarios para ofertar sobrecupos utilizando directamente el pronóstico del modelo ETS en cada médico, eliminando la repartición de la cantidad predicha. A partir del 29 de agosto, se añade una ventana de tiempo de 15 minutos entre bloques con oferta de sobrecupos contiguos.
- **Fase 3 (26 de septiembre de 2022-Actualidad):** Se reemplaza la estrategia anterior por una que define un solo bloque horario que abarca todo el intervalo diario de atención del médico, en el cual es posible reservar una cantidad determinada de sobrecupos, cantidad que es definida por la institución para cada médico en particular.

Las modificaciones en las estrategias son coordinadas en conjunto con la institución y se realizan en base al desempeño real observado al implementar cada una de estas.

Específicamente, luego de implementar la estrategia inicial, se observa una baja conversión desde las citas de sobrecupos ofertadas hacia las citas de sobrecupos atendidas, es decir, la cantidad de sobrecupos asistidos resultó considerablemente menor que la cantidad de sobrecupos ofertados, lo que se produce debido al bajo periodo de visibilidad que tienen las citas de sobrecupos en esta especialidad médica en particular, donde las citas regulares son reservadas con una baja antelación, lo que produce que una cita de sobrecupo sea visible para ser reservada durante un breve periodo de tiempo. Esto último indica que el supuesto tomado en la evaluación teórica de la estrategia se encuentra lejano de la realidad.

Frente a la baja cantidad de sobrecupos atendidos en base a la estrategia inicial, se elimina la repartición de la predicción entre los diferentes médicos utilizando la misma predicción del modelo de pronóstico para ofertar sobrecupos en cada uno de ellos. Esto último se lleva a cabo con el objetivo de lograr una mayor asignación en las citas de sobrecupos, es decir, se incrementa considerablemente la oferta con objetivo de obtener una mayor cantidad de agendamientos en las citas de sobrecupos. Transcurridas seis semanas desde el inicio de la Fase 2, se incluye una ventana de tiempo de 15 minutos entre bloques contiguos para la oferta de sobrecupos, para así evitar la posibilidad de que el médico deba atender una cantidad de pacientes mayor a su capacidad cuando se presentan bloques de atención completos y colindantes en los que se atienden sobrecupos y no se producen inasistencias.

Por último, en coordinación con los médicos del centro, se define una cantidad máxima de sobrecupos de atención, dada la solicitud de regular la cantidad de sobrecupos diarios a atender por parte de los médicos, con lo que se implementan bloques horarios que contemplan todo el intervalo de atención médica de cada especialista y que permiten la reserva del máximo de sobrecupos correspondiente para cada uno.

Para comenzar el análisis de los resultados obtenidos en cada una de las fases, en la siguiente figura se presenta la cantidad de citas de sobrecupo y citas regulares que son ofertadas durante el período previo al programa piloto y durante cada fase del piloto.

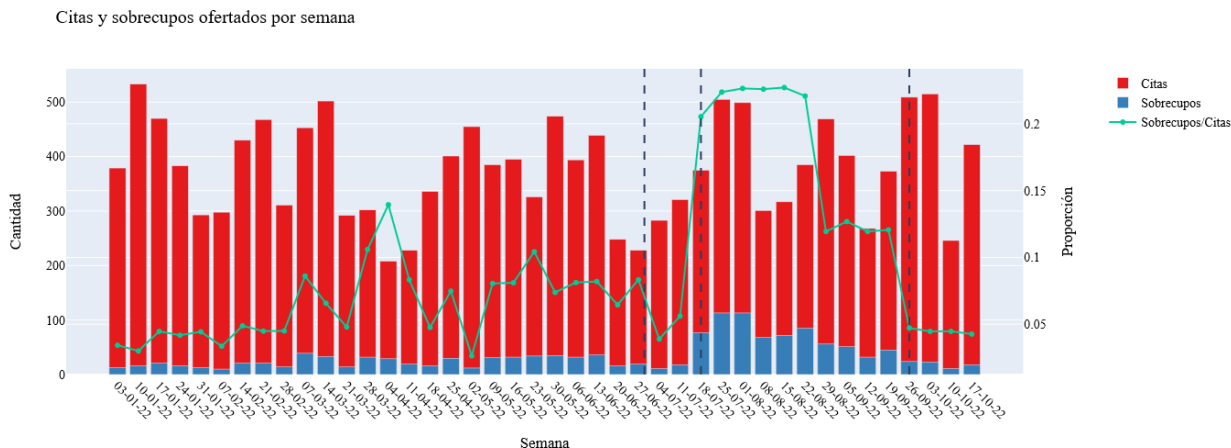
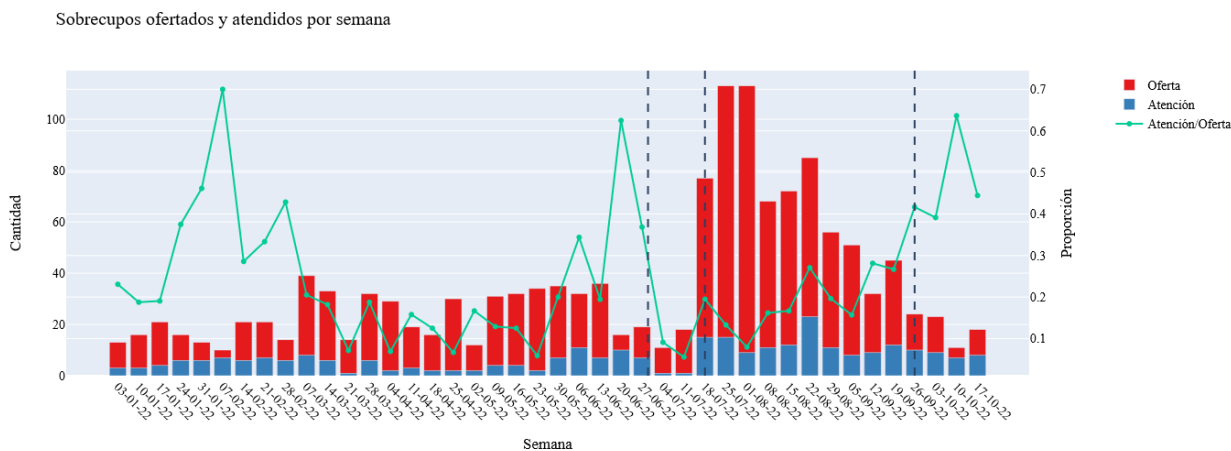


Figura 4.15: Cantidad total de citas regulares y citas de sobrecupo ofertadas por semana (barras superpuestas).

En la Figura 4.15 las líneas punteadas verticales muestran el inicio de cada fase dados los cambios en la estrategia implementada. Como se observa, durante la Fase 1 la relación entre la oferta de sobrecupos respecto a las citas normales es menor que en las semanas anteriores previas al inicio del piloto. Luego, durante la Fase 2, esta proporción aumenta, dado el considerable aumento en la cantidad de sobrecupos ofertados. Durante la Fase 2 se observa una disminución en la proporción desde la semana del 29 de agosto, lo que ocurre debido a la inclusión de la ventana de tiempo de “amortiguación” entre bloques contiguos con oferta de sobrecupos. Finalmente, durante la Fase 3, la proporción vuelve a disminuir, ya que la cantidad de sobrecupos ofertados disminuye con el uso de la nueva estrategia que incluye la restricción de disponibilizar un máximo de sobrecupos reservables para cada médico.

Continuando el análisis, en la siguiente figura se presenta una comparación entre la cantidad de sobrecupos ofertados y la cantidad de sobrecupos atendidos.



aquellos bloques con inasistencias (*Recall*) y de definir con exactitud los bloques para la atención de sobrecupos (*Precision*).

A continuación, se presentan las medidas de *Precision* y *Recall* obtenidas en cada fase.

Tabla 4.24: Resultados sobre la atención de sobrecupos (Pre piloto).

		Precision	Recall	N
Inasistencia en bloque	No	0.56	0.94	1,463
	Si	0.34	0.04	1,115
				2,578 Total

Tabla 4.25: Resultados sobre la atención de sobrecupos (Fase 1).

		Precision	Recall	N
Inasistencia en bloque	No	0.64	0.99	135
	Si	0.67	0.05	78
				213 Total

Tabla 4.26: Resultados sobre la atención de sobrecupos (Fase 2).

		Precision	Recall	N
Inasistencia en bloque	No	0.65	0.88	659
	Si	0.35	0.12	351
				1,010 Total

Tabla 4.27: Resultados sobre la atención de sobrecupos (Fase 3).

		Precision	Recall	N
Inasistencia en bloque	No	0.69	0.93	292
	Si	0.34	0.08	135
				427 Total

Tabla 4.28: Resultados sobre la atención de sobrecupos (Piloto completo).

		Precision	Recall	N
Inasistencia en bloque	No	0.66	0.91	1,086
	Si	0.36	0.10	564
				1,650 Total

En la Tabla 4.24 se presenta el desempeño de la estrategia para la oferta de sobrecupos utilizada por la organización antes del inicio del piloto. En particular, esta estrategia obtiene un *Recall* del 4 %, generando la atención de sobrecupos en el 4 % del total de bloques donde se presentan inasistencias. Asimismo, esta estrategia alcanza una *Precision* del 34 %, donde se presentan inasistencias en el 34 % de los bloques donde se atienden sobrecupos.

En la Tabla 4.25 se presenta el desempeño de la primera estrategia implementada durante el programa piloto, la cual obtiene un *Recall* del 5 % y una *Precision* del 67 %.

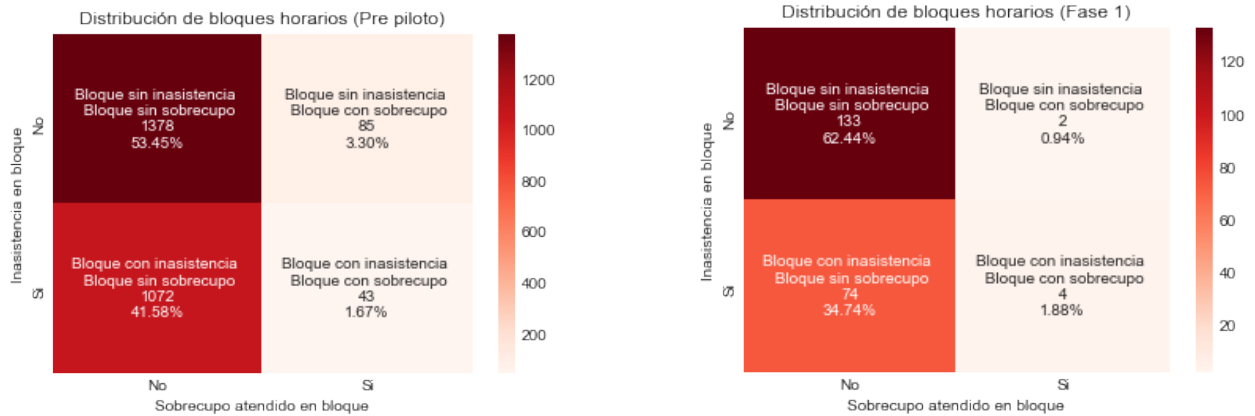
En la Tabla 4.26 se presenta el desempeño de la estrategia utilizada durante la Fase 2, la cual obtiene un *Recall* del 12 % y una *Precision* del 35 %.

En la Tabla 4.27, se presenta el desempeño de la estrategia utilizada durante la Fase 3, la cual obtiene un *Recall* del 8 % y una *Precision* del 34 %.

Por último, en la Tabla 4.28 se presenta el desempeño obtenido al agrupar los bloques horarios pertenecientes a las tres fases del piloto, donde se obtiene una *Precision* del 36 % y un *Recall* del 10 %.

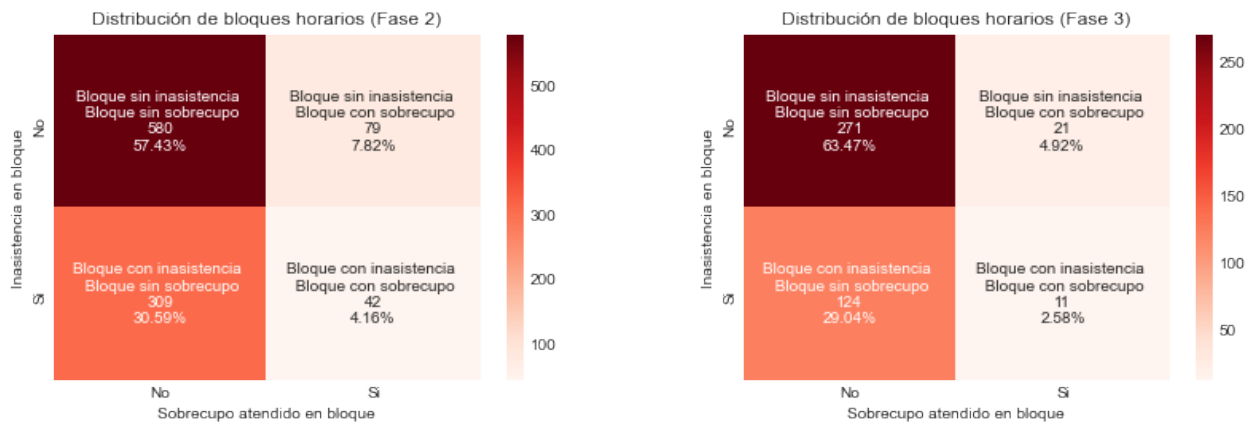
En base a los resultados obtenidos, la primera estrategia implementada es la que alcanza una mayor exactitud en la atención de sobrecupos, con una cobertura levemente mayor que la obtenida por la estrategia utilizada previo al piloto. La segunda estrategia, que elimina la repartición del pronóstico, produce un aumento considerable en la cobertura, con un *Recall* que aumenta 7 p.p. con respecto al *Recall* de Fase 1, pero, también disminuye la *Precision*, pasando de un 67 % en Fase 1 a un 35 % en Fase 2. Lo anterior es razonable dada la naturaleza de cada estrategia, donde en Fase 1 se ofertan sobrecupos en base a bloques horarios en los cuales se pronostica la inasistencia de pacientes y se distribuye este pronóstico en médicos con una mayor tasa de inasistencia, definiendo de manera más precisa aquellos bloques en los cuales ofertar sobrecupos, pero cubriendo una menor cantidad de bloques al distribuir el pronóstico entre médicos. En cambio, en Fase 2, se elimina la repartición definiendo una mayor cantidad de bloques de sobrecupo, lo que produce un aumento de la cobertura y una disminución en la exactitud al definir la misma cantidad de oferta de sobrecupos para cada médico independiente de su tasa de inasistencia. En Fase 3, la estrategia restringe la cantidad de sobrecupos máximos diarios en coordinación con cada médico, disminuyendo así la cantidad de sobrecupos ofertados y, por consiguiente, la cobertura, pasando de un *Recall* de 12 % en Fase 2 a un 8 % en Fase 3, mientras que la exactitud se mantiene en un valor similar a la estrategia previa, pasando de una *Precision* de 35 % en Fase 2 a un 34 % en Fase 3, donde la disminución en el valor de esta última medida se explica debido a una oferta de sobrecupos a lo largo de todo el intervalo de atención diario del médico, sin tomar en consideración antecedentes sobre un pronóstico de inasistencias.

En cuanto a la distribución de los escenarios que se pueden dar en un bloque horario debido a la atención de sobrecupos, a continuación se presentan matrices de confusión que detallan esta distribución de acuerdo con el total de bloques horarios de atención pertenecientes a cada fase.



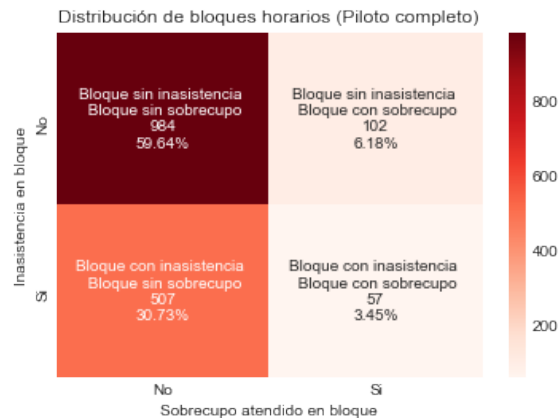
(a) Período previo al inicio del piloto.

(b) Período de Fase 1.



(c) Período de Fase 2.

(d) Período de Fase 3.



(e) Período completo del piloto.

Figura 4.17: Matrices de confusión sobre la distribución de escenarios producidos por cada estrategia implementada.

En la Figura 4.17 se observa que la atención de sobrecupos producida por las diferentes estrategias ocurre mayoritariamente en bloques horarios donde no ocurren inasistencias, a excepción de la estrategia utilizada durante la Fase 1, la cual es la única estrategia que obtiene una *Precision* mayor al 50%. Así mismo, se observa que existe una gran cantidad de bloques horarios en los que sí ocurren inasistencias y no se atienden sobrecupos, lo que se ve

directamente reflejado en los valores de *Recall* obtenidos por cada estrategia, los cuales no superan el 12%.

En particular, en la Figura 4.17 (a) se observa que en un 43% de los bloques horarios ocurren inasistencias en el período previo al inicio del piloto, mientras que, en la Figura 4.17 (e) se observa que, en el período del piloto, en un 34% de los bloques horarios ocurren inasistencias, existiendo así una menor cantidad de bloques con inasistencias durante el período del piloto que durante el período previo al piloto.

Otro punto en el análisis corresponde a la evaluación del impacto de la atención de sobrecupos en el reemplazo de las citas no asistidas. Para esto, en la evaluación se considera la totalidad de inasistencias y sobrecupos atendidos de manera agrupada, y no de manera individual a nivel de bloque horario, ya que, se busca medir la cantidad de inasistencias que son reemplazadas con sobrecupos en términos netos, donde una inasistencia no necesariamente requiere ser reemplazada por un sobrecupo atendido en exactamente el mismo bloque horario.

A continuación, se presenta la proporción de sobrecupos atendidos sobre el total de inasistencias contabilizadas por semana, o, en otras palabras, el porcentaje de inasistencias que son reemplazadas (o “recuperadas”) con la atención de sobrecupos.

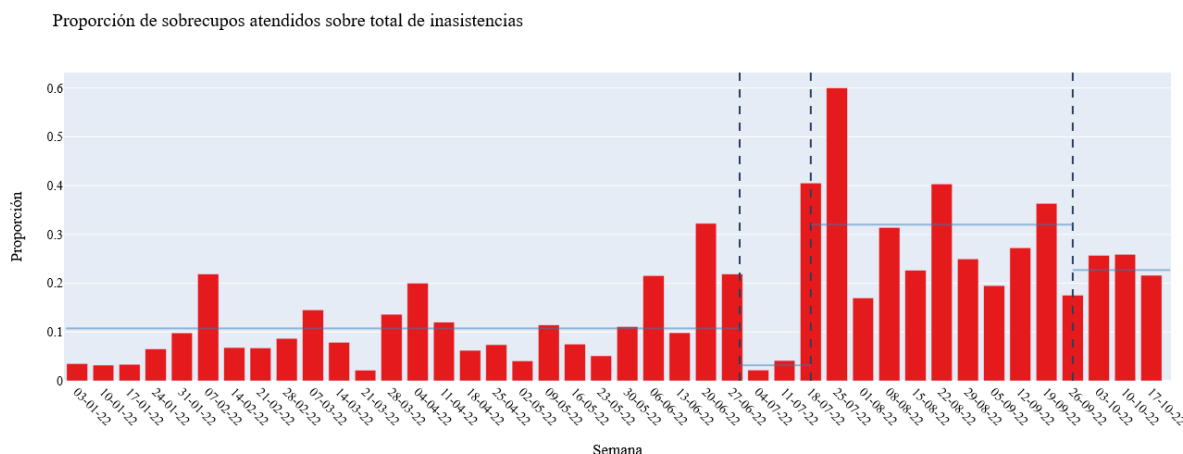


Figura 4.18: Proporción entre la cantidad de sobrecupos atendidos y la cantidad de citas no asistidas por semana.

En la Figura 4.18 las líneas continuas horizontales representan la proporción promedio por semana calculada para cada fase. En particular, se observa que previo al inicio del piloto los sobrecupos atendidos recuperan en promedio un 11% de las inasistencias, mientras que, durante la Fase 1, 2 y 3 los sobrecupos atendidos recuperan en promedio un 3%, 32% y 23% de las inasistencias por semana, respectivamente. Con esto, al comienzo del programa piloto, la proporción de inasistencias reemplazadas disminuye con el uso de la primera estrategia, luego, esta proporción aumenta notablemente al implementar una estrategia mucho más agresiva en la oferta de sobrecupos, y, posteriormente, la proporción vuelve a disminuir al implementar la última estrategia.

Así mismo, se analiza la tasa de inasistencia de acuerdo al total de citas por semana, con-

siderando esta tasa de manera natural y con un ajuste que sustrae la cantidad de sobrecupos atendidos al total de inasistencias. De esta forma, se compara una tasa de inasistencias base con una tasa de inasistencia “atenuada” que considera el efecto de la atención de sobrecupos dada cada una de las estrategias implementadas.

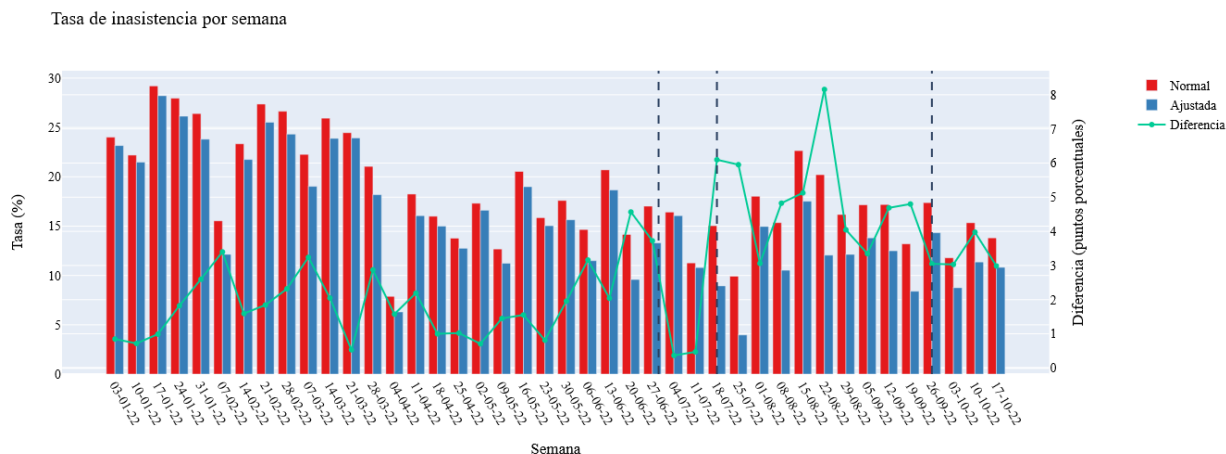


Figura 4.19: Tasa de inasistencia base y atenuada por semana.

En la Figura 4.19 se observa que la diferencia entre la tasa de inasistencia base y atenuada tiende a ser menor en el periodo previo al inicio del piloto. Específicamente, la diferencia promedio entre tasas es de 1.94 p.p. antes del inicio del piloto, 0.41 p.p. durante la Fase 1, 5 p.p. durante la Fase 2 y 3.26 p.p. durante la Fase 3. De esta forma, la diferencia entre tasas más grande se obtiene durante la Fase 2, período en el cual se utiliza la estrategia más agresiva en la oferta de sobrecupos, mientras que, la diferencia más pequeña se obtiene durante la Fase 1, período en el cual la cantidad de sobrecupos atendidos fue notoriamente bajo. Por otro lado, la estrategia utilizada durante la Fase 3 obtiene una diferencia entre tasas promedio mayor a la observada previo al inicio del período.

Últimamente, se analizan las consecuencias negativas generadas por la atención de sobrecupos en la operación del centro médico. En las matrices de confusión presentadas anteriormente se observa que, en general, los sobrecupos son atendidos mayormente en bloques horarios donde no se producen inasistencias, sin embargo, estos sobrecupos no necesariamente afectan negativamente la operación, ya que el funcionamiento de esta es perjudicado sólo cuando el médico atiende una cantidad de citas que excede su capacidad horaria en base a la duración predeterminedada de su atención (donde, como se mencionó anteriormente, ciertos médicos atienden en 15 minutos y otros en 20). De esta manera, para cuantificar el efecto negativo, se identifican los sobrecupos que producen una cantidad de atenciones que excede la capacidad de atención del médico en cada bloque horario (cuatro atenciones máximo en médicos con atención de 15 minutos y tres atenciones máximo en médicos con atención de 20 minutos). Así, se definen dos medidas: *Afectación Bruta* y *Afectación Relativa*, donde la primera mide la proporción de sobrecupos que generan un exceso de capacidad en la atención de acuerdo al total de sobrecupos, mientras que la segunda mide el porcentaje de sobrecupos que generan un exceso de capacidad en la atención y donde el médico no cuenta con holgura para recuperar la agenda en el bloque horario de atención siguiente, es decir, si en el bloque horario siguiente al bloque afectado la atención es menor a 4 citas (en caso de médicos que

atienden en 15 minutos) o menor a 3 citas (en caso de médicos que atienden en 20 minutos) entonces existe holgura y por lo tanto no se considera que el sobrecupo afectó negativamente la operación.

En la siguiente figura se presenta la *Afectación Bruta* y *Afectación Relativa* calculada por semana. Además, se incluye la cantidad de sobrecupos atendidos para dimensionar la cantidad de sobrecupos que producen afectación.

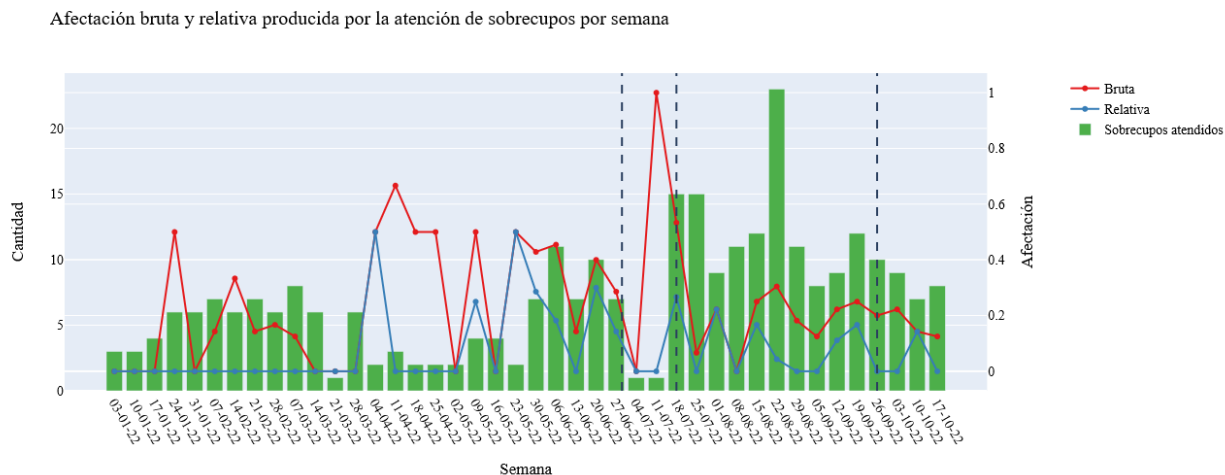


Figura 4.20: Afectación Bruta y Afectación Relativa por semana.

En la Figura 4.20 se observa que previo al inicio del piloto la *Afectación Bruta* es bastante variable, con un máximo cercano al 70 %, mientras que la *Afectación Relativa* es menor que la *Afectación Bruta* en la mayoría de las semanas. Por otra parte, al comienzo del programa piloto, se observa un pico del 100 % en la *Afectación Bruta*, donde el único sobrecupo atendido durante esa semana generó un exceso de capacidad en la atención. Posteriormente, durante la Fase 2 y 3 del piloto tanto la *Afectación Bruta* como la *Afectación Relativa* se mantienen bajo el 40 % y presentan una menor variabilidad, con valores que se concentran mayormente alrededor del 30 % y el 0 %. En general, se aprecia que la *Afectación Relativa* es menor que la *Afectación Bruta* en la mayoría de las semanas del piloto. Específicamente, la *Afectación Bruta* promedio por semana durante el período previo al inicio del piloto es de un 24 %, durante la Fase 1 de un 50 %, durante la Fase 2 de un 22 % y durante la Fase 3 de un 16 %. En cambio, la *Afectación Relativa* promedio por semana durante el período previo al inicio del piloto es de un 8 %, durante la Fase 1 de un 0 %, durante la Fase 2 de un 10 % y durante la Fase 3 de un 5 %. Así, solo en la Fase 1 del piloto la *Afectación Bruta* promedio supera al valor obtenido en el período previo al inicio del piloto, mientras que, para la *Afectación Relativa* promedio, esto ocurre solo en la Fase 2 del piloto.

Capítulo 5

Conclusiones

En este trabajo de tesis se presenta el diseño e implementación de una herramienta que permite definir la oferta de sobrecupos para citas médicas, identificando aquellos segmentos horarios con una mayor probabilidad de ocurrencia de inasistencias. En particular, mediante el uso de modelos de aprendizaje de máquinas, se generaron predicciones para la ocurrencia de inasistencias utilizando dos enfoques de modelamiento: clasificación y pronóstico. Al desarrollar y evaluar diferentes modelos, se seleccionó un método que define la oferta de sobrecupos en base a un pronóstico de la cantidad de inasistencias a ocurrir en bloques horarios de una hora, el cual es generado por un modelo ETS, donde la cantidad de inasistencias predichas por el modelo son distribuidas entre diferentes médicos, los cuales cuentan con una oferta programada de citas regulares en aquellos bloques horarios donde se producen predicciones. Los resultados de este método son evaluados favorablemente de forma teórica, lo que lleva a su implementación en un programa piloto. Así, se lleva a cabo una evaluación del desempeño real obtenido por parte del método propuesto. La implementación y posterior evaluación práctica del método posibilitó la realización de diferentes modificaciones que concluyeron en la definición de una estrategia que fue adoptada por la institución, la cual consiste en ofertar citas de sobrecupo en un único bloque horario que contempla el intervalo completo de atención diaria del médico, en el cual se establece una cantidad fija de sobrecupos que pueden ser agendados con cada médico. De esta forma, se cumple con el objetivo general y con los objetivos específicos definidos en este trabajo, presentados en la Sección 1.1. Específicamente, los resultados esperados de los objetivos específicos 1, 2 y 3 son presentados en las Secciones 2, 3.1, 3.2, 4.1, 4.2 y 4.3; Secciones 4.4 y 4.5; y Secciones 4.5.3 y 4.6, respectivamente.

En relación con la estrategia finalmente adoptada, se destaca que esta no involucra una lógica compleja en base a modelos de aprendizaje de máquinas, aunque, esta deriva de la implementación de estrategias previas que fueron disminuyendo en su nivel de complejidad, en las cuales se pudo diagnosticar el desempeño práctico para así posteriormente realizar modificaciones que concluyeron en la estrategia adoptada que fue instaurada al inicio de la Fase 3 del programa piloto. Esta última estrategia obtuvo resultados que mejoran indicadores claves de rendimiento con respecto a la lógica de sobrecupos utilizada por la institución antes del inicio del programa piloto. Concretamente, esta estrategia es capaz de cubrir un 8% de los bloques horarios donde se presentan inasistencias, mientras que la estrategia utilizada previamente por la institución cubría un 4%, incrementando así la cobertura en 4 p.p. Además, durante el período previo al inicio del programa piloto, un 11% de las citas no asistidas en promedio por semana eran reemplazadas por citas de sobrecupo, mientras que, en el período

en el cual se evalúa la estrategia adoptada, un 23% de las citas no asistidas en promedio por semana son reemplazadas por citas de sobrecupo, aumentando así el reemplazo en 12 p.p. Por otra parte, en el período previo al inicio del programa piloto, la diferencia promedio entre la tasa de inasistencia base y atenuada por semana es de 1.94 p.p., mientras que, en el período en el cual se utiliza el método finalmente seleccionado, la diferencia promedio por semana es de 3.26 p.p., aumentando así esta diferencia en 1.32 p.p.

Con respecto al método propuesto e implementado al inicio del programa piloto, se pudo observar que este obtuvo una baja cantidad de sobrecupos atendidos, lo que se debe a un bajo periodo de visibilidad de las citas de sobrecupo ofertadas, ya que, si bien se genera una predicción de aquellos bloques horarios que tienen un mayor riesgo frente a la ocurrencia de inasistencias, no se complementa esta información con una estimación de la demanda por estos mismos sobrecupos. Así, en la mayoría de los casos en los que se ofertaron sobrecupos durante Fase 1, no se presentó una demanda tal que permitiese visibilizar estas citas de sobrecupo durante un tiempo lo suficientemente grande como para que estas fuesen agendadas, donde las citas médicas de esta especialidad en particular son agendadas con un día de anticipación en promedio. Con esto, se concluye que para una especialidad médica que posee una baja antelación en la reserva de sus citas, es relevante complementar la predicción de inasistencias con una estimación de demanda para las citas de sobrecupo propuestas, ya que, si bien es clave definir donde existe un mayor riesgo de sufrir inasistencias, es de igual importancia estimar los segmentos donde existe una mayor posibilidad de que los sobrecupos ofertados sean agendados. En particular, el no considerar la demanda por sobrecupos puede derivar en un desempeño deficiente de la herramienta debido a una baja cantidad de sobrecupos atendidos, como ocurre durante la Fase 1 del programa piloto implementado en este trabajo.

Por otro lado, la estrategia adoptada e instaurada en la Fase 3 del programa piloto considera todo el intervalo diario de atención de los médicos como un bloque único para agendar citas de sobrecupo, y, de esta forma, los sobrecupos se asignan “por sí solos” en aquellos segmentos horarios donde se presenta una demanda tal que permite que la cita de sobrecupo sea visible durante un período de tiempo suficiente para que sea agendada. Además, aún cuando esta estrategia no discrimina aquellos segmentos horarios con un mayor riesgo de ocurrencia de inasistencias, no produce un efecto negativo considerable debido a una excesiva atención de sobrecupos, lo que se debe a diversas razones, entre las cuales se encuentran: la existencia de un máximo de citas de sobrecupo que pueden ser agendadas en cada médico, donde la mayoría de estos atiende un máximo de un sobrecupo diario; la alta probabilidad de que se produzca una inasistencia en los bloques horarios donde el médico cuenta con su agenda completa, tomando en cuenta que en esta especialidad aproximadamente una de cada cuatro citas termina en una inasistencia; y, por último, la posibilidad de recuperar la atención en el bloque horario siguiente cuando se producen colisiones, donde, como se observa en la *Afectación Relativa* durante la Fase 3, en promedio solo un 5% de los sobrecupos atendidos por semana producen un escenario de colisión en el cual el médico no cuenta con holgura para recuperar la atención en el bloque horario siguiente.

Se destaca también que el análisis teórico sobre los efectos de la implementación de un nuevo método permitió abordar la resistencia al cambio del área de trabajo colaboradora, respecto a la implementación de estrategias más agresivas en la oferta de sobrecupos, donde

esta resistencia al cambio poseía su origen en la preocupación de generar efectos negativos de gran magnitud en la operación del centro médico. En particular, durante la Fase 2 del programa piloto, la cantidad de sobrecupos ofertados fue considerablemente mayor que la cantidad de sobrecupos ofertados históricamente en semanas previas, donde el aumento considerable en la cantidad de sobrecupos ofertados no implicó el mismo aumento en la cantidad de sobrecupos atendidos, por lo que no se generó una crisis operacional producida por un aumento excesivo en la atención de sobrecupos, lo que se puede ver reflejado en una *Afectación Bruta* en Fase 2 menor que la observada previo al inicio del piloto, y una *Afectación Relativa* en Fase 2 ligeramente superior a la observada previo al inicio del piloto. Además, el período de Fase 2 del programa piloto permitió que el área pudiese advertir el nivel de sensibilidad de los médicos respecto a la cantidad de sobrecupos diarios dispuestos a atender por parte de estos.

Con respecto a los modelos de clasificación, se observa que estos no consiguen un buen desempeño, obteniendo un escenario de subajuste con bajos valores en las medidas de evaluación en el conjunto de entrenamiento, testing y proyección. Se concluye que este mal desempeño se debe principalmente al conjunto de datos utilizado, el cual no cuenta con variables descriptoras lo suficientemente explicativas para lograr diferenciar a la clase de interés con precisión, lo cual se puede observar en la relevancia de variables entregada por los modelos con mejor desempeño, donde gran parte de las variables poseen coeficientes cercanos a cero en el caso de Logit, indicando una contribución prácticamente nula a la probabilidad de pertenecer a la clase inasistencia por parte de las variables, mientras que, en el caso de XGBoost, son solo dos o tres variables las que poseen una ganancia de información promedio relevante, donde el resto de atributos poseen ganancias insignificantes o derechamente nulas.

Con respecto a los modelos de pronóstico, estos consiguen un buen desempeño en el modelamiento de la cantidad de inasistencias que se producen en bloques horarios, con predicciones que consiguen un bajo nivel de error. Sin embargo, este enfoque de modelamiento presenta la desventaja de entregar una directriz de información muy general en la cantidad de inasistencias predicha, donde los médicos no son recursos indistinguibles entre los cuales se puede distribuir la predicción directamente, ya que los pacientes cuentan con diferentes preferencias a la hora de reservar sus citas médicas, lo que genera la necesidad de repartir la cantidad de inasistencias predichas por los modelos de pronóstico bajo la definición de ciertas reglas. En cuanto a los modelos de pronóstico utilizados, se resalta que el uso de RNNs como un modelo de pronóstico consigue mejores resultados que modelos más tradicionales como ARIMA y ETS, aunque, la mejoría en el desempeño no es radicalmente grande como para justificar su uso, tomando en cuenta que las RNNs son modelos más complejos que requieren de una mayor calibración.

Referente al acoplamiento de modelos, se destaca que una combinación entre un modelo de clasificación y un modelo de pronóstico produce mejores resultados en la predicción de inasistencias para bloques horarios que los obtenidos mediante el uso de un modelo de clasificación por sí solo. En particular, se observa que al acoplar ambos enfoques de modelamiento se produce una sinergia que permite contrarrestar el exceso en la predicción de la clase inasistencia generada por los modelos de clasificación cuando se utilizan umbrales de clasificación pequeños, donde los modelos de pronóstico entregan una suerte de “máximo” para la cantidad de inasistencias predichas en un bloque horario. Por otra parte, resulta relevante señalar

que la utilización de la probabilidad estimada por modelos de clasificación para distribuir la cantidad de inasistencias pronosticadas produce mejores resultados que una distribución en base a una tasa de inasistencia histórica de los médicos.

Bibliografía

- [1] Fiorillo, C. E., Hughes, A. L., I-Chen, C., Westgate, P. M., Gal, T. J., Bush, M. L., y Comer, B. T., “Factors associated with patient no-show rates in an academic otolaryngology practice,” *The Laryngoscope*, vol. 128, pp. 626–631, 2017, [doi:10.1002/lary.26816](https://doi.org/10.1002/lary.26816).
- [2] Kheirkhah, P., Feng, Q., Travis, L. M., Tavakoli-Tabasi, S., y Sharafkhaneh, A., “Prevalence, predictors and economic consequences of no-shows,” *BMC Health Services Research*, vol. 16, 2015, [doi:10.1186/s12913-015-1243-z](https://doi.org/10.1186/s12913-015-1243-z).
- [3] Alaeddini, A., Yang, K., Reeves, P., y Reddy, C. K., “A hybrid prediction model for no-shows and cancellations of outpatient appointments,” *IIE Transactions on Healthcare Systems Engineering*, vol. 5, pp. 14–32, 2015, [doi:10.1080/19488300.2014.993006](https://doi.org/10.1080/19488300.2014.993006).
- [4] Moore, C., Wilson-Witherspoon, P., y Probst, J., “Time and money: Effects of no-shows at a family practice residency clinic,” *Family medicine*, vol. 33, pp. 522–7, 2001.
- [5] Molfenter, T., “Reducing appointment no-shows: Going from theory to practice,” *Substance Use & Misuse*, vol. 48, pp. 743–749, 2013, [doi:10.3109/10826084.2013.787098](https://doi.org/10.3109/10826084.2013.787098).
- [6] Nguyen, D. L., DeJesus, R. S., y Wieland, M. L., “Missed appointments in resident continuity clinic: Patient characteristics and health care outcomes,” *Journal of Graduate Medical Education*, vol. 3, pp. 350–355, 2011, [doi:10.4300/jgme-d-10-00199.1](https://doi.org/10.4300/jgme-d-10-00199.1).
- [7] Verdugo, A., “Rediseño del proceso de bloqueo de citas en un prestador de salud privada de nivel hospitalario, ambulatorio y dental.” Departamento de Ingeniería Industrial, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, 2021. [Memoria para optar al título de Ingeniera Civil Industrial].
- [8] Xie, X., Fan, Z., y Zhong, X., “Appointment capacity planning with overbooking for outpatient clinics with patient no-shows,” *IEEE Transactions on Automation Science and Engineering*, pp. 1–20, 2021, [doi:10.1109/tase.2021.3060567](https://doi.org/10.1109/tase.2021.3060567).
- [9] Huang, Y. y Zuniga, P., “Dynamic overbooking scheduling system to improve patient access,” *Journal of the Operational Research Society*, vol. 63, pp. 810–820, 2012, [doi:10.1057/jors.2011.83](https://doi.org/10.1057/jors.2011.83).
- [10] LaGanga, L. R. y Lawrence, S. R., “Clinic overbooking to improve patient access and increase provider productivity,” *Decision Sciences*, vol. 38, pp. 251–276, 2007, [doi:10.1111/j.1540-5915.2007.00158.x](https://doi.org/10.1111/j.1540-5915.2007.00158.x).
- [11] Elgueta, B., “Desarrollo de un modelo predictivo para apoyar la gestión de la agenda médica en un centro médico.” Departamento de Ingeniería Industrial, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, 2021. [Memoria para optar al título de Ingeniero Civil Industrial].
- [12] Fan, G., Deng, Z., Ye, Q., y Wang, B., “Machine learning-based prediction models for

- patients no-show in online outpatient appointments,” *Data Science and Management*, vol. 2, pp. 45–52, 2021, [doi:10.1016/j.dsm.2021.06.002](https://doi.org/10.1016/j.dsm.2021.06.002).
- [13] Kurasawa, H., Hayashi, K., Fujino, A., Takasugi, K., Haga, T., Waki, K., Noguchi, T., y Ohe, K., “Machine-learning-based prediction of a missed scheduled clinical appointment by patients with diabetes,” *Journal of Diabetes Science and Technology*, vol. 10, pp. 730–736, 2015, [doi:10.1177/1932296815614866](https://doi.org/10.1177/1932296815614866).
- [14] Mohammadi, I., Wu, H., Turkcan, A., Toscos, T., y Doebbeling, B. N., “Data analytics and modeling for appointment no-show in community health centers,” *Journal of Primary Care & Community Health*, vol. 9, p. 215013271881169, 2018, [doi:10.1177/2150132718811692](https://doi.org/10.1177/2150132718811692).
- [15] Simsek, S., Tiahr, T., y Dag, A., “Stratifying no-show patients into multiple risk groups via a holistic data analytics-based framework,” *Decision Support Systems*, p. 113269, 2020, [doi:10.1016/j.dss.2020.113269](https://doi.org/10.1016/j.dss.2020.113269).
- [16] Salazar, L. H. A., Parreira, W. D., Fernandes, A. M. d. R., y Leithardt, V. R. Q., “No-show in medical appointments with machine learning techniques: A systematic literature review,” *Information*, vol. 13, p. 507, 2022, [doi:10.3390/info13110507](https://doi.org/10.3390/info13110507).
- [17] Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez Orallo, J., Kull, M., Lachiche, N., Ramirez Quintana, M. J., y Flach, P. A., “Crisp-dm twenty years later: From data mining processes to data science trajectories,” *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–1, 2020, [doi:10.1109/tkde.2019.2962680](https://doi.org/10.1109/tkde.2019.2962680).
- [18] Simon, N., Friedman, J., Hastie, T., y Tibshirani, R., “A sparse-group lasso,” *Journal of Computational and Graphical Statistics*, vol. 22, pp. 231–245, 2013, [doi:10.1080/10618600.2012.681250](https://doi.org/10.1080/10618600.2012.681250).
- [19] Yuan, M. y Lin, Y., “Model selection and estimation in regression with grouped variables,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, pp. 49–67, 2006, [doi:10.1111/j.1467-9868.2005.00532.x](https://doi.org/10.1111/j.1467-9868.2005.00532.x).
- [20] Chen, T. y Guestrin, C., “Xgboost: A scalable tree boosting system,” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, 2016, [doi:10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785).
- [21] Box, G. E. P. y Jenkins, G. M., *Time series analysis : forecasting and control*. Holden-Day, 1976.
- [22] Hyndman, R., *Forecasting with exponential smoothing : the state space approach*. Springer, 2008.
- [23] Hyndman, R. J. y Khandakar, Y., “Automatic time series forecasting: The forecast package for r,” *Journal of Statistical Software*, vol. 27, 2008, [doi:10.18637/jss.v027.i03](https://doi.org/10.18637/jss.v027.i03).
- [24] Hochreiter, S. y Schmidhuber, J., “Long short-term memory,” *Neural Computation*, vol. 9, pp. 1735–1780, 1997, [doi:10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- [25] Elman, J. L., “Finding structure in time,” *Cognitive Science*, vol. 14, pp. 179–211, 1990, [doi:10.1207/s15516709cog1402_1](https://doi.org/10.1207/s15516709cog1402_1).
- [26] Hyndman, R. J. y Koehler, A. B., “Another look at measures of forecast accuracy,” *International Journal of Forecasting*, vol. 22, pp. 679–688, 2006, [doi:10.1016/j.ijforecast.2006.03.001](https://doi.org/10.1016/j.ijforecast.2006.03.001).

- [27] Kuhn, M. y Johnson, K., Applied predictive modeling. Springer, 2019.
- [28] Hewamalage, H., Bergmeir, C., y Bandara, K., “Recurrent neural networks for time series forecasting: Current status and future directions,” *International Journal of Forecasting*, vol. 37, pp. 388–427, 2021, [doi:10.1016/j.ijforecast.2020.06.008](https://doi.org/10.1016/j.ijforecast.2020.06.008).
- [29] Branco, P., Torgo, L., y Ribeiro, R., “A survey of predictive modelling under imbalanced distributions,” 2015.
- [30] He, H., Ma, Y., y Wiley, J., Imbalanced learning : foundations, algorithms, and applications. John Wiley & Sons, Cop, 2013.

Anexos B. Importancia de atributos.

Tabla B.1: Coeficientes estimados para cada variable por modelo Logit para predicción de inasistencias de pacientes nuevos.

Variable	Valor	Coeficiente	Odds (e^{coef})								
<i>Variables categóricas</i>											
Canal creación	Portal Web	-0.70	0.50	Día creación cita	13	-0.09	0.91	Hora creación cita	11	-0.27	0.76
Canal creación	Call Center	-1.18	0.31	Día creación cita	9	-0.13	0.88	Hora creación cita	20	-0.37	0.69
Canal creación	Recepción	-1.78	0.17	Día creación cita	22	-0.14	0.87	Hora creación cita	10	-0.40	0.67
Día cita	21	0.20	1.22	Día creación cita	31	-0.15	0.86	Hora creación cita	9	-0.41	0.66
Día cita	14	0.19	1.21	Día creación cita	19	-0.16	0.85	Hora creación cita	8	-0.44	0.65
Día cita	28	0.19	1.21	Día creación cita	2	-0.18	0.84	Mes cita	4	0.50	1.64
Día cita	4	0.16	1.18	Día creación cita	18	-0.20	0.82	Mes cita	5	0.38	1.46
Día cita	15	0.16	1.17	Día creación cita	21	-0.20	0.82	Mes cita	8	0.36	1.43
Día cita	18	0.16	1.17	Día creación cita	27	-0.21	0.81	Mes cita	10	0.21	1.23
Día cita	19	0.15	1.16	Día creación cita	26	-0.27	0.76	Mes cita	3	0.11	1.12
Día cita	13	0.13	1.14	Día creación cita	14	-0.30	0.74	Mes cita	2	0.01	1.01
Día cita	26	0.09	1.10	Día semana cita	Sábado	0.01	1.01	Mes cita	7	-0.02	0.98
Día cita	27	0.09	1.10	Día semana cita	Miércoles	-0.07	0.93	Mes cita	9	-0.02	0.98
Día cita	29	0.09	1.09	Día semana cita	Jueves	-0.14	0.87	Mes cita	11	-0.02	0.98
Día cita	2	0.07	1.07	Día semana cita	Martes	-0.22	0.80	Mes cita	6	-0.31	0.73
Día cita	31	0.07	1.07	Día semana cita	Lunes	-0.23	0.79	Mes cita	12	-0.60	0.55
Día cita	30	0.06	1.06	Día semana creación cita	Sábado	0.32	1.37	Mes creación cita	12	0.82	2.28
Día cita	12	0.05	1.06	Día semana creación cita	Domingo	0.19	1.21	Mes creación cita	6	0.34	1.40
Día cita	22	0.04	1.04	Día semana creación cita	Jueves	-0.03	0.97	Mes creación cita	11	-0.01	0.99
Día cita	24	0.02	1.02	Día semana creación cita	Martes	-0.05	0.95	Mes creación cita	9	-0.06	0.94
Día cita	23	0.01	1.01	Día semana creación cita	Lunes	-0.08	0.92	Mes creación cita	3	-0.23	0.80
Día cita	3	-0.03	0.97	Día semana creación cita	Miércoles	-0.09	0.92	Mes creación cita	10	-0.26	0.77
Día cita	10	-0.03	0.97	Duración	20	-0.13	0.87	Mes creación cita	2	-0.27	0.76
Día cita	20	-0.05	0.95	Hora cita	19	0.14	1.14	Mes creación cita	4	-0.30	0.74
Día cita	11	-0.08	0.92	Hora cita	16	0.06	1.06	Mes creación cita	5	-0.30	0.74
Día cita	25	-0.12	0.88	Hora cita	15	-0.03	0.97	Mes creación cita	7	-0.30	0.74
Día cita	8	-0.13	0.88	Hora cita	17	-0.03	0.97	Mes creación cita	8	-0.55	0.58
Día cita	7	-0.15	0.86	Hora cita	18	-0.04	0.96	Plan de salud	Fonasa	0.37	1.45
Día cita	9	-0.16	0.85	Hora cita	13	-0.07	0.93	Plan de salud	Particular	0.23	1.26
Día cita	5	-0.17	0.84	Hora cita	14	-0.08	0.93	Plan de salud	Isapre	0.01	1.01
Día cita	17	-0.17	0.85	Hora cita	12	-0.19	0.83	Sexo	Masculino	-0.05	0.95
Día cita	6	-0.26	0.77	Hora cita	9	-0.33	0.72	Tiempo visita	PM	-0.24	0.78
Día cita	16	-0.30	0.74	Hora cita	11	-0.42	0.66	<i>Variables numéricas</i>			
Día creación cita	25	0.25	1.29	Hora cita	10	-0.43	0.65	Tasa inasistencia médico		0.65	1.92
Día creación cita	5	0.18	1.20	Hora creación cita	5	0.37	1.44	Inasistencias médico		0.10	1.10
Día creación cita	24	0.18	1.20	Hora creación cita	1	0.20	1.22	Antelacion cita		0.02	1.02
Día creación cita	16	0.14	1.15	Hora creación cita	3	0.07	1.08	Demanda médico		-0.02	0.98
Día creación cita	11	0.12	1.13	Hora creación cita	17	0.03	1.03	Atenciones médico		-0.03	0.97
Día creación cita	3	0.11	1.12	Hora creación cita	19	-0.06	0.94	Edad paciente		-0.16	0.85
Día creación cita	8	0.09	1.09	Hora creación cita	16	-0.07	0.94				
Día creación cita	7	0.08	1.08	Hora creación cita	18	-0.08	0.93				
Día creación cita	4	0.04	1.04	Hora creación cita	6	-0.09	0.91				
Día creación cita	6	0.03	1.03	Hora creación cita	23	-0.12	0.89				
Día creación cita	10	0.03	1.04	Hora creación cita	4	-0.13	0.88				
Día creación cita	17	0.03	1.04	Hora creación cita	13	-0.13	0.88				
Día creación cita	12	0.02	1.02	Hora creación cita	2	-0.15	0.86				
Día creación cita	20	0.02	1.02	Hora creación cita	22	-0.15	0.86				
Día creación cita	28	-0.01	0.99	Hora creación cita	15	-0.16	0.85				
Día creación cita	29	-0.03	0.97	Hora creación cita	21	-0.18	0.84				
Día creación cita	30	-0.04	0.96	Hora creación cita	12	-0.20	0.82				
Día creación cita	15	-0.07	0.93	Hora creación cita	14	-0.25	0.78				
Día creación cita	23	-0.07	0.93	Hora creación cita	7	-0.27	0.77				

Tabla B.2: Promedio de coeficientes estimados para variables categóricas por modelo Logit para predicción de inasistencias de pacientes nuevos.

Variable	Promedio Coeficientes	Promedio Odds
Canal creación	-1.22	0.33
Día cita	0.01	1.02
Día creación cita	-0.03	0.98
Día semana cita	-0.13	0.88
Día semana creación cita	0.04	1.06
Duración	-0.13	0.87
Hora cita	-0.13	0.89
Hora creación cita	-0.14	0.88
Mes cita	0.05	1.10
Mes creación cita	-0.10	0.98
Plan de salud	0.20	1.24
Sexo	-0.05	0.95
Tiempo visita	-0.24	0.78

Tabla B.3: Ganancia promedio de información por variable en base a modelo XGBoost para predicción de inasistencias de pacientes recurrentes.

Variable	Valor	Ganancia Promedio			
<i>Variables categóricas</i>					
Canal creación	Portal Web	37.88	Hora cita	9	0.63
Canal creación	Call Center	0.23	Hora creación cita	21	6.79
Día cita	16	5.78	Hora creación cita	4	5.84
Día cita	31	4.03	Hora creación cita	13	4.95
Día cita	22	3.54	Hora creación cita	12	3.65
Día cita	15	2.98	Hora creación cita	17	3.38
Día cita	7	2.83	Hora creación cita	20	3.10
Día cita	27	2.47	Hora creación cita	10	2.95
Día cita	4	2.42	Hora creación cita	8	2.77
Día cita	3	1.09	Hora creación cita	23	2.58
Día cita	23	1.06	Mes cita	5	4.93
Día cita	21	1.02	Mes cita	9	3.82
Día cita	29	0.85	Mes cita	12	3.75
Día cita	30	0.20	Mes cita	2	2.13
Día creación cita	11	5.58	Mes cita	7	1.16
Día creación cita	7	5.30	Mes cita	6	0.18
Día creación cita	23	2.86	Mes creación cita	4	4.77
Día creación cita	21	2.65	Mes creación cita	11	4.23
Día creación cita	13	2.45	Mes creación cita	10	3.44
Día creación cita	27	2.32	Mes creación cita	9	2.92
Día creación cita	22	0.41	Mes creación cita	3	2.13
Día semana cita	Martes	5.93	Plan de salud	Particular	7.91
Día semana cita	Jueves	5.08	Plan de salud	Isapre	3.90
Día semana cita	Lunes	3.84	Tiempo visita	PM	3.12
Día semana cita	Sábado	2.02	<i>Variables numéricas</i>		
Día semana creación cita	Lunes	5.16	Tasa inasistencia paciente		55.24
Día semana creación cita	Jueves	4.51	Demanda paciente		10.97
Día semana creación cita	Sábado	2.21	Tasa inasistencia médico		9.24
Día semana creación cita	Martes	1.13	Edad paciente		8.89
Duración	20	6.91	Inasistencias paciente		7.70
Hora cita	11	4.96	Antelación cita		5.74
Hora cita	10	3.92	Atenciones médico		4.10
Hora cita	15	3.84	Inasistencias médico		3.89
Hora cita	19	3.52	Demanda médico		3.76
Hora cita	16	1.94	Atenciones paciente		2.12
Hora cita	17	1.55			