UNIVERSIDAD DE CHILE
FACULTAD DE MEDICINA
ESCUELA DE POSTGRADO

EFECTOS DE LA PANDEMIA DE COVID-19 EN MODELOS DE PREDICCIÓN DE PACIENTES QUE NO SE PRESENTAN A CITAS PRESENCIALES Y TELECONSULTAS

NATALIA ALEJANDRA GILLIES RIVAS

TESIS PARA OPTAR AL GRADO DE MAGISTER EN INFORMÁTICA MÉDICA

Directora de Tesis: Prof. Jocelyn Dunstan

Co-Directores: Víctor Riquelme - Héctor Ramírez

2022

AGRADECIMIENTOS

Al Servicio de Salud Metropolitano Sur por creer en este proyecto y colaborar en su realización. Espero que el resultado les sea útil para continuar aportando a la salud pública.

A mi Directora de Tesis, Jocelyn Dunstan y Co-directores Victor Riquelme y Héctor Ramírez, por su dedicación, enseñanzas, apoyo incondicional y orientación, que ciertamente contribuyeron a lograr terminar este trabajo de la mejor manera posible

A mis compañeros del Magister en Informática Médica cohorte 2020, ya que sin su compañía nada habría sido lo mismo.

A Pablo, por su paciencia y amor incondicionales.

Finalmente, al FONDEF ID19I10271, “Soluciones tecnológicas, basadas en técnicas matemáticas avanzadas de aprendizaje de máquinas, para aumentar la eficiencia en la gestión hospitalaria” que ayudaron a financiar este trabajo.

TABLA DE CONTENIDOS

ÍNDICE DE TABLAS

ÍNDICE DE ILUSTRACIONES

RESUMEN

Problema: Los pacientes que no se presentan a sus citas ambulatorias (NSP) alcanzan un 23% a nivel global y un 27,8% en Latinoamérica. En Chile, corresponden a un 16,5% de las citas. El NSP genera una sobrecarga sobre los sistemas de salud, con impacto a nivel de costos, ganancias y utilización de recursos. La pandemia de COVID-19 ha hecho migrar las consultas desde una modalidad presencial a teleconsultas, lo que presenta desafíos a los modelos actuales de predicción de NSP.

Solución: Se entrenan modelos de aprendizaje de máquina que permitan estimar la probabilidad de NSP de pacientes en teleconsultas, en base a características de la persona, datos de la cita y datos del prestador. Se espera obtener un modelo para predecir la probabilidad de inasistencia del paciente a consultas generadas en contexto de pandemia de COVID-19 y estudiar la diferencia con los algoritmos entrenados en consultas pre-pandemia.

Método: Se utilizan datos de cuatro hospitales, entre 2018 y 2020, y se comparan según el tipo de atención y el NSP. Se entrenan modelos de aprendizaje de máquinas para predecir la probabilidad de que un paciente no asista a una cita en contexto de pandemia y se compararán las variables que tienen mayor influencia.

Variables y Métricas: Se consideran características de la persona (edad, sexo, previsión, pueblo originario, nacionalidad, historial de NSP, comuna), de la cita (fecha, hora, resultado, especialidad, tiempo entre agendamiento y cita, modalidad de atención, atención en pandemia, tipo de profesional) y del prestador (establecimiento y comuna). Como evaluación se utiliza el *F1-Score,* que combina precisión y sensibilidad.

Resultados: El algoritmo *XGBoost* en la predicción con doce meses previos a la cita es el algoritmo con mejores resultados, alcanzando un *F1-Score* de 0,32. Las variables con mayor importancia a la hora de predecir el NSP son los días entre el agendamiento y la cita y el histórico de NSP de los últimos 365 días. El que las citas sean realizadas por teleconsultas sólo tiene importancia en algunos modelos específicos, pero no genera cambios a nivel general.

ABSTRACT

Problem: Patients who do not show up for their outpatient appointments or no-show (NSP in Spanish) reached 23% globally and 27,8% in Latin America. In Chile, they reach 16,5% of the outpatient appointments. NSP generates an overload on health systems, impacting the costs, profit and use of resources. The COVID-19 pandemic has changed consultations from face-to-face to remote, presenting challenges to current NSP prediction models.

Solution: We seek to train machine learning models that allow estimating the probability of NPS of patients in teleconsultations, based on characteristics of the person, appointment and provider data. We expect to obtain a model to predict the probability of patient non-attendance at appointments generated in the context of the COVID-19 pandemic and to study the difference with the algorithms trained in pre-pandemic appointments.

Method: Data from four hospitals between 2018 and 2020 was used and compared according to the appointment type and NSP. Machine learning models were trained to predict the probability that a patient will not attend an appointment in the context of the pandemic and the variables that have the most significant influence will be compared.

Variables and Metrics: Characteristics of the patient (age, sex, insurance, migrant, nationality, history of no-show, commune), of the appointment (date, time, result, specialty, time between scheduling, appointment modality, pandemic attention, type of professional) and the provider (establishment and commune) were be considered. The F1-Score was used as an evaluation, which combines precision and recall.

Results: The XGBoost algorithm, in the prediction with twelve months prior to the appointment, is the algorithm with the best results, reaching an F1-Score of 0,32. The most important characteristics when predicting the NSP are the days between the scheduling and the appointment and the NSP history of the last 365 days. The fact that appointments are online is only important in some specific models but does not generate changes at a general level**.**

1. INTRODUCCIÓN

1.1. Antecedentes

1.1.1. Sistema de salud chileno

El sistema de salud chileno consta de dos sectores, público y privado. El primero, el Fondo Nacional de Salud (FONASA), cubre a 80% de la población a través del Sistema Nacional de Servicios de Salud (SNSS) y su red de 29 servicios de salud regionales, sumado al sistema municipal de atención primaria, con lo que cubren a alrededor de 70% de la población nacional. Un 3% adicional está cubierto por los Servicios de salud de las fuerzas armadas y el 7% restante son trabajadores independientes y sus familias que no cotizan a FONASA y que, en caso de necesidad, utilizan los servicios del sector público. El sector privado está constituido por las Instituciones de Salud Previsional (ISAPRE), que cubren aproximadamente a 17,5% de la población y proveen servicios a través de instalaciones tanto privadas como públicas. Un reducido sector de la población paga por la atención a la salud directamente de su bolsillo. Además de FONASA y de las ISAPRE, tres mutuales ofrecen cobertura exclusiva para accidentes de trabajo y enfermedades profesionales a los trabajadores afiliados (sin incluir a sus familias), los cuales representan cerca de 15% de la población. Estas mutuales prestan servicios dentro de sus propias instalaciones y, en caso de contar con capacidad ociosa, ofrecen atención a población no afiliada a cambio de un pago por servicio (1).

El sistema de salud en Chile presenta dos componentes, uno que apunta a la complejidad social, y el otro a la complejidad asistencial, entendiendo que esto forma parte de un continuo. En cada uno de ellos se realizan actividades insertas en el área social o en la red técnico asistencial, respectivamente. En este diseño, la Atención Primaria de Salud (APS) participa por su condición, tanto en la complejidad social como en la complejidad asistencial puesto que interactúa fuertemente con la comunidad a través de las acciones de prevención y promoción, interactuando incluso con otros sectores (ministerios, ONG, organizaciones sociales, consejos, municipios), y por otra parte corresponde a la principal puerta de entrada hacia la red técnico asistencial, permitiendo resolver un porcentaje

importante de la demanda o siendo el que origina el proceso de referencia contrarreferencia, hacia las especialidades médicas o hacia las hospitalizaciones. La complejidad técnico asistencial está orientada a resolver los problemas de salud que requieren de una mayor especificidad técnica y tecnológica, garantizando la resolución, asegurando la continuidad de la atención. Todas estas acciones además deben contemplar estrategias de prevención y promoción alineadas con prioridades sanitarias del país (2).

1.1.2. Sistema de salud pública

El SNSS, a través de los 29 Servicios de salud en todo el territorio, provee servicios ambulatorios y hospitalarios para los afiliados a FONASA. Estos se articulan en una red asistencial que permite entregar mejores condiciones de salud a sus integrantes, dependiendo del nivel de complejidad que se requiera. Para organizar de mejor manera y atender según necesidades se creó el sistema de niveles de atención, compuesto de la siguiente manera:

1. Nivel Primario: Se atiende en centros de salud familiar (CESFAM), centros de salud rural, consultorios generales urbanos, postas rurales y servicios de atención primaria de urgencia, dependientes de los municipios, y 105 hospitales de menor complejidad tipo 4 dependientes de SNSS, bajo la supervisión del Ministerio de Salud, el cual establece las normas técnicas de funcionamiento. En estos pueden existir desde un auxiliar rural hasta un equipo formado por médicos generales y técnicos, e incluso hoy en día, otros profesionales como odontólogos, kinesiólogos y psicólogos, dependiendo de la magnitud de la población de cobertura, y la cercanía a un centro asistencial de mayor complejidad. La estrategia de la atención primaria es otorgar cobertura de atención a mayor volumen poblacional, resolviendo enfermedades de menor complejidad, con un mayor impacto a nivel país. En este nivel se llevan a cabo los programas básicos de salud orientados a la prevención y promoción de salud. Se cuenta con medios simples de apoyo diagnóstico y un arsenal terapéutico establecido de acuerdo con las patologías básicas que se pueden solucionar a este nivel. Las actividades que se efectúan en este nivel son fundamentalmente: controles, consultas al policlínico, visitas domiciliarias, educación de grupos, vacunaciones y

alimentación complementaria (3). Como en este nivel no existe la atención por parte de especialistas es que se realizan derivaciones a establecimientos de atención secundaria o terciaria que puedan dar respuesta a esta demanda. La atención de salud primaria es gratuita para los afiliados a FONASA, ya que su forma de financiamiento es a través del pago de per cápita (por cada usuario inscrito en el establecimiento) y a través de programas ministeriales (4).

2. Nivel Secundario: Está formado por centros de referencia de salud, consultorios adosados a especialidades, además de hospitales de mediana complejidad tipo 3. Por lo general, involucran establecimientos hospitalarios que adosados tienen un centro de atención ambulatoria (consultorios adosados), quienes derivan a los pacientes por referencia, por lo cual existe mayor actividad profesional y de especialistas, con mayor proporción de elementos diagnósticos y terapéuticos (3).

3. Nivel Terciario: Compuesto por centros de diagnóstico y tratamiento, consultorios adosados a especialidades y hospitales de mayor complejidad tipo 1 y 2. Se caracterizan por su alta complejidad y baja cobertura. Están destinados a resolver los problemas que no se han podido solucionar a nivel primario ni secundario. Deben actuar como centros de referencia no solo para la derivación de pacientes de su propio servicio de salud, sino que son centros de derivación con carácter regional, supra regional y en ocasiones nacional. Esto no siempre se cumple, debido a la consulta espontánea que se da en los servicios de urgencia hospitalaria de los centros nivel 1 y 2. Sus recursos humanos son los más especializados y tienen el apoyo diagnóstico y terapéutico de mayor complejidad. Muchas veces sus acciones se entrelazan con las del nivel secundario, ya que tienen también la responsabilidad de solucionar problemas de frecuencia intermedia por la alta demanda (3).

Los beneficiarios de FONASA tienen acceso a dos modalidades de atención: la Modalidad de Atención Institucional (MAI) y la Modalidad de Libre Elección (MLE). La primera comprende la atención que brindan de forma directa las instituciones públicas de salud, con cierta limitación en la capacidad de elección del

prestador, como son la comuna de residencia para los usuarios de la atención primaria. Al momento de recibir la atención, los usuarios deben realizar copagos que van de 10% a 20% del precio del servicio fijado por FONASA de acuerdo con su nivel de ingresos, excepto los más pobres, los mayores de 60 años y los portadores de algunas patologías específicas.

FONASA agrupa a sus beneficiarios según su ingreso en los Tramos A, B, C y D. Esto es fijado cada año según las variaciones del ingreso mínimo. Es importante saber que sólo las personas con tramo B, C y D pueden atenderse a través de bonos en la Red de libre elección en prestadores en convenio con FONASA. Al año 2022 la distribución de los tramos se resume en la siguiente tabla (5):

Tramo	Beneficiarios	Bonificación y copago
Tramo A	Personas carentes de recursos y personas migrantes. Causantes de subsidio familiar (Ley 18.020).	Bonificación del 100% en las atenciones de salud en la Red Pública (MAI)
Tramo B	Personas que perciben un ingreso imponible mensual menor o igual a $350.000**.-**	Bonificación del 100% en las atenciones de salud en la Red Pública (MAI) y acceso a compra de bonos en establecimientos privados en convenio con Fonasa (MLE)
Tramo C	Personas que perciben un ingreso imponible mensual mayor a $350.000.- y menor o igual a $511.000.- Nota: Con 3 o más cargas familiares pasará a Tramo B.	Bonificación del 90% en las atenciones de salud en la Red Pública (MAI) y acceso a compra de bonos en establecimientos privados en convenio con Fonasa (MLE)
Tramo D	Personas que perciben un ingreso imponible mensual mayor a $511.000.- Nota: Con 3 o más cargas familiares pasará a Tramo C.	Bonificación del 80% en las atenciones de salud en la Red Pública (MAI) y acceso a compra de bonos en establecimientos privados en convenio con Fonasa (MLE)

Tabla 1: Clasificación de beneficiarios de FONASA a según su tramo (5).

La modalidad institucional suele ser usada por los ciudadanos de menores recursos. Los beneficiarios de FONASA, de acuerdo con los tramos indicados anteriormente, pueden elegir la MLE y, mediando el copago equivalente a la diferencia entre el precio fijado por los prestadores para cada prestación y la cantidad fija aportada por FONASA, puede elegir el prestador dentro del sector

privado. A esta modalidad suelen recurrir los beneficiarios del FONASA de mayores ingresos (1)

Los servicios de salud pública (como vacunas y control de la tuberculosis) se proveen a toda la población sin restricción o discriminación. En el caso de la tuberculosis, la entrega de los medicamentos es gratuita y se debe hacer en los establecimientos del sector público bajo supervisión. Las vacunas son entregadas por el Estado y su aplicación en el sector público es gratuita para toda la población, aunque en el sector privado la aplicación puede tener un costo (1).

1.1.3. Proceso de atención ambulatoria

La Consulta Ambulatoria en Red es un proceso complejo en el cual participan directamente una serie de profesionales y técnicos de los establecimientos de salud y que requiere para su normal operación del apoyo de otros procesos que proveen de los insumos necesarios y de las regulaciones que deben ser consideradas. La principal puerta de entrada al sistema corresponde a la atención primaria, que es donde se resuelven la mayor cantidad de los problemas de salud que presenta la población (2)

En la Ilustración 1 se grafica un flujo estándar de derivación de interconsulta en la cual el paciente es evaluado en la APS, priorizado y luego derivado a la atención secundaria, donde ingresa a lista de espera y es atendido, para luego retornar a la atención primaria.

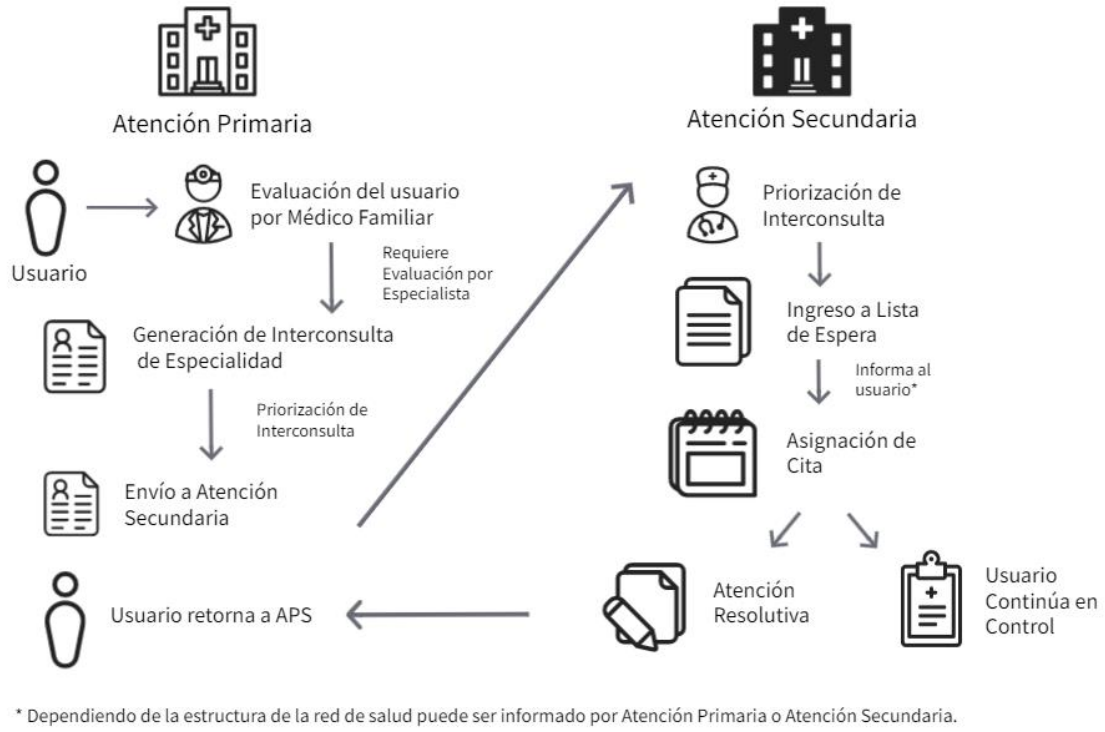


Ilustración 1: Representación gráfica proceso de referencia y contra referencia. Elaboración propia

El principal punto de entrada al proceso de atención de consulta ambulatoria se produce en un establecimiento de APS. Esta fase se inicia una vez que, evaluado el paciente en la atención primaria, el profesional médico decide su derivación a un especialista. Este requerimiento, antes de ser enviado al especialista definido, es revisado, evaluado y priorizado por un comité de coordinación de red, que es el encargado de determinar si éstas se realizan en forma adecuada y con pertinencia según los protocolos definidos en cada red, generalmente por contralores que se encuentran en los establecimientos de salud primaria (2). Si es pertinente la evaluación por un especialista, esta necesidad puede ser resuelta de diferentes formas:

- La atención por un especialista en el mismo establecimiento de APS, a través de las consultorías (capacitación gestionada),

- En general lo más frecuente es la derivación a otro establecimiento de la red que disponga de la especialidad requerida, ya sea hospitales, centros de referencia de salud o centros de diagnóstico y tratamiento.

- A través del programa de resolución de especialidades, en el cual los establecimientos de APS determinan la demanda de consultas de especialidad y compran considerando la integralidad y resolutividad, además de buscar mantener la continuidad de la atención.

La siguiente fase es la derivación de la interconsulta hacia el establecimiento de destino, donde ésta es nuevamente priorizada de acuerdo con las características de la derivación. Si se encuentra correcta, esta ingresa al registro de lista de espera que le corresponda. Una vez que se habilita un cupo prosigue la asignación de la hora y citación del paciente al establecimiento y al especialista que corresponda. Para la asignación de las horas, se deben considerar los tiempos de espera, además de los criterios de prioridad sanitarios que se hayan establecido. Al menos deben considerarse el diagnóstico, sexo y edad del paciente, las condiciones de ruralidad y marginalidad, y la concomitancia de otros problemas de salud. Al momento de contactar al paciente se le debe informar la fecha y hora de la atención, la hora y el lugar en que se debe presentar, las indicaciones específicas que se requieran para esa atención y los documentos que debe llevar. En ese mismo momento se le deberá indicar al paciente que en caso de no poder asistir en la fecha definida dé aviso al establecimiento correspondiente para poder reasignar esa hora. El contacto con el paciente puede ser realizado, dependiendo de cómo se coordine esta red, por el establecimiento hospitalario o por la atención primaria. Esto puede ser directamente o a través de llamadas telefónicas, cartas, SMS u otro sistema. Todos estos sistemas diferidos deben acompañarse de un registro de la acción que permita su posterior monitoreo (2).

1.1.4. Pacientes que no se presentan

Un paciente que no se presenta, o *no-show patient* (NSP) se define como un paciente que falla en la asistencia a una cita agendada sin notificar previamente al prestador de salud (6). Estos alcanzan un 23% a nivel global y un 27,8% en Latinoamérica (7). En Chile, los pacientes que no se presentan a citas de atención ambulatoria corresponden a un 16,5% de las citas, con un rango entre regiones que varía entre un 8,8 a 20,2%. La inasistencia es mayor en las especialidades de dermatología, geriatría y nutrición (8).

La ausencia de pacientes a sus citas impacta a la salud pública de múltiples maneras. Primero, generando una infra utilización de recursos de los establecimientos, lo que extiende tiempos de espera de otros pacientes, además de un impacto económico que limita contrataciones y mejoras en infraestructura (9). Adicionalmente, los pacientes que mantienen sus citas perciben efectos negativos, como insatisfacción y altos tiempos de espera que dan la percepción de un decaimiento en la calidad de servicio (7).

Sin embargo, existe consenso que el NSP no es un fenómeno aleatorio, y que existen variables que influyen en la probabilidad que un paciente falte a su cita. Se destacan los principales factores influyentes:

- Especialidad: se identifican diferencias en el comportamiento del NSP en pacientes de distintas especialidades, siendo las con mayor promedio identificadas fisioterapia (57,3%), endocrinología (36%) y cardiología (30%)(10,11)
- Edad: es un factor significativo, de tal forma que es inversamente proporcional a la probabilidad de NSP, en donde los adultos jóvenes son quienes tienen más probabilidad de perder su cita, mientras que en pediatría la probabilidad de NSP se ve aumentada con la edad del niño (11–16)
- Nivel socioeconómico: Se identifica una correlación entre menor nivel socioeconómico y una mayor probabilidad de NSP (13,15).
- Tiempo entre agendamiento y cita: el tiempo de espera del paciente, desde a fecha en que es citado a la fecha en que realmente se agenda a cita, es uno de los factores predictores más importantes según la literatura, siendo a mayor cantidad de días que pasan entre el agendamiento y la cita, mayor es la probabilidad de no presentarse (12,14,17)
- Historial de NSP previo: también ha sido considerado como un fuerte predictor de NSP, siendo aquellos pacientes con historial previo de no presentarse a sus citas los con mayor probabilidad de no presentarse a sus próximas citas (12–14,16,17)

- Tipo de Seguro: se encontró una correlación entre los pacientes con seguro privado y la menor probabilidad de NSP (18,19)
- Distancia al establecimiento: se ha encontrado una relación entre una mayor distancia entre el establecimiento y la residencia del paciente y una mayor probabilidad de NSP (14,16–19)

Al ser posible determinar variables relacionadas con el fenómeno del NSP es que se han generado múltiples algoritmos de aprendizaje de máquinas que permiten predecir con mayor o menor certeza la probabilidad de que un paciente se presente o no a su cita. Una revisión sistemática (9) analizó un total de 50 artículos publicados en los últimos 10 años, de los cuales 40 eran relacionados con algoritmos de predicción de NSP, buscando determinar la correlación entre variables y buscar una única métrica que permita predecir el comportamiento de los pacientes frente a una consulta médica.

1.1.5. Telemedicina

La telemedicina se define como el intercambio de información médica entre dos actores que no están ubicados en el mismo espacio, que pueden ser médico/paciente, o médico/médico a través de algún medio de comunicación electrónico, con el objeto de mejorar el estado de salud de un paciente. Incluye una variedad cada vez mayor de aplicaciones y servicios que utilizan videoconferencias, correo electrónico, teléfonos inteligentes, comunicaciones inalámbricas y otras formas de tecnología de las telecomunicaciones (25). Esta estrategia de atención se implementó formalmente en Chile en el año 2018 con el surgimiento de la estrategia de Hospital Digital. Hospital Digital se trata de “un nuevo modelo de atención en salud, centrado en el paciente, y que aprovecha el potencial de las tecnologías para acercar la atención a las personas, instalando una alternativa al modelo tradicional” (21). Esta estrategia fue la que permitió que se iniciaran y formalizaran las estrategias de salud digital que se habían estado gestando en el país para permitir un mayor acceso a atenciones de salud. Sin embargo, se encontró con variados contratiempos que hicieron que la estrategia no prosperara con el impacto esperado (27). A la fecha, se encuentra implementada la atención de forma telemática para consultas de especialidad como dermatología, patología oral y nefrología, entre otras.

Una teleconsulta corresponde a una consulta a la distancia con intercambio de información realizada a través de tecnologías de la información y telecomunicaciones en modalidad sincrónica (tiempo real) entre un paciente y uno o más miembros del equipo de salud. Para efectos de esta tesis se considerará como teleconsulta aquellas realizadas vía telefónica o videollamada entre paciente y un miembro del equipo de salud. Las teleconsultas tuvieron un auge durante la pandemia de COVID-19, debido a que fue la forma más sencilla y rápida con la que los profesionales podían establecer contacto con los pacientes y efectuar el seguimiento de sus patologías sin tener que realizar una consulta presencial.

1.1.6. Aprendizaje de máquinas

El aprendizaje de máquinas, o *machine learning* en inglés, es un área dentro de la inteligencia artificial que se encarga de aprender de los datos sin que el programador o programadora tenga que explícitamente definir las reglas. Técnicas basadas en *machine learning* han sido aplicadas satisfactoriamente en distintos campos, desde el reconocimiento de patrones, visión computacional, ingeniería en naves espaciales, finanzas, entretenimiento, y aplicaciones en medicina y biomedicina (28). Cuando se quiere predecir un fenómeno del cual tenemos historia de su comportamiento, es necesario tener un conjunto de datos etiquetados de la forma $\{(X_i, y_i)\}_{i=1}^{N}$ donde: X_i son las variables tales como edad, NSP en los últimos 30 días, comuna del establecimiento, entre otras. N es el número de ejemplos de nuestro conjunto e y_i es si el paciente se presentó o no a la cita. El objetivo es usar el conjunto de datos para producir un modelo que sea capaz de recibir un conjunto de variables X y entregar una probabilidad $(\hat{p})$ que nos permita deducir la etiqueta y. A este tipo de modelos se les llama supervisados (24).

1.1.7. Algoritmos de predicción de NSP

Múltiples estudios a nivel internacional han intentado predecir el comportamiento de los pacientes frente a sus citas médicas (7). Sin embargo, no todos los estudios son comparables con este trabajo ya que difieren en ciertas características basales, como son el tamaño de las muestras utilizadas y el porcentaje de NSP. Esto ocasiona que los resultados de un estudio no puedan ser extrapolados entre sí, debido a que ciertas métricas son sensibles al porcentaje de

NSP, por ejemplo. Para efectos del análisis consideraremos estudios que tengan características similares a las que se comprenden en este trabajo, ya sea por su porcentaje de NSP, tamaño muestral u origen de los datos ya que en todos los casos corresponden a estudios realizados en centros de especialidad.

Artículo	N° Pacientes	N° Citas	Meses	% NSP	Modelo	Medida de Desempeño
Daggy et al, 2010	5.446	32.394	36	15.2	Regresión logística	0,82 (AUC)
Huang y Hanauer, 2014	7.998	104.799	120	11.2	Regresión logística	82,1 (Accuracy)
Huang y Hanauer, 2016	7.291	93.206	120	17	Regresión logística	0,706 (AUC)
Goffman et al, 2017	-	21.551.572	48	13.87	Regresión logística	0,713 (AUC)
Mohammadi et al, 2018	73.811	73.811	27	16.7	Regresión logística, Redes neuronales	0,86 (AUC)
Elvira et al, 2018	323.664	2.234.119	20	10.6	*Gradient boosting*	0,74 (AUC)
Lin et al, 2019		2.000.000	36	18	*Bayesian lasso*	0,70 – 0,92 (AUC)

Tabla 2: Tabla resumen de estudios comparables. Extracto tabla Carreras-García D et al (4)

De la Tabla 2 es posible identificar que la medida más utilizada es el área *AUC*, que corresponde al área bajo la curva de la curva *ROC*. La definición de esta y otras métricas se encuentra detallada en la sección 5.2.

1.1.8. Pandemia COVID-19 y su impacto

La pandemia de COVID-19 se declaró a nivel mundial el 11 de marzo de 2020 (25) y tuvo su primer caso en Chile el 3 de marzo del mismo año (26). A la fecha el COVID-19 ha afectado a todo el mundo, superando los 480 millones de contagios, con 6.1 millones de muertes a nivel mundial a marzo del 2022. En Chile, el total de pacientes contagiados asciende a 3.443.018, mientras que las muertes alcanzan un total de 44.940 personas al 26 de marzo del 2022 (27). La alta contagiosidad del virus de transmisión vía gotas o aerosoles a través de estornudos y tos, entre otros (28,29), ha obligado a la sociedad a tomar medidas de resguardo adicionales para evitar la propagación. Entre éstas se encuentra el lavado frecuente de manos, el uso de mascarilla y el distanciamiento físico, con lo que se han definido aforos limitados para cada lugar y se ha obligado a restringir los espacios cerrados, como cines, pubs, colegios, entre otros. También se han instaurado otros mecanismos como son las cuarentenas obligatorias y el establecimiento de toques de queda, entre otras medidas que buscan disminuir el desplazamiento de las personas (30). Se ha dado énfasis en no salir de casa y por lo mismo muchas empresas han pasado a tener una modalidad de trabajo remota y servicios como el despacho a domicilio se han vuelto esenciales.

Dentro de las consecuencias de la pandemia nos encontramos con varios efectos en relación con la salud pública. Se establecieron medidas a nivel central, como la suspensión de un grupo de Garantías Explícitas en Salud (GES) para reorientar los esfuerzos al control de la pandemia (31). En muchos establecimientos se debieron cancelar citas médicas, por diversos motivos. Uno de ellos corresponde a la necesidad de establecer turnos rotativos en el personal de salud para prevenir que, en el caso de existir un contagiado, el servicio dejase de operar por cuarentenas de la totalidad del personal. Con esto, la cantidad de profesionales disponibles para atención a público se redujo, con muchos profesionales que no podrían ejercer sus labores de forma remota. Otro motivo tiene que ver con la reasignación de personal médico a unidades de cuidados críticos, lo que hizo que muchos servicios quedaran desprovistos de personal para realizar las atenciones. Adicionalmente, y desde el punto de vista de los pacientes, estos comenzaron a faltar a las citas médicas para evitar la exposición a ambientes con aglomeraciones que propiciaran el contagio.

Todo esto en su conjunto generó modificaciones en el quehacer de las unidades de atención ambulatorias. En concreto, obligó a implementar diversas medidas sanitarias para evitar contagios, dentro de ellas, la suspensión de múltiples agendas y cupos en establecimientos públicos y privados para cumplir con los aforos indicados, favorecer el distanciamiento físico y disminuir las aglomeraciones, y la implementación de citas de forma remota para mantener la atención respetando las normas sanitarias y la seguridad tanto para pacientes como para funcionarios (30). Según datos de I-Med en junio de 2020 se generaron un 62% menos de consultas ambulatorias que en la misma fecha del 2019 (32). La Fundación Politopedia calculó que 2,3 millones de consultas podrían postergarse, además de 125 mil cirugías y 181 mil casos GES, a lo que se suman todas las patologías que surEne por el encierro, como los problemas de salud mental (32).

1.2. Problema

Los pacientes que no se presentan a sus consultas presenciales representan en Chile en promedio un 16,5%, con un rango entre regiones desde 8,8 a 20,2% (8). Esto genera una carga sobre los sistemas de salud, con impactos significativos a nivel de costos, ganancias y utilización de recursos. El impacto afecta tanto a los prestadores de servicios sanitarios como a los mismos pacientes que pierden su cita. Algunos de estos efectos corresponden a contar con un proceso terapéutico descontinuado, mayores tiempos de espera para la atención y un aumento de ingresos de pacientes a través de servicios de urgencia. Estos ingresos inesperados llevan a un aumento en los gastos, ya que el coste de las urgencias es mayor y entrega pocos servicios de tipo preventivo. Sin ir más lejos, causa una disminución en el acceso a atención de otros pacientes, lo que puede llevar a una insatisfacción entre los mismos y los prestadores de salud. Adicionalmente, el que los pacientes no se presenten genera una pérdida de ganancias a los establecimientos sanitarios. Un estudio en un laboratorio vascular identificó que un NSP de un 12% puede costarle al laboratorio unos USD$89.107 anuales (unos 64.157.000 de pesos chilenos) (6). Otro estudio en una clínica de veteranos que incluyó a 10 sucursales de la misma estimó una pérdida promedio de USD$14.58 millones en promedio por año por concepto de pacientes que no se presentan (10).

Si bien existen múltiples ejemplos en la literatura de algoritmos de aprendizaje de máquinas que permiten predecir y caracterizar los motivos por los que un paciente no se presenta a su cita (9), éstos han sido entrenados en ambientes y con variables pensadas en los que la atención ocurre de forma presencial. Debido a la pandemia de COVID-19 y a las restricciones que se han debido imponer para disminuir los contagios a través de reducción de aforos y cuarentenas, es que se han aumentado las consultas de forma telemática, por lo que es un desafío interesante el identificar si las características que influyen en el NSP de forma presencial son los mismos que influyen en el NSP de en época de pandemia.

Buscando en la literatura, existe un estudio sobre los factores que influyen en la asistencia a controles post-operatorios vía telemedicina entre los que se concluye que los pacientes atendidos por teleconsulta tienen el doble de probabilidad de no presentarse, en comparación a aquellos que se agendan en una cita tradicional (33). Esto nos entrega algunas luces de las diferencias que podría tener el NSP una vez que se comparen ambos contextos. No se han entrenado algoritmos de aprendizaje de máquinas para predecir ausencia a consultas remotas en el contexto de pandemia. Los resultados obtenidos en este contexto podrían ser muy distintos a lo que se conoce en una época de normalidad relativa, es decir, sin pandemia de COVID-19 en curso.

1.3. Motivación

La principal motivación para este trabajo tiene que ver con mejorar el aprovechamiento de los recursos en la salud pública, evitando la pérdida de los mismo en pacientes que no se presentan a sus citas y que generan que los profesionales clínicos pierdan tiempo productivo.

Si bien la escala de tiempo de esta tesis no permite contemplar una implementación ni evaluación de la solución a elaborar, se espera que contribuya a mitigar el NSP en el Servicio de Salud Metropolitano Sur (SSMS) mediante la entrega de un algoritmo que al ser ejecutado sobre las citas agendadas permita anteponerse al comportamiento de los pacientes, en este caso, a que éstos puedan

no presentarse a su cita. El trabajo se orienta a evitar la pérdida de horas de atención clínica en el sistema de salud público, mediante el entrenamiento de un algoritmo que permita identificar previamente a aquellos pacientes que podrían no presentarse a su cita médica. Con esto, es posible realizar acciones tendientes a evitar las inasistencias, con estrategias de contactabilidad de pacientes mediante mensajes de texto o llamados telefónicos, o la reorganización de las agendas médicas para asignar sobrecupos de acuerdo con los pacientes probablemente inasistentes. Esto tiene un impacto en la salud pública y, además de evitar los problemas ya antes mencionados, corresponde a un aporte en mejorar la oportunidad y el acceso a atenciones de salud en un sistema que se encuentra al límite de sus capacidades.

Adicionalmente, se permite modernizar el trabajo dentro de los servicios públicos, inyectando tecnologías nuevas que permiten mejorar la eficiencia del trabajo y llevar a la organización a una nueva etapa en su desarrollo tecnológico. Finalmente, este trabajo pretende ser un aporte a la transformación digital de este organismo.

2. HIPÓTESIS

Debido a la pandemia existe un cambio en el comportamiento de las personas que no se presentan a su cita, lo cual podemos evidenciar en las variables que utilizamos para explicar el fenómeno.

3. OBJETIVO

1.4. Objetivo general

Analizar las características relevantes y los resultados de modelos de aprendizaje automático para identificar si existe un cambio de comportamiento en las personas que no se presentan a sus citas en el contexto de pandemia de COVID-19 con respecto a las personas que no se presentan a citas presenciales.

1.5. Objetivos específicos

1. Describir cómo se caracteriza el NSP de pacientes en época de pre-pandemia, pandemia vía presencial y pandemia vía teleconsulta.
2. Identificar las diferencias entre atenciones realizadas de forma presencial y aquellas realizadas por teleconsulta en pandemia.
3. Diseñar y construir modelos de aprendizaje de máquinas que permitan predecir el NSP en tiempos pre-pandemia y en consultas en época de pandemia.
4. Comparar las características que determinan el NSP prepandemia y pandemia.

4. MATERIAL Y MÉTODO

Para el desarrollo de este trabajo se utilizaron datos anonimizados extraídos, provenientes de 4 hospitales del SSMS, que comprenden los años 2018 a 2020. Para la obtención de estos datos se solicitó acceso a través de solicitud al Comité Ético Científico del SSMS donde se expuso el proyecto y se entregó la documentación requerida de acuerdo con el protocolo de la institución. Luego de revisados los antecedentes se dictaminó la aprobación del estudio vía Memorándum N°118 el 7 de julio de 2021.

Se utilizó información administrativa de pacientes, con datos relacionados al paciente, a la cita y al prestador de salud. No se utilizó ningún tipo de información personal de los pacientes y éstos no son identificables, solo se asocia la información determinante que corresponda a la atención de cada paciente de forma anónima. No fueron considerados en este estudio datos como nombres, identificadores personales, datos de contactabilidad como direcciones o teléfonos, fecha de nacimiento, relaciones familiares, nivel socioeconómico, nivel educacional ni estado civil. El motivo de esta decisión radica en la falta de acceso unificado a los datos clínicos de pacientes, debido a la falta de implementación de ficha clínica electrónica de forma transversal en la red del servicio sur.

Para la obtención de los datos se consideró la información almacenada en el Registro Clínico Electrónico TrakCare, de la empresa Intersystems, a cuyo espejo

de base de datos cuenta con acceso personal de la Dirección del SSMS, concretamente la Unidad de Ciencia de Datos del Departamento de Gestión de Tecnologías de la Información y Comunicación.

Se puede categorizar los datos entregados en tres categorías: paciente, cita y establecimiento. El detalle de cada una de las variables se encuentra en la Tabla 3. Adicionalmente se generaron variables calculadas con el fin de capturar temporalidad las cuales se detallan en la Tabla 4.

Tipo de Variable	Variable	Definición
Paciente	Edad	Número entero, corresponde al número de años desde el nacimiento del paciente al momento de la extracción de la información.
	Sexo	Categórico que detalla el sexo de un paciente.
	Pueblo Originario	Pueblo originario que el paciente declara.
	Nacionalidad	País de nacimiento del paciente.
	Previsión	Indica el régimen previsional de salud al que se encuentra adherido el paciente, FONASA o ISAPRE
	Comuna	Comuna de residencia del paciente
Cita	Fecha de la cita	Fecha en que fue agendada la cita, en formato año mes día.
	Hora de la cita	Hora en que fue agendada la cita, en formato hora minuto.
	Resultado de la cita	El resultado del evento. Una cita pasa de estar en estado agendada (estado inicial) a un estado final (atendido, no atendido, cancelada, etc)
	Especialidad	Especialidad del profesional que realiza la cita
	Atención en Pandemia	Si la cita fue o no realizada en contexto de pandemia de COVID-19. Se considera como fecha de inicio de la pandemia el día 11 de marzo del 2020 de acuerdo con lo establecido por la Organización Mundial de la Salud (18).
	Tipo de profesional	Profesión de quien realiza la consulta.
	Prestación	Prestación asignada a la consulta.
	Descripción	Descripción en texto libre que sirve para indicar alguna observación realizada a la consulta agendada.
Establecimiento	Establecimiento	Institución donde se realiza la atención al paciente.
	Comuna	Comuna de la institución que realiza la atención

Tabla 3: Detalle variables entregadas por el SSMS

Variable	Definición
Target	Valor binario que representa si la persona asistió (0) o no asistió (1) a su cita, esto se identificará como etiqueta del modelo supervisado.
Día de la Semana	Obtenido desde la fecha, corresponde día de la semana.
Mes	Obtenido desde la fecha, corresponde al número de mes del año.
Día del Año	Obtenido desde la fecha, corresponde día del año.
Día del Mes	Obtenido desde la fecha, corresponde día del mes.
Minuto del día	Corresponde al minuto del día, sacado desde la hora de la cita.
Está en la misma comuna	Relación entre la comuna del establecimiento y la comuna del paciente. Marca si el paciente vive o no en la misma comuna del establecimiento de la atención.
Atención a distancia	Filtrado manual, extraído desde la prestación y la descripción de la cita, de citas que refieren haber sido realizadas “a distancia”. En éstas no se especifica el método mediante el cual se realizaron dichas citas y no fue posible garantizar si se habían realizado de forma telefónica, por lo que se dejaron clasificadas de forma independiente.
Atención videollamada	Filtrado manual de citas, extraído desde la prestación y la descripción, que refieren haber sido realizadas a través de videollamada.
Atención telefónica	Filtrado manual de citas, extraído desde la prestación y la descripción, que refieren haber sido realizadas por vía telefónica.
Citas Previas 30, 60, 90, 120 y 365D	Número total de citas previas por los periodos de 30, 60, 90, 120 y 365 días anteriores a la fecha de la cita.
NSP 30, 60, 90, 120 y 365D	Porcentaje de NSP previo por los periodos de 30, 60, 90, 120 y 365 días anteriores a la fecha de la cita.
Dif. Fecha Cita	Corresponde a la diferencia en días de la fecha de agendamiento con la fecha de la cita.

Tabla 4: Detalle variables calculadas

Adicionalmente, se filtraron las citas canceladas o no atendidas por condiciones distintas a no se presentó, además de unificar tipo de profesional.

4.1. Análisis descriptivo

El conjunto de datos se compone por 3.787.110 citas, con 598.627 pacientes únicos. En la Tabla 5 se encuentran los datos anualizados, donde se observa una disminución en todas las variables para el año de la pandemia. En total, la base cuenta con un promedio de NSP de un 12,6%, lo cual se encuentra en el rango de lo descrito dentro de la literatura nacional, donde se indica que el NSP en el año 2005 bordeaba el 13% y al 2010 alcanzaba un 16% (8).

	2018	2019	2020	Total
N° CItas	1.452.117	1.437.987	897.006	3.787.110
N° Pacientes	219.676	212.329	166.622	598.627
% NSP	13,083%	13,839%	10,899%	12,607%

Tabla 5: Distribución global del set de datos

Al realizar un análisis mensual acorde a la Ilustración 2, se observa una fuerte correlación entre el número de citas y el porcentaje de NSP (p=0,76). Se destaca que el NSP tiene un alza marcada en octubre y noviembre 2019, donde alcanza el 17%, lo cual puede ser atribuido fenómenos sociales ocurridos en Chile durante dicho periodo. Por otro lado, se evidencia una baja importante tanto de las citas como del NSP durante el 2020, alcanzando su mínimo con un 6,31% en abril, en pleno inicio de pandemia, el cual se va recuperando y cierra en diciembre con un 12,18%, alcanzando el promedio de NSP de la muestra.

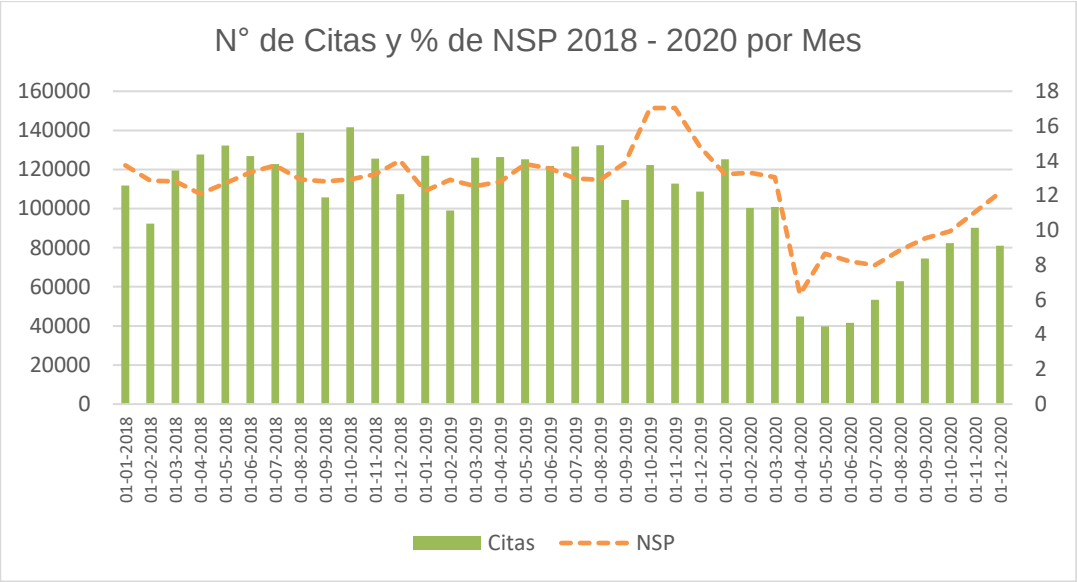


Ilustración 2: Número de citas y porcentaje de NSP 2018-2020 por mes

4.2. Análisis de variables numéricas

En la Tabla 6 se evidencia la distribución de las variables numéricas de forma general en el conjunto de datos. Para efectos de visualización de la información se realizaron ciertos ajustes, como son el quitar el 1% superior en las variables Edad y Citas Previas, y el 1% inferior en la variable Diferencia entre días de agendamiento y cita. Esto se debe a que existen errores en el registro de estas variables donde, por ejemplo, existen gran cantidad de pacientes sobre los 120 años o pacientes con casi 300 citas previas.

En la distribución de la variable edad se aprecia que las atenciones se concentran entre los pacientes que tienen entre 0 a 20 y luego 50 a 80 años. En lo que respecta a las variables asociadas a citas previas se puede apreciar que el valor más común es cero, es decir pacientes sin citas previas. Si bien en todos los análisis de tiempo el valor común sigue siendo 0, este disminuye paulatinamente siendo los pacientes con 1 cita el segundo caso más común.

En relación con el Histórico de NSP, independiente del periodo de análisis, siempre prevalecen aquellos pacientes que no tienen NSP. Hay que recordar que esta medida considera el total de citas a las que el paciente no se presenta sobre las citas totales del paciente, por lo que sus valores se distribuyen con valores entre 0 y 1. Si bien al analizar el NSP de 30 días existen pocos registros, en la medida

que aumenta el periodo de análisis van aumentando todos los rangos de NSP. Por último y en relación con la variable que mide la diferencia entre los días del agendamiento y la cita se destaca algunos registros con agendamiento positivo, es decir, citas que fueron agendadas posterior a su fecha efectiva, así como que la mayoría de las citas aparecen como agendadas con cero días de diferencia, lo cual según el esquema de flujo de agendamiento que se detalle en la sección 1.1.3 carece de sentido.

El análisis descriptivo de las variables numéricas por año se encuentra en la Tabla 7, donde se calcularon la completitud de la variable (el porcentaje de datos no nulos), su promedio, desviación estándar, valor mínimo, valor máximo, percentiles y el valor más común (VMC). Se puede identificar que no existen mayores diferencias entre los 3 años, a excepción de tres de las variables, las cuales son: el número de citas previas, el NSP previo y los días de diferencia entre el agendamiento y cita, donde todos los valores disminuyen el año 2020. Esto se evidencia tanto en los valores promedio como en los valores máximos registrados.

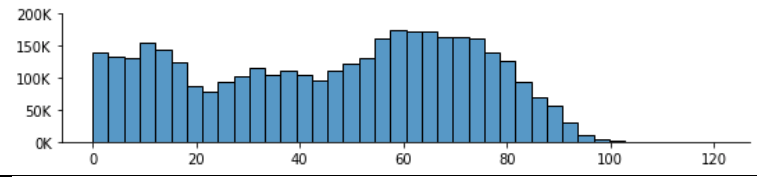
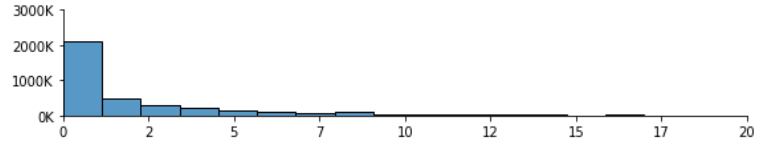
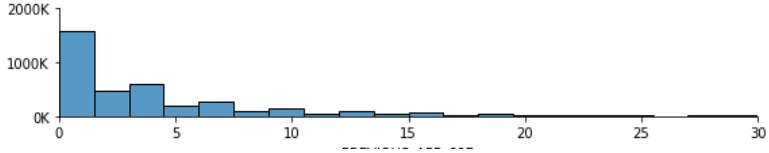
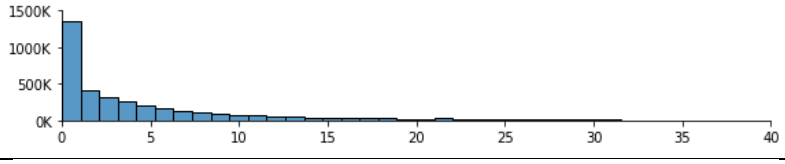
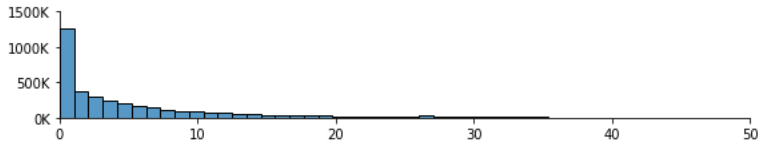
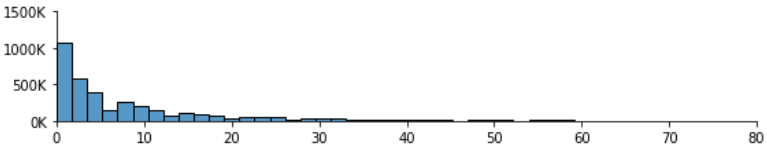
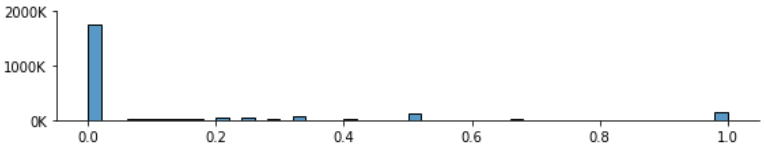
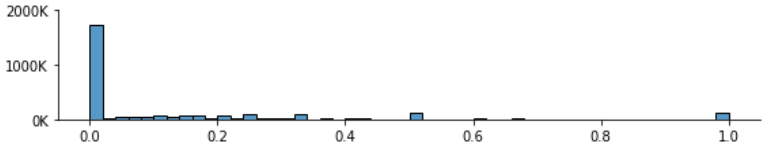
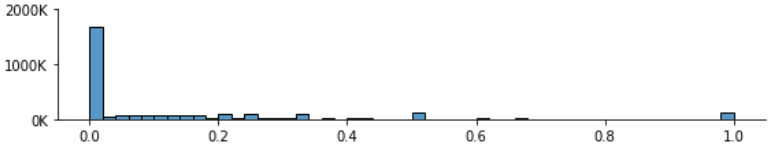
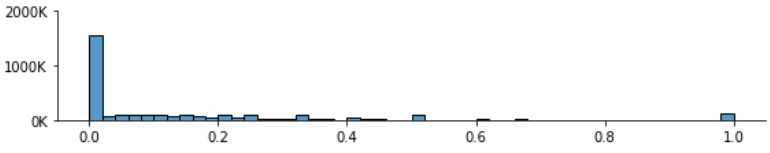
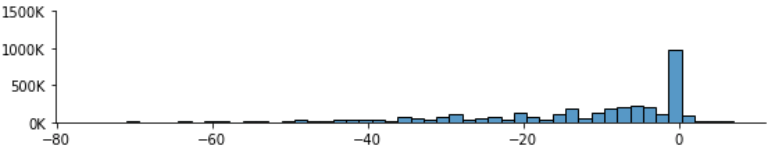
	Distribución
Edad	
Citas Previas 30D	
Citas Previas 60D	
Citas Previas 90D	
Citas Previas 120D	
Citas Previas 365D	
NSP 30D	
NSP 90D	
NSP 120D	
NSP 365D	
Dif. Fecha Cita	

Tabla 6: Histogramas de variables numéricas

Variable	Año	Completitud	Promedio	DS	Min	Máx	P25	P50	P75	P99	VMC
Edad	2018	1	46,30	26,35	2	121	21	50	69	92	3
	2019	1	46,59	25,82	0	115	24	51	68	91	2
	2020	1	45,43	25,13	0	121	24	49	66	90	1
Citas previas 30D	2018	1	2,55	3,87	0	83	0	1	3	18	0
	2019	1	2,59	3,85	0	59	0	1	3	18	0
	2020	1	2,37	3,68	0	61	0	1	3	17	0
Citas previas 60D	2018	1	4,48	6,53	0	110	1	2	6	32	0
	2019	1	4,60	6,66	0	114	1	2	6	32	0
	2020	1	4,02	6,24	0	110	0	2	5	30	0
Citas previas 90D	2018	1	6,02	8,83	0	155	1	3	7	43	0
	2019	1	6,22	9,06	0	142	1	3	8	44	0
	2020	1	5,18	8,25	0	134	0	2	6	41	0
Citas previas 120D	2018	1	7,31	10,85	0	161	1	4	9	53	0
	2019	1	7,57	11,18	0	158	1	4	9	55	0
	2020	1	6,10	9,98	0	135	1	3	7	50	0
Citas previas 365D	2018	1	11,01	17,71	0	247	1	5	13	88	0
	2019	1	11,45	18,42	0	333	1	5	14	91	0
	2020	1	9,56	16,69	0	247	1	4	11	83	0
NSP 30D	2018	0,65	0,14	0,27	0	1	0	0	0,14	1	0
	2019	0,66	0,14	0,28	0	1	0	0	0,17	1	0
	2020	0,62	0,11	0,25	0	1	0	0	0	1	0
NSP 60D	2018	0,76	0,14	0,26	0	1	0	0	0,20	1	0
	2019	0,76	0,15	0,26	0	1	0	0	0,20	1	0
	2020	0,71	0,11	0,23	0	1	0	0	0,11	1	0
NSP 90D	2018	0,79	0,14	0,25	0	1	0	0	0,20	1	0
	2019	0,80	0,15	0,25	0	1	0	0	0,22	1	0
	2020	0,74	0,11	0,23	0	1	0	0	0,13	1	0
NSP 120D	2018	0,81	0,15	0,24	0	1	0	0	0,20	1	0
	2019	0,81	0,15	0,25	0	1	0	0	0,22	1	0
	2020	0,76	0,11	0,22	0	1	0	0	0,13	1	0
NSP 365D	2018	0,83	0,15	0,23	0	1	0	0,04	0,20	1	0
	2019	0,84	0,16	0,24	0	1	0	0,06	0,24	1	0
	2020	0,80	0,11	0,21	0	1	0	0	0,14	1	0
Dif. Fecha Cita	2018	1	-14,48	18,55	-276	420	-23	-8	0	7	0
	2019	1	-14,99	18,34	-238	472	-24	-8	-1	8	0
	2020	1	-10,17	15,92	-210	389	-15	-5	0	11	0

Tabla 7: Total de variables numéricas por año

Con el fin de evidenciar diferencias una vez ocurrido el fenómeno de la pandemia se realiza una división artificial de un periodo de 8 meses, entre noviembre del año anterior y junio del siguiente año, para poder visualizar las diferencias de forma equitativa entre ambos periodos. Esto nos permite tener dos periodos temporalmente comparables: uno como control que abarca noviembre de 2018 a junio de 2019 y otro, que abarca noviembre del 2019 y junio de 2020 y que evidencia la ocurrencia del fenómeno considerando meses previos y posteriores al inicio de la pandemia. El total de las tablas se resume en el Anexo 10.1, pero se deja un ejemplo en la Tabla 8 para entender mejor la forma del análisis.

Del análisis general de los dos periodos podemos encontrar las siguientes observaciones:

1. En el periodo de la pandemia, la disminución del número de citas es notoria en los meses de abril mayo y junio, alcanzando su máximo el mes de mayo con una disminución de 85.428 citas (68,2%) en relación con el año control.

2. En la variable edad el promedio no presenta diferencias entre ambos periodos, manteniéndose estable. Se destaca que se modifica el valor más común en los meses de marzo y abril, donde inicialmente los valores correspondían a menores de 2 años y se ven modificados por los pacientes de 60 años.

3. Número de citas previas: El promedio de citas previas disminuye en el periodo de pandemia, con mayor diferencia en el mes de marzo en cada uno de los cortes. La diferencia se va haciendo mayor en la medida que va aumentando el periodo del análisis, partiendo en el promedio de 30 días en 1.16 puntos en marzo y terminando en el promedio de 365 días en el mismo mes con 5.98 puntos promedio. Esto da a entender que la cantidad de citas fueron disminuyendo en los meses de pandemia, lo que reitera lo discutido en el primer punto. Adicionalmente, el valor más común de todas las mediciones es cero, es decir, pacientes que no tienen citas previas.

Variable	Periodo	Mes	N° Filas	Comp	Prom	DS	Mín	Máx	Perc 25	Perc 50	Perc 75	Perc 99	VMC
Citas previas 30 días	Control	11	125.499	100%	2,67	3,81	0	45	0	1	3	18	0
		12	107.328	100%	1,14	2,20	0	51	0	0	1	10	0
		1	127.022	100%	2,60	3,85	0	53	0	1	3	17	0
		2	99.028	100%	2,77	3,91	0	58	0	1	4	18	0
		3	126.119	100%	2,76	4,10	0	59	0	1	4	18	0
		4	126.444	100%	2,67	3,92	0	49	0	1	3	19	0
		5	125.200	100%	2,86	4,11	0	51	0	1	4	20	0
		6	121.770	100%	2,86	4,21	0	50	0	1	4	20	0
	Pandemia	11	112.770	100%	2,71	3,69	0	38	0	1	4	17	0
		12	108.630	100%	1,29	2,50	0	39	0	0	2	12	0
		1	125.259	100%	2,71	3,96	0	47	0	1	4	18	0
		2	100.409	100%	2,69	3,90	0	61	0	1	4	18	0
		3	100.805	100%	1,60	2,83	0	58	0	1	2	13	0
		4	44.761	100%	2,07	3,29	0	32	0	1	3	15	0
		5	39.772	100%	2,34	3,47	0	34	0	1	3	16	0
		6	41.604	100%	2,43	3,50	0	30	0	1	3	16	0

Tabla 8: Análisis variable numérica Citas previas 30 días. Destacado en verde mayores variaciones del periodo.

4. El NSP es notoriamente menor en época de pandemia, con la mayor diferencia en el mes de abril en todos los periodos. Esto se evidencia tanto en los valores promedio como en los percentiles, en especial en el percentil 75.

5. Diferencia de días entre agendamiento y cita: La diferencia disminuye en la pandemia, con su valor máximo en el mes de junio 2020 con 8,92 días menos de diferencia entre el agendamiento y la cita en promedio. Al analizar los percentiles se evidencia una disminución en esta misma cifra, pero también un aumento en los días de agendamiento que ocurren en el futuro. La diferencia es especialmente perceptible en el percentil 99, donde el valor pasa de 5 días de agendamiento en el futuro a un total de 46 para el mes de junio. El valor más común sigue siendo cero, es decir, las citas agendadas el mismo día, lo cual es destacable ya que los establecimientos de atención secundaria no suelen tener agendamientos espontáneos que justifiquen este valor.

4.3. Análisis de variables categóricas

Para el análisis de variables categóricas se consideraron los mismos márgenes de tiempo que para las variables numéricas. El detalle de forma anual se presenta en la Tabla 9.

	Año	Completitud	Cardinalidad	VMC	% VMC
Sexo	2018	100%	3	Mujer	58%
	2019	100%	4	Mujer	58%
	2020	100%	3	Mujer	58%
Pueblo originario	2018	54%	14	Ninguna	79%
	2019	70%	14	Ninguna	81%
	2020	77%	14	Ninguna	78%
Nacionalidad	2018	100%	46	Chile	98%
	2019	100%	44	Chile	98%
	2020	100%	46	Chile	97%
Previsión	2018	100%	6	Fonasa	93%
	2019	100%	5	Fonasa	94%
	2020	100%	5	Fonasa	94%
Tipo profesional	2018	84%	22	Médico	54%
	2019	88%	21	Médico	55%
	2020	91%	23	Médico	60%
Prestación	2018	100%	359	Consulta Integral De Especialidades	16%
	2019	100%	397	Consulta Integral De Especialidades	17%
	2020	100%	318	Consulta Integral De Especialidades	16%
Comuna	2018	100%	301	San Bernardo	21%
	2019	100%	290	San Bernardo	21%
	2020	100%	280	San Bernardo	23%
Especialidad	2018	67%	78	Med, Interna	8%
	2019	70%	78	Med, Interna	9%
	2020	74%	80	Psiquiatría Adulto	11%
Descripción cita	2018	6%	78	Consulta Repetida	35%
	2019	5%	77	Consulta Repetida	38%
	2020	6%	66	Consulta Repetida	36%
Establecimiento	2018	100%	4	Hospital Barros Luco Trudeau	54%
	2019	100%	4	Hospital Barros Luco Trudeau	55%
	2020	100%	4	Hospital Barros Luco Trudeau	48%
Comuna Establecimiento	2018	100%	3	San Miguel	72%
	2019	100%	3	San Miguel	71%
	2020	100%	3	San Miguel	65%

Tabla 9: Detalle variables categóricas

Respecto al género se puede observar que este es constante, con un 58% en mujeres en todos los años, en donde el segundo mayor valor es hombre, dejando menos de 1% a otros valores. Sobre la nacionalidad esta presenta una alta cardinalidad en todos los años, pero el porcentaje mayor (~98%) siempre corresponde a Chile. Por su parte, la previsión se correlaciona con lo esperado a

un hospital público, con altos valores de pacientes pertenecientes al seguro público de salud FONASA, lo que se condice con lo indicado en la sección 1.1.1. Se aprecia que el 2020 disminuye la cardinalidad de la variable Prestación, lo que da a entender que la variedad de la oferta durante ese año se vio disminuida en relación con los anteriores. Esto mismo se ve reforzado si analizamos la variable comuna del paciente, que también presenta una disminución en su cardinalidad en el mismo año. Si bien llama la atención la alta cardinalidad de esta variable y es posible que exista algún error de registro, se da a entender que los hospitales que son de referencia nacional del SSMS no pudieron atender a tantos pacientes provenientes de otras regiones que no fueran la Región Metropolitana. Por su parte en especialidad se ve un cambio en la variable con mayor porcentaje, ya que en el año 2020 esta pasa a ser Psiquiatría Adulto en vez de Medicina como en los dos años anteriores. Hay que destacar que la variable Descripción de la cita cuenta con muy baja completitud, esto se debe a que es un campo opcional de registro de información en texto libre que es adicional a la agenda de la cita.

Usando la misma metodología que en las variables numéricas se elaboraron de forma artificial dos divisiones temporales artificiales de 8 meses de duración: una como control y otra como pandemia, para analizar mejor el fenómeno de pandemia (análisis completo se encuentra en el Anexo 10.2). Se destaca que las consultas de profesional médico, valor más común en todos los meses, aumentan en proporción en la pandemia, con un 62 % promedio sobre el 55% del año anterior como se detalla en la Tabla 10.

Periodo/Mes	11	12	1	2	3	4	5	6
Control 2018-2019 % VMC (Médico)	54%	54%	56%	54%	55%	54%	54%	55%
Pandemia 2019-2020 % VMC (Médico)	57%	56%	55%	56%	62%	66%	62%	62%

Tabla 10: Consultas por profesional médico, periodo control v/s pandemia

Esto da a entender que, si bien todas consultas disminuyeron en el periodo, lo hicieron en mayor medida las consultas correspondientes a otros profesionales de la salud no médicos. Por otro lado en la Tabla 11, se aprecia una baja en la cardinalidad de la comuna del paciente, en donde la comuna de San Bernardo, el valor más común en todos los meses, aumenta su proporción sobre el total. Esto

puede dar a entender que las atenciones fueron centradas de mayor manera en la población asignada a cada establecimiento por la disminución de la movilidad del periodo.

Periodo/Mes		11	12	1	2	3	4	5	6
Control 2018-2019	Cardinalidad	243	226	229	222	232	237	234	229
	% VMC (San Bernardo)	21%	20%	21%	21%	21%	21%	21%	20%
Pandemia 2019-2020	Cardinalidad	225	212	234	211	210	163	149	154
	% VMC (San Bernardo)	21%	22%	21%	21%	22%	24%	24%	24%

Tabla 11: Variable Comuna del establecimiento, periodo control v/s. pandemia

Respecto a las especialidades, inicialmente el valor que tiene mayor cita correspondiente a medicina interna, pero luego este se modifica y pasa a ser psiquiatría adulto con un 17% y un 19% para los meses de mayo y junio del 2020, tal como se ve en la Tabla 12.

Periodo/Mes		11	12	1	2	3	4	5	6
Control 2018-2019	VMC	Psiquiatría Adulto	Med. Interna	Med. Interna	Med. Interna	Med. Interna	Med. Interna	Med. Interna	Med. Interna
	% VMC	8%	8%	10%	9%	8%	8%	8%	8%
Pandemia 2019-2020	VMC	Med. Interna	Med. Interna	Med. Interna	Med. Interna	Med. Interna	Med. Interna	Psiquiatría Adulto	Psiquiatría Adulto
	% VMC	12%	11%	10%	11%	12%	14%	17%	19%

Tabla 12: Variable Especialidad periodo control v/s pandemia

Por último, el Hospital Barros Luco Trudeau (HBLT) presenta una disminución en su representatividad (porcentaje de citas sobre el total) en la época de pandemia, llegando hasta un 43% por sobre los demás establecimientos, lo que da a entender que las citas de mayor complejidad disminuyen en la red del SSMS en mayor proporción que otro tipo de consultas ofertadas de forma más genérica por los demás establecimientos de la red. El detalle de cada mes se puede ver en la Tabla 13.

Periodo/Mes	11	12	1	2	3	4	5	6
Control 2018-2019 % VMC (HBLT)	53%	55%	55%	56%	54%	55%	57%	56%
Pandemia 2019-2020 % VMC (HBLT)	52%	51%	54%	54%	51%	43%	44%	46%

Tabla 13: Hospital Barros Luco Trudeau periodo control v/s pandemia

4.4. Análisis de variables calculadas

De las variables asociadas a atenciones realizadas por teleconsultas, podemos encontrar el resumen en la Tabla 14. En ésta se consideran todas las atenciones que fueron realizadas en el periodo de pandemia. Como se comentó anteriormente en la Tabla 4, las atenciones a distancia corresponden a un grupo de atenciones que no es posible determinar si son realizadas de forma telefónica o por videollamada de forma efectiva, por lo que se deEne en una categoría diferente.

De la Tabla 14 podemos interpretar que en general no existen mayores diferencias en el NSP en las consultas realizadas a distancia en relación con el NSP de las consultas que son de forma presencial. Por su parte las atenciones por vía telefónica, algo menos de un tercio que las anteriores, tienen un NSP levemente mayor. Finalmente, las atenciones realizadas por videollamada tienen un NSP cerca de un 70% mayor que las consultas realizadas de forma presencial, aunque representan a un número menor dentro de la muestra.

Atributo	Total Citas	% NSP
Distancia	53.726	9,72%
Videollamada	3.299	16,82%
Telefónica	15.596	10,39%
Presencial	550.319	9,68%

Tabla 14: Total teleconsultas periodo de pandemia

En la Tabla 15 se aprecia el análisis mensual, donde se evidencia que en todos los tipos de atenciones por teleconsultas éstas van aumentando en la medida que avanza el tiempo. Asimismo, el NSP de los 3 tipos presenta un alza en la medida que avanzan los meses, con el valor más alto correspondiente a las videollamadas con un 26,4% de NSP en el mes de noviembre. Sin embargo, al ser un número menor de consultas no afectan al total del NSP que es llevado

principalmente por las consultas presenciales, que tienen una disminución marcada en su NSP en relación con el promedio, el cual se va estabilizando hasta el mes de diciembre donde alcanza el valor promedio de los años previos a la pandemia según lo evidenciado en la Ilustración 2 y en la Tabla 15.

	Distancia		Videollamada		Telefónica		Presencial	
Mes	N	% NSP	N	% NSP	N	% NSP	N	% NSP
3*	200	0,0%	8	0,0%	2	0,0%	52.197	11,6%
4	792	0,3%	44	0,0%	45	0,0%	43.880	6,4%
5	2.185	7,6%	56	0,0%	197	8,6%	37.334	8,7%
6	5.501	7,5%	117	9,4%	2.362	7,5%	33.624	8,4%
7	6.039	10,2%	375	19,7%	2.789	7,3%	44.226	7,6%
8	7.288	8,8%	571	15,4%	2.443	9,2%	52.593	8,8%
9	8.930	9,5%	696	15,7%	2.161	10,9%	62.674	9,4%
10	8.009	10,8%	582	15,1%	2.109	11,1%	71.656	9,8%
11	8.384	11,3%	502	18,5%	1.928	14,0%	79.434	10,9%
12	6.398	11,3%	348	26,4%	1.560	16,7%	72.701	12,1%
Total	53.726	9,73%	3.299	16,82%	15.596	10,39%	550.319	9,7%

*Tabla 15: Total de atenciones por teleconsulta y NSP por mes * Marzo considera desde el día 12 en adelante*

4.5. Correlaciones entre variables

Una correlación es una medida de asociación entre variables numéricas. En datos correlacionados, el cambio en la magnitud de una variable es asociado con el cambio en la magnitud de otra variable, ya sea en el mismo sentido (correlación positiva) o en el sentido opuesto (correlación negativa). La correlación de Pearson es usada típicamente para datos que se distribuyen de forma normal. Para datos que no se distribuyen de forma normal, o datos con alto número de *outliers* se puede usar la correlación de Spearman. Ambos coeficientes de correlación se encuentran en el rango entre -1 y +1, donde el 0 indica que no existe asociación entre las variables. Si la relación se vuelve más fuerte hasta alcanzar una línea recta (para el caso de Pearson) o existe una curva monótona que aumenta o disminuye (para Spearman) en la medida que el coeficiente se acerca al valor absoluto de 1 (34).

En la Tabla 16 se encuentran los resultados de ambos tipos de correlaciones con el NSP divididas por año. Destaca que ninguna variable demuestra una correlación que tenga una asociación fuerte con el fenómeno (NSP), siendo los

mayores valores encontrados en el histórico de NSP, que tienen siempre mayores valores de correlación en los años 2018-2019, y en los días de diferencia entre el agendamiento y la cita.

Variable	Año	Pearson	Spearman
Edad	2018	-0,07	-0,08
	2019	-0,10	-0,11
	2020	-0,07	-0,08
Citas Previas 30D	2018	-0,05	-0,06
	2019	-0,05	-0,07
	2020	-0,05	-0,06
Citas Previas 60D	2018	-0,05	-0,08
	2019	-0,06	-0,08
	2020	-0,05	-0,07
Citas Previas 90D	2018	-0,05	-0,08
	2019	-0,06	-0,08
	2020	-0,05	-0,08
Citas Previas 120D	2018	-0,05	-0,08
	2019	-0,06	-0,09
	2020	-0,05	-0,08
Citas Previas 365D	2018	-0,05	-0,08
	2019	-0,06	-0,09
	2020	-0,04	-0,08
NSP 30D	2018	0,16	0,14
	2019	0,17	0,15
	2020	0,16	0,14
NSP 60D	2018	0,17	0,15
	2019	0,18	0,16
	2020	0,16	0,14
NSP 90D	2018	0,18	0,15
	2019	0,19	0,16
	2020	0,16	0,14
NSP 120D	2018	0,18	0,16
	2019	0,19	0,16
	2020	0,16	0,13
NSP 365D	2018	0,19	0,16
	2019	0,19	0,16
	2020	0,16	0,13
Dif. Fecha Cita	2018	-0,13	-0,18
	2019	-0,12	-0,16
	2020	-0,13	-0,18

Tabla 16: Correlaciones entre variables por año. Todas las correlaciones son estadísticamente significativas

Al igual que en las variables numéricas y categóricas, se hizo una división artificial de periodos temporales para evidenciar mejor el fenómeno de pandemia.

Los resultados se encuentran en el Anexo 10.3. De éstas se puede concluir que la edad presenta un cambio en el periodo de pandemia, bajando la correlación con el target a un 0,02 en abril en comparación al 0,1 de abril del periodo de control. Respecto a las variables de citas previas y NSP histórico todas tienen bajas en correlaciones en el periodo de marzo-abril 2020, tanto en comparación con los meses de control como con el resto de los meses de pandemia.

4.6. Modelos de predicción de NSP

Con la finalidad de tener un mayor entendimiento de como el conjunto de variables influye en el fenómeno NSP y el impacto de la pandemia se elaboraron modelos de aprendizaje de máquina.

Se escogieron 3 algoritmos para realizar el entrenamiento, según lo expuesto en la sección 1.1.7 y en el trabajo realizado por Ramirez et al (35). Para este caso, fueron Regresiones Logísticas, *Random Forests* y *XGBoost*, donde éste último fue seleccionado basado en el trabajo de Chen et al (36) donde obtuvieron un ROC AUC 0,90 utilizando este tipo de algoritmos por sobre otros más tradicionales. El objetivo fue elegir algoritmos que permitan cierto grado de interpretabilidad para analizar el fenómeno de mejor forma. Se explicará brevemente de qué se trata cada algoritmo:

1. Regresión Logística: Nos permite estimar la probabilidad $(\hat{p})$ de pertenencia a la clase positiva $(y = 1)$ a través de una suma ponderada de las variables representadas como $(\theta_i X_i)$, un término de *bias* (θ_0) y una función logística, como se muestra en la siguiente fórmula:

$$Logit(p) = Log\left(\frac{p}{1-p}\right) = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \ldots + \theta_n X_n$$

Donde el *n* es el número de variables, X_i es el valor de la variable *i,* por ejemplo la edad del paciente y (θ_i) son los parámetros del modelo y p es la probabilidad de pertenencia a la clase positiva. Luego para obtener la probabilidad se calcula la función inversa del Logit. Con esto si la $(\hat{p} \geq 0{,}5)$ se dice que el paciente pertenece a la clase positiva, es decir, el paciente va a faltar a su cita, en caso contrario $(\hat{p} < 0{,}5)$ el paciente asistirá. Por ejemplo en la Ilustración 3, si el resultado de una predicción

es un -1,34 al aplicar la inversa nos quedaría una probabilidad de 0,21 y utilizando el umbral de 0,5 el registro quedaría clasificado en la clase negativa, es decir, asistirá a su cita.

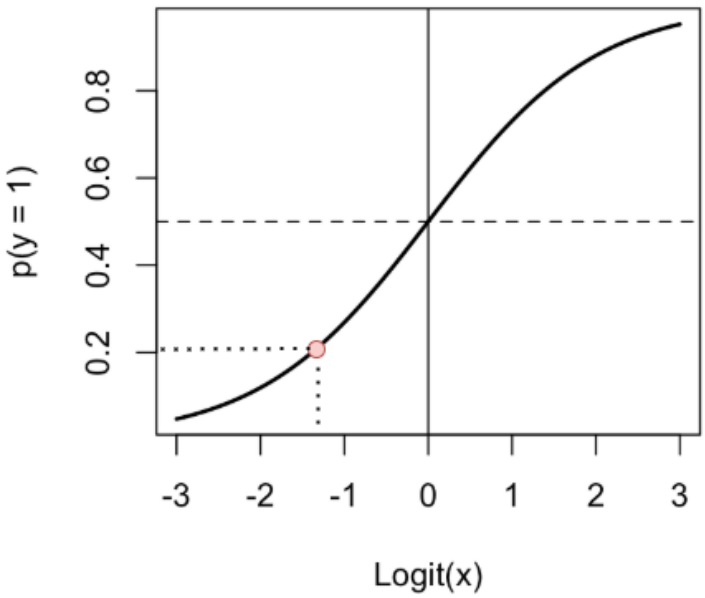


Ilustración 3: Ejemplo gráfico función logística. Elaboración propia

Notar que las regresiones logísticas nos permiten explicar a través de los parámetros del modelo θ_i cómo afectan los cambios en las variables independientes, lo que lo hace una herramienta poderosa para estudios epidemiológicos o que necesitan ser explicativos (37).

2. *Random Forests*: Para entender este modelo primero es necesario entender qué son los árboles de decisión. Un árbol de decisión es una estructura similar a un diagrama de flujo en el que cada nodo representa una condición en un atributo, cada rama representa el resultado de la condición y cada hoja representa una etiqueta de clase (decisión tomada después de calcular todos los atributos). Las ramificaciones desde el inicio del árbol hasta la hoja representan reglas de clasificación. Un ejemplo de árbol de decisión se puede observar en la Ilustración 4

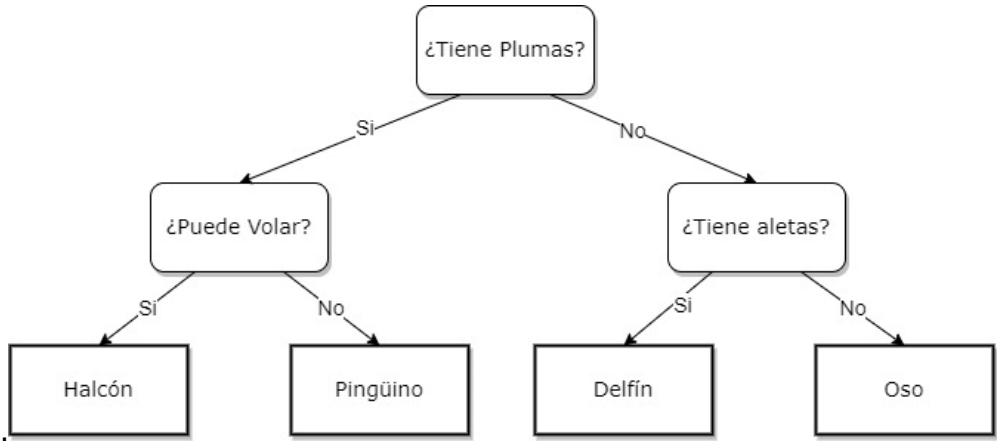


Ilustración 4: Ejemplo de árbol de decisión.

Los *Random Forests* son un conjunto de árboles de decisión de modo que cada árbol recibe un conjunto de variables aleatorias y emite un voto, luego, al combinar estos resultados, se obtiene el resultado de clasificación final. (38,39). Finalmente para consolidar los resultados se puede utilizar una estrategia de mayoría simple (*hard-voting*) o una media de probabilidades como se muestra en la siguiente fórmula:

$$\hat{p} = \hat{f}(X) = \frac{1}{B}\sum_{b=1}^{B} f_b(X_b),$$

Donde *B* es el número de árboles, f_b es la instancia del árbol *b* y X_b el sub-conjunto de variables del modelo *b*. La estrategia de media de probabilidad es la que implementa por defecto la biblioteca utilizada en este trabajo.

3. *XGBoost* o *Extreme Gradient Boosting*: son una implementación escalable de los *tree boosting* (40). Para entender estos últimos podemos partir por establecer que el enfoque tradicional para la construcción de modelos es construir un solo modelo predictivo robusto. Un enfoque diferente es construir un conjunto de modelos como los *Random Forests* y luego unificar el resultado mediante el promedio simple. Por otro lado, la familia de los *boosting* busca entrenar modelos de forma secuencial donde cada iteración aprende del error anterior (41). Para esto iniciamos con un modelo inicial f_0 entrenado con variables X_i y que aprende de las etiquetas Y_i. Luego para los siguientes modelos se modifican las etiquetas de la siguiente forma $\widehat{y_i} = y_i - f(X_i)$, donde $f(X_i)$ es la probabilidad ($\hat{p}$) del modelo *i*, y_i es la etiqueta con que se entrenó el

modelo e $\hat{y}_i$ las etiquetas del modelo siguiente. Finalmente para tener una predicción se unifican los modelos, multiplicados por un hiperparámetro α de la siguiente forma:

$$\hat{p} = \hat{f}(xX) = f_0(X) + \alpha_1 f_1(X) + \cdots + \alpha_n f_n(X)$$

Para el análisis del fenómeno se utilizará el modelo con mejor rendimiento, evaluado a través el *F1-Score,* el cual es explicado en detalle en la Tabla 18. Adicionalmente, cada algoritmo tiene parámetros que pueden ser modificados, llamados hiperparámetros. La búsqueda de hiperparámetros óptimos es un proceso costoso en tiempo y recursos computacionales, por lo cual se utilizaron los definidos por defecto, con la excepción de los que nos permite tratar el desbalance de clases.

5. MODELAMIENTO DEL PROBLEMA

Como se detalló en la sección 1.2, que un paciente no se presente a su cita ambulatoria genera múltiples problemáticas tanto médicas como de salud pública.

Con el fin generar acciones preventivas para evitar que los pacientes falten a sus citas es que se busca elaborar un modelo de aprendizaje de máquinas que en base a características de la persona, la cita y el establecimiento pueda predecir la probabilidad de que el paciente no se presente. Luego, dado el fenómeno de la pandemia y el auge de las teleconsultas, se desea identificar si las variables que explican este comportamiento se mantienen. En esta línea se decide generar una metodología innovadora que permita utilizar de la mejor forma posible la gran cantidad de datos disponibles.

Para esto, se implementa una metodología de encasillamiento temporal que considera la elaboración de divisiones por segmentos temporales delimitados con una historia de 12 meses previos. A cada uno de los puntos generados lo llamaremos Punto de Tiempo (PT). Cada PT corresponde a un periodo de un mes, el cual busca predecir el comportamiento de dicho mes con la información de un año anterior a ese mismo mes. Para esto, se considera como data de entrenamiento los datos de los 12 meses correlativos anteriores, la cual es dividida en data de entrenamiento y validación, en proporción 80 – 20 %, respectivamente. De esta forma, cada mes del set de datos se transforma en un PT que es entrenado

por un algoritmo que considera los 12 periodos anteriores de igual duración. Como ejemplo, para predecir el mes de abril del año 2019, se entrena con los datos de abril del 2018 a marzo del 2019.

Debido a que no se cuenta con información anterior al año 2018 para llevar a cabo esta metodología, se realiza un total de 24 puntos de tiempo que comienzan en enero del año 2019 que generarán igual cantidad de modelos, como se aprecia en la Ilustración 5. Así, se genera un modelo específico por cada PT que sirve para predecir el mes correspondiente en base a la data de los 12 meses anteriores.

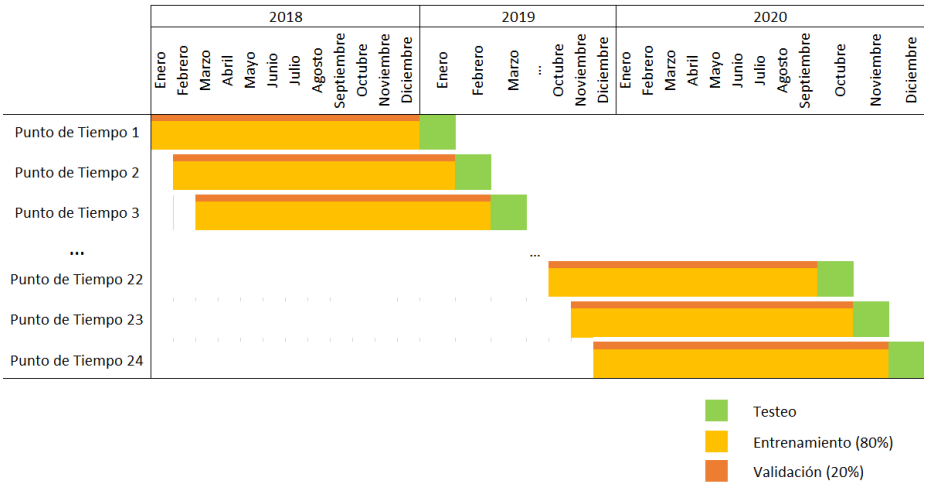


Ilustración 5: PT generados para los años 2019 a 2020. Elaboración propia.

De forma paralela se implementa una variación de la metodología de los PT, donde se considera de forma exclusiva el periodo de pandemia. Se realizó una modificación temporal en la data de entrenamiento pensando en generar un modelo que se entrene exclusivamente con datos acuñados en pandemia, bajo la premisa de que el entregar solo estos datos podría generar un modelo más limpio y específico, ya que en los análisis descriptivos se evidenció de que a raíz de la pandemia se modificaba de forma importante el comportamiento del NSP. Para esto, en vez de considerar la historia de 12 meses previos a cada mes como se explicó anteriormente, solo se entrena con los meses anteriores disponibles a cada punto de tiempo, con lo que en la medida que avanzan los meses la cantidad de datos de entrenamiento va aumentando gradualmente. Así, abril 2020 es entrenado con el mes de marzo 2020, pero mayo 2020 se entrena con los meses de marzo y abril 2020 juntos. Con este esquema se generan 9 PT solo de la época de pandemia 2020 que darán resultado a igual cantidad de modelos, como se observa en la Ilustración 6.

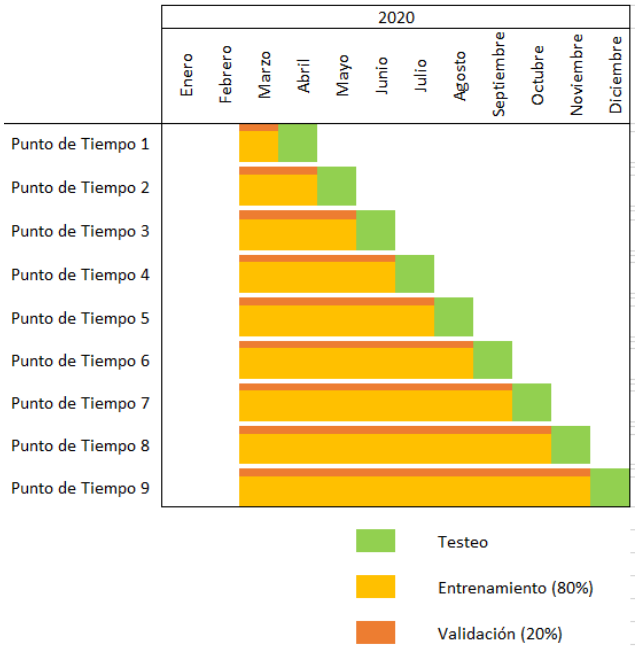


Ilustración 6: PT para periodo de pandemia

5.1. Manejo de variables

Dado que no todos los algoritmos son capaces de manejar por defecto las variables categóricas, es necesario aplicar algún tipo de transformación que nos permitan trabajar con ellas. Para esta tesis vamos a ocupar la estrategia de One-Hot Enconding, que transforma una variable con *n* categorías distintas en *n* variables binarias, de tal forma que cada variable binaria asociada a una categoría toma valor 1 si indica la presencia o 0 en caso de ausencia de la categoría (42). Por ejemplo si tenemos una variable categórica llamada “Colores” y posee tres posibles valores: Rojo, Amarillo y Verde, está se transformaría en tres columnas distintas con la siguiente distribución respecto a la variable original:

$$Rojo = [1, 0, 0]$$

$$Amarillo = [0, 1, 0]$$

$$Verde = [0, 0, 1]$$

Debido a que existen variables con alta cardinalidad (con alto número de categorías), como especialidad, tipo de profesional y prestación, es que se decidió definir un umbral tras el cual las variables cuyos valores representaran un menor

porcentaje del total de la muestra quedaran agrupadas en una nueva categoría llamada Otros. Se hicieron pruebas de diferentes umbrales: 10%, 5%, 1% y 0,5% y se ejecutaron los modelos con cada uno. El mejor desempeño se evidenció en el umbral del 1%, por lo que en adelante este será utilizado para los análisis.

Finalmente para entrenar se utilizó el siguiente subgrupo de variables:

Edad, Sexo, Nacionalidad, Pueblo originario, Previsión, Está en la misma comuna, Fecha de la cita, Mes, Minuto del Día, Día de la semana, Atención a distancia, Atención telefónica, Atención Videollamada, Tipo profesional, Prestación, Especialidad, Atención en pandemia, Comuna establecimiento, NSP 30D, NSP 60D, NSP 90D, NSP 120D, NSP 365D, Citas previas 30D, Citas previas 60D, Citas previas 90D, Citas previas 120D, Citas previas 365D, Diferencia fecha agendamiento/cita, Target (Se presentó o no)

5.2. Métricas utilizadas

Para la medición del rendimiento de cada modelo se toman como base los análisis que se pueden hacer de la Matriz de Confusión, la cual nos permite cuantificar de forma sencilla las predicciones de cada modelo con sus valores reales como se encuentra en la Tabla 17. Por ejemplo, un Verdadero Positivo, es el número de instancias en donde el modelo dijo que la persona no se iba a presentar y efectivamente no se presentó en su cita.

		Condición Predicha	
		Positivo (PP)	**Negativo (PN)**
Condición Real	**Positivo (P)**	Verdadero Positivo (TP)	Falso Negativo (FN)
	Negativo (N)	Falso Positivo (FP)	Verdadero Negativo (TN)

Tabla 17: Matriz de confusión

Para la evaluación de los modelos se ocuparán métricas más complejas que se derivan de la matriz de confusión y nos permiten un mejor entendimiento del rendimiento de los modelos. En la Tabla 18 se listan las métricas y sus fórmulas.

Nombre Métrica	Fórmula
Accuracy (ACC)	$Accuracy = \frac{TP + TN}{P + N}$
Precisión (PPV)	$Precisión\ (PPV) = \frac{TP}{PP}$
Sensibilidad (TRP)	$Sensibilidad\ (TPR) = \frac{TP}{P}$
F1-Score	$F1 - Score = 2 * \frac{precisión * sensibilidad}{presición + sensibilidad}$
Geometric Mean (GM)	$GM = \sqrt{TPR * TNR}$, donde $TNR = \frac{TN}{N}$
Matthews Coefficient Correlation (MCC)	$MCC = \sqrt{TPR * TNR * PPV * NPV} - \sqrt{FNR * FPR * FOR * FDR}$, donde $NPV = \frac{TN}{PN}$, $FNR = \frac{FN}{P}, FPR = \frac{FP}{N}, FOR = \frac{FN}{PN}, FDR = \frac{FP}{PP}$

Tabla 18: Tabla de métricas

Adicionalmente, se utilizará la métrica de ROC AUC, la cual no está definida en base a la matriz de confusión, si no, como el área bajo la curva ROC. La curva ROC busca representar la combinación entre los verdaderos positivos (la proporción de ejemplos positivos predichos de forma correcta) y los falsos positivos (proporción de ejemplos negativos predichos incorrectamente con respecto a un umbral de clasificación), para construir un gráfico que resume el rendimiento del modelo. Por consiguiente, las curvas ROC solo pueden generarse con modelos que devuelven algún tipo de *score* de confianza (o probabilidad) de predicción (23).

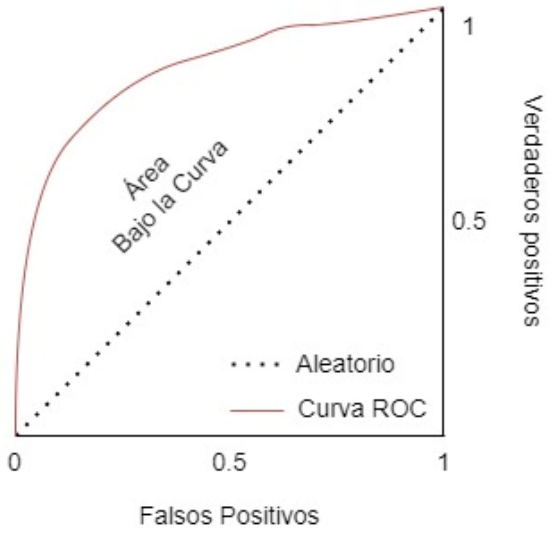


Ilustración 7: Gráfico curva ROC. Elaboración propia.

En la Ilustración 7 se aprecia la diferencia en las curvas ROC entre un clasificador aleatorio (ROC AUC 0,5) y un modelo en particular. El área bajo la curva marcada con color rojo es el ROC AUC del modelo. Los rangos de valores son entre 0 y 1.

Para efectos de esta tesis, la métrica principal será dada por el *F1-Score*, esto por su capacidad de generar una relación entre la precisión y la sensibilidad con rangos de 0 a 1, donde entre mayor es el valor mejor es el modelo. Igualmente, en los análisis se presentará el detalle de las otras métricas formuladas.

6. RESULTADOS

6.1. Resultados generales

Se realizó la ejecución de los tres de algoritmos con la totalidad de las variables como se describe en sección 5.1.

Se calcula el promedio de *F1-Score* sobre el conjunto de datos de testeo, es decir, con datos que no estuvieron presentes en la fase de entrenamiento. Para

decidir a qué clase pertenece cada paciente, se trabaja con un umbral de probabilidad de 0.5, es decir:

$$\hat{y} = \left\{ \begin{array}{l} Se\ presenta\ (0)\ Si\ \hat{p}\ < 0{,}5 \\ No\ se\ presenta\ (1)\ Si\ \hat{p}\ \geq 0{,}5 \end{array} \right\}$$

Luego, todas las métricas son calculadas para la clase positiva (no se presenta). Se observa que el modelo con mejor *F1-Score* corresponde al entrenado utilizando el algoritmo *XGBoost* con un valor de 0,28. El promedio de las métricas para cada algoritmo se detalla en la Tabla 19, donde el Número de Positivos corresponden a las personas que el modelo indicó que iban a faltar.
Se destaca la precisión que tiene el Algoritmo *Random Forests*, sin embargo, tiene una muy baja sensibilidad ya que selecciona una pequeña cantidad de pacientes.

Modelo	**Positivos ($\hat{p}$ > 0.5)**	***F1-Score***	**ROC AUC**	**Sensibilidad**	**Precisión**
Regresión Logística	111.5028	0,26	0,63	0,61	0,17
Random Forests	10.079	0,02	0,72	0,01	0,38
XGBoost	805.033	0,28	0,68	0,53	0,20

Tabla 19: Resultados generales modelos

Respecto a las métricas mensuales, los resultados del Algoritmo *XGBoost* se encuentran en la Tabla 20. Para las Regresiones Logísticas y *Random Forests* las tablas de resultados pueden ser encontrados en los Anexos 10.4 y 10.5, respectivamente.

Modelo	Año	Mes	% NSP	Positivos ($\hat{p} > 0.5$)	*F1 - Score*	ROC AUC	Precisión	*Sensibilidad*	GM	MCC
XGBoost	**Total**	**-**	**12%**	**805.033**	**0,28**	**0,68**	**0,20**	**0,53**	**0,60**	**0,16**
	2019	Ene	12%	40.794	0,32	0,72	0,22	0,58	0,65	0,21
		Feb	13%	36.525	0,35	0,75	0,24	0,68	0,68	0,25
		Mar	13%	47.564	0,28	0,66	0,19	0,56	0,60	0,15
		Abr	13%	47.095	0,35	0,75	0,24	0,69	0,68	0,25
		May	14%	42.396	0,30	0,65	0,21	0,53	0,60	0,16
		Jun	14%	47.363	0,30	0,66	0,20	0,58	0,61	0,16
		Jul	13%	48.850	0,34	0,74	0,23	0,66	0,67	0,23
		Ago	13%	47.534	0,29	0,66	0,20	0,56	0,61	0,16
		Sept	14%	37.079	0,29	0,65	0,21	0,53	0,60	0,15
		Oct	17%	51.607	0,34	0,65	0,24	0,59	0,60	0,16
		Nov	17%	49.016	0,34	0,66	0,24	0,62	0,61	0,17
		Dic	15%	50.410	0,31	0,66	0,21	0,65	0,61	0,16
	2020	Ene	13%	32.898	0,33	0,71	0,25	0,49	0,62	0,20
		Feb	13%	32.891	0,32	0,70	0,23	0,56	0,63	0,19
		Inicio de la Pandemia								
		Mar	13%	40.518	0,27	0,64	0,18	0,56	0,59	0,12
		Abr	6%	10.365	0,17	0,66	0,11	0,42	0,57	0,11
		May	9%	4.397	0,21	0,69	0,19	0,24	0,47	0,13
		Jun	8%	7.583	0,23	0,70	0,17	0,39	0,57	0,16
		Jul	8%	12.830	0,22	0,67	0,15	0,46	0,60	0,15
		Ago	9%	9.272	0,22	0,67	0,18	0,30	0,51	0,14
		Sept	10%	20.699	0,23	0,66	0,16	0,46	0,58	0,13
		Oct	10%	32.274	0,24	0,67	0,15	0,60	0,61	0,14
		Nov	11%	32.455	0,24	0,62	0,16	0,51	0,58	0,11
		Dic	12%	22.618	0,25	0,63	0,18	0,42	0,55	0,11

Tabla 20: Resultados XGBoost 24 PT. El umbral establecido para las métricas definidas por umbral fue de 0,5.

Tal como se comentó en la sección 4.1 el NSP disminuye en especial a inicios de la pandemia. Esto afecta directamente a las métricas que son sensibles al desbalance de clases, como precisión, sensibilidad y el *F1-Score*. Esto se ve evidenciado, con el cambio en el promedio de la sensibilidad, que pasa desde un 0,58 prepandemia a un 0,43 en pandemia, o en el *F1-Score* que desciende hasta un 0,17 en abril del 2020.

Por su parte podemos considerar como medida comparable entre los periodos el ROC AUC, ya que no es una métrica sensible al desbalance de clases. Al comparar los periodos prepandemia y pandemia existe una leve disminución en el promedio del ROC AUC, lo que nos indica que la capacidad del modelo de

predecir el fenómeno disminuye. Adicionalmente, si bien en ambos existe variabilidad entre los distintos meses, las fluctuaciones son menores en el periodo pandemia lo que nos indica que existe un fenómeno más estable. Los valores más altos, de 0,75, se encuentran en los meses de febrero y abril 2019.

El MCC, al igual que ROC AUC, nos muestra que el modelo es mejor que una predicción aleatoria ya que alcanzan valores superiores a 0, cuando sus rangos son entre -1 y 1 donde 0 indica que no existe una relación. Sin embargo, se ve más afectado que el ROC AUC en el periodo de pandemia dado que toma un umbral fijo para su cálculo, el cual se ve afectado por el cambio en el fenómeno, que en este caso es la disminución en el NSP.

En general es posible identificar que luego del inicio de la pandemia existe un cambio que impacta de forma negativa en las métricas de los modelos, afectando el desempeño. Esto debido a que, hasta esa fecha, el modelo se encontraba siendo entrenado con datos más bien estables en términos de número de citas, NSP y distribución en general y, luego del inicio de la pandemia, éstos se modificaron abruptamente. Es importante notar que, con el paso del tiempo, los modelos son capaces de adaptarse a este nuevo escenario llegando al mes de diciembre de 2020 a tener un valor de *F1-Score* de 0,25, mucho más cercano al 0,28 del promedio del periodo.

Utilizando la estrategia de entrenar solo con los datos del periodo de la pandemia en la Tabla 21 se evidencia el resultado del algoritmo de 9 PT, con las mismas métricas utilizadas anteriormente. Se puede observar que al realizar una comparación mes a mes se obtienen peores métricas, por ejemplo, el *F1-Score* baja de un 0,23 a 0,15 analizando el promedio en el periodo de pandemia.

Modelo	Año	Mes	% NSP	Positivos ($\hat{p} > 0.5$)	*F1 - Score*	ROC AUC	Precisión	Sensibilidad	GM	MCC
XGBoost	**Total**	**-**	**12%**	**48.914**	**0,15**	**0,63**	**0,19**	**0,14**	**0,35**	**0,09**
	2020	Abr	6%	2.143	0,16	0,67	0,18	0,14	0,36	0,11
		May	9%	1.229	0,12	0,70	0,23	0,08	0,28	0,09
		Jun	8%	1.072	0,11	0,65	0,23	0,07	0,27	0,09
		Jul	8%	1.459	0,11	0,64	0,22	0,07	0,27	0,09
		Ago	9%	4.022	0,12	0,60	0,15	0,11	0,32	0,06
		Sept	10%	6.935	0,14	0,58	0,15	0,14	0,36	0,05
		Oct	10%	9.746	0,20	0,62	0,18	0,22	0,44	0,10
		Nov	11%	6.958	0,13	0,56	0,16	0,11	0,32	0,05
		Dic	12%	15.350	0,26	0,65	0,22	0,34	0,53	0,14

Tabla 21: Resultados XGBoost 9 PT

Cuantificando la estrategia de 12 meses, o modelo de 24 PT, en comparación con el modelo de 9 PT que utiliza solo los datos de pandemia, el primer modelo logra identificar correctamente (precisión) 24.327 pacientes que van a faltar a su cita mientras que el segundo solo 9.124, lo que nos da una mejora considerable en términos de *F1-Score* promedio, en donde el modelo de 24 PT tiene un promedio de 0,28 por sobre el 0,15 del modelo de 9 PT. Dado esto, podríamos inferir que si existe un cambio en la forma en que las variables independientes explican el fenómeno esto siempre da mejor complementado con la información de los 12 meses anteriores al fenómeno.

Dado los resultados anteriores los análisis de las importancias de las variables se enfocarán en la estrategia de 12 meses.

6.2. Análisis de importancia de las variables

Una herramienta que entrega el algoritmo *XGBoost*, por pertenecer a la familia de los árboles de decisión, es obtener un listado de importancia de cada variable. Existen 3 formas para calcular la importancia de una variable*:*

1. *Gain* (Ganancia): corresponde a la contribución relativa de cada variable respecto a cada árbol en el modelo. Una variable con un valor más alto implica que es más importante para la generación de la predicción. El concepto es que antes de generar una división en una variable determinada habría elementos mal clasificados, y luego de añadir la

división se generan dos nuevas ramas donde cada una de esas ramas tiene mayor *accuracy* que las ramas anteriores. Es considerada el atributo más importante para interpretar la importancia relativa de cada variable.

2. *Coverage* (Cobertura): Número de veces que una variable fue usada para dividir la data a través de todos los árboles, medido por el número de datos de entrenamiento que pasan por esas divisiones. Esta métrica corresponde al número relativo de observaciones relacionadas con esta variable. Por ejemplo, partiremos asumiendo que se tienen 100 observaciones, 4 variables y 3 árboles. Podemos suponer que la variable 1 es usada para decidir en el nodo para 10, 5 y 2 observaciones en los árboles 1, 2 y 3 respetivamente. Entonces, la métrica considerará la cobertura como 10+5+2 = 17 observaciones. Esto será calculado para las 4 variables y la cobertura será 17 expresado como porcentaje de la cobertura de todas las variables presentes.

3. *Weight* (Peso): Es el porcentaje que representa el número relativo de veces que una variable aparece en las divisiones de los árboles del modelo. Como ejemplo, si la variable 1 se encuentra en 2 divisiones, 1 división y 3 divisiones en los árboles 1, 2 y 3, entonces el peso de la variable 1 será 2+1+3 =6. La frecuencia de la variable 1 como porcentaje se calcula entre el peso de todas las variables.

Notar que estas definiciones solo nos permiten ordenar y asignar un valor a las variables, pero no asegura que exista una relación directa con el fenómeno que estamos tratando de explicar ya que se tratan de modelos no lineales. Para este trabajo se decidió utilizar la medida de ganancia a la hora de calcular la importancia de las variables del *XGBoost*.

Luego de generar el *one-hot encoding* el conjunto de datos cuenta con un total de 114 variables. Para entregar interpretabilidad a la información se calculó el porcentaje relativo de aparición de cada variable con relación a todas las variables disponibles en cada uno de los PT, entregando para cada variable un valor porcentual que representaba el peso en el dataset. Luego, se consideró el promedio

de cada variable a lo largo de los 24 PT y se elaboró un ranking, del que fueron seleccionadas las 10 variables con mayor valor para realizar el análisis explicativo. Como ejemplo, podemos revisar lo expuesto en la Tabla 22, donde se muestran 3 variables y su ganancia por cada PT.

	Variable 1	Variable 2	Variable 3
PT 1	40	12	15
PT 2	5	25	12
PT 3	14	34	27

Tabla 22: Ejemplo distribución peso de variables

Luego, sacamos el valor porcentual que representan dichas ganancias con relación a cada PT, como se muestra en la Tabla 23. Desde acá se promedia el porcentaje de cada variable y según ese valor se seleccionan las variables que se consideran importantes para el modelo. Para el caso del ejemplo el orden de importancia de las variables es 2, 1 y 3.

	Variable 1	Variable 2	Variable 3
PT 1	60%	18%	22%
PT 2	12%	60%	29%
PT 3	19%	45%	36%
Promedio	30%	41%	29%

Tabla 23: Ejemplo distribución peso porcentual de variables

Las variables que presentaron el mayor valor promedio y que por tanto fueron seleccionadas se encuentran en la Ilustración 8, en donde se expresa de forma gráfica el porcentaje de peso de la variable y su evolución a lo largo de los 24 PT. Para las variables binarias se incluye el porcentaje que representa cada variable cuando es positiva en el set de datos completo, el porcentaje de NSP de las citas que pertenecen a esa clase (ejemplo, a la especialidad hematología) y el NSP de las que no pertenecen (que pertenecen a una especialidad distinta a hematología).

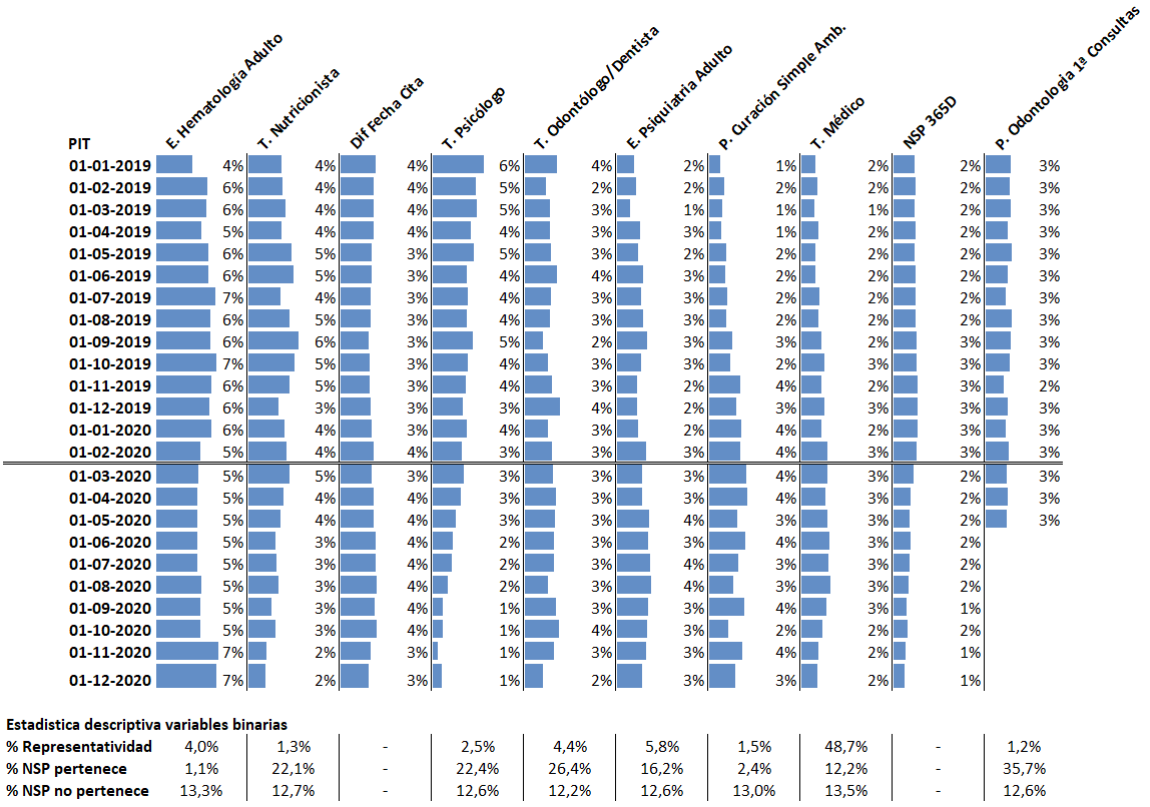


Ilustración 8: Top 10 Importancia de Variables XGBoost. La línea horizontal marca el inicio de la pandemia.

Del análisis de la Ilustración 8 podemos obtener algunas conclusiones, Por ejemplo, la variable con mayor importancia corresponde a la especialidad de Hematología, que tiene una importancia de magnitud estable en los 24 PT. Se destaca el bajo porcentaje de NSP (1,1%) que tienen las citas que pertenecen a esta categoría, lo cual da a entender el motivo de la importancia que le asigna el modelo.

Los tipos de profesional nutricionista y psicólogo tienen un comportamiento similar en relación con su importancia como variable dentro de la predicción, con una disminución en su importancia una vez llegada la pandemia. Esto se explica por una disminución brusca que tuvo en NSP en el periodo de pandemia, en relación con las otras variables del modelo. Se destaca el alto porcentaje de NSP (22,1% y 22,4%) de las citas que se encuentran dentro de estas clases.

La diferencia de días entre el agendamiento y la cita aparece como una variable con alta importancia, lo cual se condice con las correlaciones obtenidas en la sección 4.5. Hay que destacar que su importancia disminuye en la época de pandemia.

El tipo de profesional odontólogo/dentista presenta una importancia estable a lo largo de los 24 PT. Una variable parecida que aparece es la prestación Odontología primeras consultas, que tiene similares porcentajes de importancia hasta su desaparición en la pandemia. Se destaca en ambos un NSP (26,4% y 35,7%) muy por sobre el promedio de la muestra.

Para la especialidad psiquiatría adulto se observa una importancia con tendencia al alza en el periodo de pandemia, con un NSP (16,2%) levemente mayor al promedio de la muestra.

La prestación Curación simple ambulatoria presenta un aumento de importancia en la pandemia, con un NSP (2,4%) muy bajo con relación a los que no pertenecen a dicha clase.

El tipo de profesional médico se destaca por su alta representatividad dentro del conjunto de datos, con un 48%. Si bien el NSP (12,2%) es muy similar al promedio, su importancia se incrementa con la llegada de la pandemia.

El NSP en los últimos 365 días es otra de las variables que obtuvieron un valor de correlación relevante en los análisis previos. Si bien tiene alta importancia en los primeros PT, con la llegada de la pandemia esta va en disminución.

6.3. Resultado modelo de selección de variables

Con el fin de tener un modelo con mejores métricas y más explicativo, se generó un nuevo flujo de entrenamiento basado en el *XGBoost* de la sección 6.1, utilizando el método *RecursiveFeatureAddition* de biblioteca *feature-engine* el cual nos permite seleccionar las variables que más aportan a través de un proceso recursivo que consta de los siguientes pasos:

1. Entrena un modelo inicial utilizando todas las variables.
2. De los resultados del primer modelo, hacer un ranking de las variables de acuerdo con su importancia utilizando el peso.

3. Luego, entrena un nuevo modelo solo con la variable más importante y determina su desempeño (tomando como medida el *F1-Score*).

4. Agregar la segunda variable más importante y entrena nuevamente el modelo.

5. Calcula la diferencia en desempeño entre ambos modelos.

6. Si el *F1-Score* aumenta en al menos 0.01, entonces la variable se agrega al modelo. De lo contrario, la variable es removida.

7. Se repiten los pasos 4 a 6 hasta que todas las variables han sido evaluadas.

El resultado es un modelo depurado en donde sólo se consideran variables importantes, dejando de lado las que entregan poco valor.

Utilizando el algoritmo se vuelven a obtener 24 modelos, el detalle de las métricas se observa en la Tabla 24.

Modelo	Año	Mes	% NSP	Positivos ($\hat{p} > 0.5$)	*F1 - Score*	ROC AUC	Precisión	Sensibilidad	GM	MCC
XGBoost	**Total**	**-**	**12%**	**900.965**	**0,32**	**0,74**	**0,22**	**0,65**	**0,66**	**0,22**
	2019	Ene	12%	47.995	0,33	0,73	0,22	0,67	0,66	0,22
		Feb	13%	38.262	0,34	0,74	0,23	0,68	0,67	0,23
		Mar	13%	49.625	0,34	0,74	0,22	0,70	0,68	0,24
		Abr	13%	50.420	0,33	0,73	0,22	0,68	0,66	0,22
		May	14%	51.408	0,35	0,73	0,24	0,70	0,67	0,24
		Jun	14%	51.912	0,34	0,73	0,23	0,71	0,66	0,23
		Jul	13%	51.076	0,34	0,73	0,23	0,67	0,66	0,23
		Ago	13%	53.344	0,33	0,72	0,22	0,68	0,66	0,22
		Sept	14%	44.390	0,35	0,73	0,23	0,71	0,67	0,24
		Oct	17%	53.317	0,39	0,73	0,27	0,70	0,66	0,24
		Nov	17%	50.081	0,40	0,74	0,28	0,73	0,67	0,26
		Dic	15%	59.568	0,35	0,72	0,22	0,81	0,64	0,22
	2020	Ene	13%	39.755	0,35	0,74	0,25	0,60	0,66	0,23
		Feb	13%	35.039	0,35	0,74	0,24	0,63	0,66	0,23
		Inicio de la Pandemia								
		Mar	13%	39.502	0,33	0,72	0,22	0,66	0,65	0,21
		Abr	6%	12.461	0,20	0,74	0,12	0,55	0,64	0,15
		May	9%	9.645	0,30	0,78	0,20	0,57	0,67	0,24
		Jun	8%	9.852	0,27	0,75	0,18	0,52	0,64	0,20
		Jul	8%	9.087	0,29	0,76	0,21	0,45	0,62	0,22
		Ago	9%	17.726	0,29	0,76	0,19	0,61	0,68	0,23
		Sept	10%	19.116	0,28	0,73	0,19	0,52	0,64	0,20
		Oct	10%	28.823	0,28	0,73	0,18	0,63	0,66	0,20
		Nov	11%	40.468	0,29	0,73	0,18	0,74	0,66	0,21
		Dic	12%	38.093	0,31	0,73	0,20	0,76	0,66	0,21

Tabla 24: Resultados XGBoost 24 PT con Modelo de selección de variables

Si se compara con el *XGBoost* original, nos encontramos con que existe una mejora en todas las métricas evaluadas, en donde se destaca el *F1-Score* que aumenta de un 0,28 a un 0,32 y un ROC AUC que sube de 0,68 a 0,74.

6.4. Análisis de importancia de las variables en el modelo de selección de variables

Como se explicó en la sección 6.2. existen diferentes formas de calcular la importancia de las variables. Para el caso de la biblioteca *RecursiveFeatureAddition* se utiliza la medida de peso para listar la importancia, lo que hará cambiar el listado de variables seleccionadas por el modelo. Como resultado del algoritmo de adición

de variables el total de variables disminuye de las 114 iniciales a un total de 54. El umbral utilizado para seleccionar si una variable es o no incluida en el modelo corresponde a un aumento de desempeño del orden de 0,01 puntos de *F1-Score.*

Con el fin de tener mayor interpretabilidad se realiza un análisis mensual de los cambios de las variables con relación al mes anterior y se agrega el porcentaje de importancia de la variable para el modelo del PT correspondiente. El total de los resultados se puede revisar en la Ilustración 9 e Ilustración 10.

01-01-19		01-02-19		01-03-19		01-04-19		01-05-19		01-06-19		
Diff Fecha Cita	-	41% Diff Fecha Cita	→	24% Diff Fecha Cita	→	26% Diff Fecha Cita	→	37% Diff Fecha Cita	→	38% Diff Fecha Cita	→	24%
T.Odontologo/Dentista	-	16% T. Psicologo	→	18% T. Psicologo	→	17% T. Psicologo	→	16% T. Psicologo	→	17% T. Psicologo	→	16%
NSP 365D	-	13% T. T. Ocupacional	→	11% P. Odon. Primeras Con.	↗	12% NSP 365D	↗	13% NSP 365D	→	13% NSP 365D	→	13%
T. Nutricionista	-	12% NSP 365D	↘	9% NSP 365D	→	11% T. T. Ocupacional	↗	11% T.Odontologo/Dentista	↗	12% T. T. Ocupacional	↗	13%
T. Cirujano Dentista	-	10% T. Nutricionista	↘	8% T. T. Ocupacional	→	11% T.Odontologo/Dentista	↗	8% T. T. Ocupacional	↘	11% T.Odontologo/Dentista	↘	11%
P. Consulta Psiquiatria	-	4% P. Odon. Primeras Con.	→	8% P. Odon. Repetidas	↗	9% T. Nutricionista	→	8% T. Nutricionista	→	8% T. Nutricionista	→	10%
NSP 90D	-	4% P. Consulta Esp. CDT	→	7% T.Odontologo/Dentista	↗	8% P. Odon. Primeras Con.	↘	7%		P. Odon. Primeras Con.	→	10%
		P. Odon. Repetidas	→	7% P. Consulta Control	→	4%				Minuto del día	→	2%
		T.Odontologo/Dentista	↘	5% Minuto del día	→	2%				Día del Mes	→	2%
		Día Semana	→	1%								

01-07-19		01-08-19		01-09-19		01-10-19		01-11-19		01-12-19		
Diff Fecha Cita	→	37% Diff Fecha Cita	→	27% Diff Fecha Cita	→	21% Diff Fecha Cita	→	20% Diff Fecha Cita	→	17% Diff Fecha Cita	→	20%
T. Psicologo	→	17% T. Psicologo	→	16% T. Psicologo	→	13% T. Psicologo	→	15% P. Odon. Primeras Con.	→	15% T. Nutricionista	↗	15%
NSP 365D	→	15% P. Odon. Primeras Con.	→	12% NSP 365D	↗	12% NSP 365D	→	13% T. Nutricionista	↗	12% NSP 365D	↗	13%
T.Odontologo/Dentista	↗	10% NSP 365D	↘	12% T. T. Ocupacional	→	12% T. Nutricionista	↗	11% NSP 365D	↘	10% T. Psicologo	↗	12%
T. Nutricionista	↗	10% T. Nutricionista	→	12% T. Nutricionista	→	11% T. T. Ocupacional	↘	10% T. Psicologo	↘	8% T.Odontologo/Dentista	→	11%
T. Médico	→	8% P. Odon. Repetidas	→	8% P. Consulta Esp. CDT	→	8% T.Odontologo/Dentista	↗	9% T. T. Ocupacional	↘	7% T. T. Ocupacional	→	10%
P. Con. Psiquiatrica	→	4% T. Médico	↘	6% T.Odontologo/Dentista	→	8% P. Consulta Esp. CDT	↘	8% P. Odon. Repetidas	→	7% P. Consulta Esp. CDT	↗	7%
		E. OtorrinoLaringolo	→	3% E. OtorrinoLaringolo	→	7% T. Médico	↗	7% P. Consulta Esp. CDT	↘	7% NSP 30D	↗	3%
		P. Con. Psiquiatrica	↘	3% T. Médico	↘	5% T. Otro	→	3% T. Médico	↘	5% Edad	↗	3%
				Minuto del día	→	2% Minuto del día	→	2% P. Consulta CRS	→	3% E. Med. Interna	→	3%
						Día Semana	→	1% P. Otro	→	2% Prev. Citas 60D	→	2%
								Edad	→	2% Minuto del día	→	2%
								NSP 30D	→	2%		
								T. Enfermera	→	2%		

Simbología

→ Sigue en el mismo puesto que el periodo anterior

→ Aparece respecto al periodo anterior

↗ Sube lugares respecto al periodo anterior

↘ Baja lugares respecto al periodo anterior

Ilustración 9: Resultados 12 PT iniciales XGBoost RecursiveFeatureAddition

01-01-20		01-02-20		01-03-20		01-04-20		01-05-20		01-06-20		
Diff Fecha Cita	→	19% Diff Fecha Cita	→	22% Diff Fecha Cita	→	17% Diff Fecha Cita	→	19% Diff Fecha Cita	→	18% Diff Fecha Cita	→	23%
NSP 365D	↗	14% T.Odontologo/Dentista	↗	15% T.Odontologo/Dentista	→	12% P. Odon. Primeras Con.	→	14% T.Odontologo/Dentista	→	13% T.Odontologo/Dentista	→	14%
T. Psicologo	↗	12% NSP 365D	↘	12% T. Psicologo	↗	10% T. Psicologo	→	10% T. Psicologo	→	11% T. Psicologo	→	12%
T. Nutricionista	↘	12% T. T. Ocupacional	↗	12% NSP 365D	↘	10% P. Odon. Repetidas	→	10% T. T. Ocupacional	↗	9% NSP 365D	↗	11%
T.Odontologo/Dentista	→	9% T. Psicologo	↘	11% T. T. Ocupacional	↘	10% T. T. Ocupacional	→	9% NSP 365D	↗	9% T. T. Ocupacional	↘	10%
T. T. Ocupacional	→	8% T. Nutricionista	↘	10% P. Consulta Esp. CDT	↗	7% NSP 365D	↘	9% P. Odon. Primeras Con.	↘	8% T. Nutricionista	→	9%
P. Consulta Esp. CDT	→	7% P. Consulta Esp. CDT	→	9% P. Consulta Control	→	7% P. Consulta Esp. CDT	↘	8% P. Consulta Esp. CDT	→	7% P. Consulta Esp. CDT	→	8%
P. Kine Ambulatoria	→	4% P. Kine Ambulatoria	→	4% T. Cirujano Dentista	→	6% P. Consulta Control	↘	7% P. Consulta Control	→	6% P. Consulta Gine/Trau	↗	4%
P. Otro	→	3% NSP 30D	↗	3% T.T. Paramédico	→	6% T. Asistente Social	→	4% T. T. Paramédico	→	5% NSP 30D	↗	3%
NSP 30D	↘	3% Minuto del día	↗	2% T. Kinesiologo	→	4% NSP 60D	→	3% P. Consulta Gine/Trau	→	4% Edad	↗	3%
Edad	↘	3%		NSP 30D	↘	3% C. San Bernando	↗	3% P. Kine Ambulatoria	→	3% Minuto del día	↗	2%
P. Consulta CRS	→	3%		Edad	→	2% Edad	→	2% NSP 30D	→	3%		
Día del año	→	2%		C. San Bernando	→	2% Minuto del día	↗	1% Edad	↘	3%		
Minuto del día	↘	2%		Prev. Citas 60D	→	2%		Minuto del día	↘	1%		
				Minuto del día	↘	1%						

01-07-20		01-08-20		01-09-20		01-10-20		01-11-20		01-12-20		
Diff Fecha Cita	→	34% Diff Fecha Cita	→	29% Diff Fecha Cita	→	28% Diff Fecha Cita	→	25% Diff Fecha Cita	→	22% P. Desp. Receta Cronicos	↗	19%
T. Médico	→	15% NSP 365D	↗	10% T.Odontologo/Dentista	↗	19% E. Hematología Adulto	→	20% P. Desp. Receta Cronicos	→	20% E. Hematología Adulto	→	16%
NSP 365D	↗	10% T. Médico	↘	9% NSP 365D	↘	12% P. Curación Simple	→	11% P. Odon. Repetidas	→	9% Diff Fecha Cita	↘	13%
C. San Miguel	→	8% T. Nutricionista	→	9% P. Consulta Esp. CDT	→	9% T. Médico	→	9% P. Consulta Esp. CDT	→	8% P. Curación Simple	→	7%
P. Odon. Repetidas	→	8% T.Odontologo/Dentista	→	9% E. Rehabilitación Protesis	→	6% NSP 365D	↘	7% T. Médico	↘	7% NSP 365D	↗	6%
Prev. Citas 365D	→	7% C. San Miguel	↘	8% C. San Miguel	→	6% Prev. Citas 365D	→	5% NSP 365D	↘	6% P. Odon. Repetidas	↘	6%
P. Consulta salud mental	→	5% P. Odon. Repetidas	↘	7% Atención Pandemia	→	6% E. Obstetricia	→	5% Prev. Citas 365D	↘	4% T. Médico	↘	4%
Atención Pandemia	→	4% P. Psicologo clínico	→	5% T.T. Paramédico	→	5% Cita Distancia	→	5% P. Otros	→	4% E. Neurología Adulto	→	4%
Cita a Distancia	→	3% E. Neurologica Adulto	→	4% E. Psiquiatria Infantil	→	3% NSP 30D	→	4% C. Buin	→	4% C. San Miguel	→	3%
C. Buin	→	3% Edad	→	3% Edad	→	3% P. Venosa Adulto	→	3% Atención Pandemia	→	4% NSP 120D	→	3%
Día del mes	→	2% NSP 30D	→	3% Minuto del día	→	2% Edad	↘	3% NSP 60D	→	3% Atención Pandemia	↘	3%
		C. San Bernando	→	2%		Minuto del día	↘	2% P. Consulta Psiquiatrica	→	3% P. Psicologo clínico	→	3%
						Prev. Citas 120D	→	2% NSP 30D	↘	3% P. Otros	↘	3%
								Edad	↘	2% P. Consulta Esp. CDT	↘	2%
										NSP 30D	↘	2%
										Cita a Distancia	→	2%
										Edad	↘	2%
										Minuto del día	→	1%

Simbología

→ Sigue en el mismo puesto que el periodo anterior

→ Aparece respecto al periodo anterior

↗ Sube lugares respecto al periodo anterior

↘ Baja lugares respecto al periodo anterior

Ilustración 10: Resultados 12 PT finales XGBoost RecursiveFeatureAddition

Para interpretar las ilustraciones debe considerar a siguiente simbología:

1. ↑: La variable no se encontraba en el modelo del mes pasado y aparece en el mes del análisis.
2. →: La variable se encontraba en el modelo del mes pasado y se ubica en la misma posición. Esto no significa que tenga el mismo porcentaje de importancia, sino que solo se rescata su posición con relación al total de las variables.
3. ↗: La variable se encontraba en el modelo del mes pasado, pero aumentó su posición.
4. ↘: La variable se encontraba en el modelo del mes pasado, pero disminuyó su posición

De los resultados podemos concluir que una de las variables con mayor importancia desde los primeros modelos corresponde a la diferencia entre la fecha de agendamiento y cita, la cual se mantiene con altos porcentajes de importancia a lo largo de los 24 modelos. Las variables que más se repiten entre los 24 modelos, independiente de su posición o porcentaje de importancia, son el NSP de 365 días que aparece en todos los modelos y la diferencia entre el día de agendamiento y la cita que aparece en 23 modelos. Con valores menores aparecen el tipo de profesional Psicólogo y Odontólogo, cada uno con 17 apariciones, y finalmente el tipo de profesional Nutricionista que cuenta con 15 apariciones. Si se comparan estos resultados con los obtenidos en el *XGBoost* original nos encontramos con que coinciden las mismas variables.

Con menores valores aparecen variables como los tipos de profesional Nutricionista y Terapeuta Ocupacional. Adicionalmente, se destaca la variable Minuto del día, que presenta 14 apariciones. Con 13 apariciones se encuentra la prestación Consulta de Especialidad en CDT, lo que nos da a entender que las atenciones de especialidad tienen influencia superior sobre aquellas atenciones que no son por especialista. Por último, se presenta la edad con 12 apariciones.

Con relación al porcentaje relativo de importancia de cada variable la Diferencia de días entre el agendamiento y la cita aparece nuevamente como la variable más importante, con un 25% promedio en los 24 PT. La sigue la prestación Despacho receta crónicos, que si bien cuenta con pocas apariciones en los últimos meses, ya que es una prestación que se gestó debido a la pandemia, cuenta con un 20% promedio. Al igual que en el *XGBoost* original aparece nuevamente la especialidad hematología adulto en tercer lugar, con un 18%. Se repite en cuarto y quinto lugar los tipos de profesional psicólogo y odontólogo con un 14% y 11%, respectivamente.

Se puede observar que desde el mes de noviembre 2019 comienza a aparecer el NSP de 30 días como variable relevante, lo cual nos podría indicar que comienza a tomar relevancia la información más reciente del paciente, complementando la información histórica que nos puede entregar variables como el NSP de 365 días.

Dentro de las variables que aparecen en época de pandemia nos encontramos con las variables Atención en pandemia y Atención a distancia, lo que nos da luces de que estas variables tienen importancia, aunque ésta es de forma muy menor. También se evidencia que las comunas del establecimiento comienzan a tomar importancia, en concreto San miguel con 4 apariciones del total de 9 de meses de pandemia, lo que nos puede indicar que hay un fenómeno asociado con los establecimientos que pertenecen o no pertenecen a estas comunas por sobre las demás comunas.

En contraparte, la variable Tipo de profesional Psicólogo tiene alta importancia en los primeros PT, pero desde el mes de julio del 2020 desaparece por completo como variable. Esto puede deberse a que antes de la pandemia el NSP del psicólogo sobresalía del promedio, con valores de hasta 25%. Con la llegada de la pandemia éstos descendieron hasta un 12% en promedio un valor mínimo de 5% en abril 2020, sin que las citas tuvieran una disminución importante, ajustándose más al promedio de aquella época.

6.5. Resultados modelos específicos

Con el objetivo de tener modelos con mejores indicadores y que las conclusiones del análisis de las variables sean más robustas se generaron modelos por subgrupos específicos, es decir, considerando determinadas variables presentes como categoría. Para esto se utilizaron las variables categóricas de Especialidad, Tipo de Profesional y Prestación y se tomó como medida el ROC AUC para elegir el top 3. No es posible continuar ocupando el *F1-Score* como se venía haciendo hasta ahora para evaluar los modelos, ya que este se ve influenciado por el desbalance de clases entre los modelos.

Luego de ejecutar los 3 modelos por cada categoría se analizaron los resultados. Tanto para los modelos de Especialidad como Prestación no se presentaron resultados muy superiores a los del modelo general, con valores de ROC AUC similares a los discutidos previamente, es decir, alrededor del 0,74, con lo que no existía una ganancia importante en términos de desempeño. Sin embargo, algunos modelos asociados a Tipo de Profesional obtuvieron valores de ROC AUC por sobre el 0,74 los cuales corresponden a los tipos de profesional Matrona, Enfermera y Asistente Social. El resultado promedio para para cada modelo especifico se observa en la Tabla 25.

Modelo	Positivos ($\hat{p} > 0.5$)	*F1-Score*	ROC AUC	Sensibilidad	Precisión
Matrona	11.208	0,386	0,79	0,52	0,31
Enfermera	43.100	0,283	0,80	0,62	0,18
Asistente Social	5.887	0,407	0,81	0,47	0,36

Tabla 25: Resultados modelos por tipo de profesional

Utilizando la misma metodología anterior para el análisis de la importancia de las variables se encuentran los siguientes hallazgos:

1. Para los modelos de matrona, el sexo empieza a tener importancia que se ve acentuada con la pandemia. El detalle de porque existen diversos sexos en el Tipo de profesional matrona, es que los hombres se atienden generalmente para consultoría ETS y seguimiento de programa VIH. Adicionalmente, destacan las citas previas (30 días y 60 días) y el minuto del día en los modelos entrenados.

2. Para los modelos de Enfermera, la prestación de "Consulta integral de especialidades en medic interna y subesp oftalmo neurolo oncologia en CDT" aumenta su importancia considerablemente a medida que se avanzan en los puntos de tiempo, con un crecimiento adicional en la época que empieza la pandemia. Adicionalmente, destacan la especialidad de cirugía adulto y urología en los modelos entrenados.

3. Para los modelos de asistente social, desde agosto del 2020 aparece la variable “Atención en pandemia” hasta los últimos modelos entrenados, esto quiere decir que existe un cambio en el fenómeno desde el inicio de la pandemia que los modelos están considerando a través de esta *variable*. Adicionalmente, destacan la prestación “Consulta de enfermera, matrona o Nutricionista”, “Visita domiciliaria integral de salud mental a domicilio, trabajador social” y la especialidad de asistente social.

7. DISCUSIÓN

7.1. Aspectos metodológicos

De este trabajo se destaca el tamaño de la muestra, que, en comparación a trabajos similares, alcanza a 3.787.110 millones de datos, además de considerar tanto datos de hospitales adultos como pediátricos, de múltiples especialidades, lo que entrega muchas opciones a la hora de realizar los análisis.

Adicionalmente, se implementa una metodología de Puntos de Tiempo (PT) que considera la información de los meses previos de cada cita para modelar el problema, lo que generó distintas predicciones y que permite que el entrenamiento y métricas de predicción sea lo más cercano a una implementación real. El desarrollo de esta metodología permite realizar análisis históricos y es posible a futuro replicarlo para otros periodos de tiempo, como podría ser el año 2021 si es que se cuenta con dicha información disponible.

Por otro lado y si bien, no fueron impedimento para realizar el trabajo, si se debieron hacer adecuaciones para solventar malas prácticas a nivel de registros.

El ejemplo más claro tiene que ver con el análisis que se realizó de los días entre el agendamiento y la cita. Cerca del 4% de la base tenía días en positivo, es decir, citas que fueron agendadas posterior a su realización, con hasta 200 días de diferencia, y un alto porcentaje de citas con cero días de diferencia, es decir, que fueron ingresados el mismo día. Lo primero podría ser explicado por un desfase en la actualización de los datos de las agendas, pero lo segundo deja de tener sentido ya que en los hospitales públicos en general no existen en general mecanismos de demanda espontánea que representen el fenómeno, por lo que puede ser atribuido al primer punto. También se destacan otros potenciales errores de registro, como son la cantidad de pacientes mayores de 120 años de edad y los pacientes con un histórico de citas anual superior a 350 citas, lo que nos hace pensar en la existencia de RUN de pacientes que están siendo utilizados con registros genéricos y sin identificadores personales. De cualquier manera, más allá de las inferencias, se recomienda una revisión de los procesos y de los sistemas de agendamiento de los hospitales, de tal forma de que, al continuar con este trabajo y su implementación, sea posible que los modelos sean entrenados de forma más fehaciente y con mayor robustez.

Es importante destacar que este trabajo considera la información recopilada entre los años 2018 a 2020. Esto trae algunas dificultades, como que el año 2019 es un año complejo si se considera que los meses de octubre y noviembre presentan alteraciones debido al estallido social ocurrido en Chile, que alteró la asistencia a las citas programadas lo cual se evidenció en los análisis. En el año 2020, si bien fue el año en que se desató la pandemia de COVID-19, la respuesta de los establecimientos públicos a la hora de realizar la transformación digital requerida para la implementación de atenciones tomó a las unidades por sorpresa, por lo que las consultas telemáticas se generaron de manera más bien improvisada y de forma paulatina a lo largo del año. Por tanto, sería interesante el poder realizar este mismo trabajo con los datos recopilados entre los años 2021 – 2022 con el fin de obtener conclusiones más robustas ya que en este periodo las consultas telemáticas se encontraban con mayores niveles de implementación.

Si consideramos que una de las variables que más importancia tuvo en el estudio tiene que ver con los días entre el agendamiento y la cita, vale la pena señalar que en la gran mayoría de los servicios clínicos las citas médicas son

asignadas con hasta 3 meses de antelación, lo cual según lo recopilado en la literatura y evidenciado en esta tesis es un factor de riesgo a la hora de no presentarse a una cita. Sería interesante el poder modernizar los sistemas de agendamiento de tal forma de hacerlos dinámicos y más flexibles o, si no, implementar estrategias que sean más transparentes y cercanas al paciente para que se encuentren en contacto directo con el establecimiento y no prolongar los tiempos. En el mismo contexto se podrían potenciar estrategias como Portales de pacientes, que a través de aplicaciones web permitan que el paciente visualice las citas programadas a demanda, sin tener a los establecimientos como intermediarios de entrega exclusiva de información.

7.2. Aspectos éticos

El dilema ético en lo relacionado con algoritmos de aprendizaje de máquinas y su vínculo con aspectos clínicos no es nuevo. Existe una discusión importante en torno a que, el desplegar algoritmos de aprendizaje de máquinas en los sistemas sanitarios mejoraría los procesos de toma de decisiones, lo que se traduciría en una mejora tanto en la velocidad como en la confiabilidad de las decisiones clínicas. Así mismo, se resolverían problemas asociados a sesgos y a las alteraciones de capacidades cognitivas que pudieran llevar a errores en diagnósticos. A nivel de sistemas de salud, podría resolver ineficiencias en los flujos de trabajo, inequidades y potenciales derroches de recursos. Los algoritmos de aprendizaje de máquinas se encuentran popularizados especialmente debido a la información disponible, la cual ha aumentado exponencialmente entre más se implementan sistemas digitales que guardan grandes cantidades de información, lo que genera material para ser trabajado con algoritmos de aprendizaje.

Para analizar los aspectos éticos involucrados se revisará el problema desde las aristas más importantes utilizando la metodología del análisis es basada en el aprendizaje obtenido del curso de capacitación “Formulación Ética de Proyectos de Ciencia de Datos” impartido por la Universidad Adolfo Ibáñez, el cual puede encontrarse descrito en su manual (43). En este apartado se revisarán tanto las consideraciones éticas de la etapa de diseño como también se dejarán sugerencias al equipo que procederá a su implementación.

7.2.1. Proporcionalidad

La utilización de un modelo de aprendizaje de máquinas es apropiado para resolver el problema, ya que permite identificar a los pacientes que no se presentan de forma concreta y disminuir su incidencia de forma directa. Los otros métodos disponibles consideran una mayor utilización de recursos ya que implican una confirmación de horas global, o el sobre agendamiento con sobrecupos. Como parte de los impactos negativos se puede encontrar que el algoritmo presente errores que impidan que se contacte a los pacientes correctos.

7.2.2. Licencia social

Con relación a los usuarios afectados y el uso de sus datos para resolver el problema creo que considerarían su uso aceptable ya que en primera instancia los datos para entrenar no consideran información sensible. Así mismo, la divulgación del uso de un algoritmo de estas características sería aceptado ya que es un aporte a la gestión clínica y administrativa, que aporta en mejorar la calidad de la atención y el uso de los recursos públicos. Adicionalmente, puede ser una medida que aporte a áreas específicas como es la reducción de listas de espera.

Desde el punto de vista legislativo dentro de las normas que justifican la implementación de la iniciativa se debe considerar la Ley 19.937 sobre Autoridad sanitaria y el Reglamento de los servicios de salud (decreto 140 de 2004). En esta se describe que a los Servicios de Salud les corresponderá la “Articulación, gestión y desarrollo de la red asistencial correspondiente, para la ejecución de las acciones integradas de fomento, protección y recuperación de la salud, como también la rehabilitación y cuidados paliativos de las personas enfermas”, con lo que la utilización del modelo aporta en el cumplimiento de este objetivo. Otras regulaciones que pueden impactar en el proyecto son la Ley 19.628 Sobre protección de la vida privada y la Ley N°20.584 que Regula los derechos y deberes que tienen las personas en relación con acciones vinculadas a su atención en salud.

7.2.3. Transparencia

Dentro de los organismos o entidades interesadas que deberían estar al tanto del proyecto se encuentran el mismo Servicio de Salud en especial Departamento de Gestión TIC y Departamento de Gestión de Redes, los distintos

hospitales en específico los profesionales involucrados en las unidades de gestión de redes, admisión y SOME, la comunidad a través del Consejo de la Sociedad Civil (COSOC) y el Consejo Integrado de la Red Asistencial en Salud (CIRA). Para generar una comunicación fluida con estas instancias sería recomendable solicitar audiencia en las reuniones programadas para dar a conocer el proyecto y su impacto, además de capacitar a profesionales de la OIRS en caso de que existan pacientes que quieran acercarse a los establecimientos con dudas. En estas instancias sería recomendable explicar la forma simple como se ejecuta el modelo y las variables que se usan para su construcción.

7.2.4. Discriminación / Equidad

Mediante la realización de este trabajo se genera una recomendación que orienta a la estratificación de pacientes que se presentarán o no a su cita, lo cual puede ser cuestionable. Esto debido a que, si los establecimientos quisieran reducir a cero su porcentaje de NSP bastaría con entregar las citas a aquellos pacientes con menores probabilidades de no presentarse. Esto generaría una exclusión de aquellos pacientes con más probabilidad de ausentismo, dejándolos en una segunda categoría, poniéndolos en una situación de desmedro que puede impactar en términos concretos en su salud, ya que los motivos por los cuales se entrega una cita médica y que se encuentran establecidos por normativa contemplan la edad, el sexo, el diagnóstico, la gravedad y/o las patologías concomitantes del paciente y no cuan probable sea que el mismo asista.

Frente a esto es importante recordar las fases de la entrega de una cita en el proceso ambulatorio, descritas en la sección 5.1.3. Acá se evidencia que existe un proceso de agendamiento de citas que es independiente del de la confirmación de éstas. Por lo tanto vale la pena reforzar que el objetivo en el uso de esta herramienta es su implementación en el proceso de confirmación de citas, no en el agendamiento. El proceso de agendamiento de pacientes debe llevarse a cabo de acuerdo con los criterios descritos de forma precedente, ya que el algoritmo no tiene la posibilidad de entregar mayores luces de a qué pacientes citar o no, ya que para obtener las variables necesarias requiere un análisis de diversos datos, entre ellos históricos y de la cita misma, que no hacen factible su implementación en una fase previa. Es decir, para el entrenamiento del algoritmo se requiere como dato de

base el dato de la cita ya agendada con todas las características de la misma (fecha, hora, tipo de profesional, especialidad, prestación, entre otras), información que no puede ser proporcionada de forma artificial de forma anticipada con el fin de favorecer a uno u otro tipo de pacientes.

7.2.5. Rendición de cuentas

En caso de requerirse información respecto al trabajo acá realizado me encontraré disponible para resolver consultas o dudas una vez entregados los códigos. Se recomienda que la implementación de este tenga lugar en la Unidad de Ciencia de Datos en el Departamento de Gestión TIC, pero dependerá del Servicio de Salud Metropolitano Sur la estrategia de su implementación. Sería importante elaborar una estrategia para el caso en que el algoritmo presente errores o requiera ser optimizado, así como un mecanismo de control de este. Se debe considerar durante este proceso que se desconoce en concreto cómo funciona el algoritmo, lo que podría estar generando una discriminación por sesgo de la población o decisiones tomadas en la implementación, lo cual tendría que ser evaluado durante una implementación.

7.3. Trabajos futuros

Se debe considerar como una segunda etapa la implementación del algoritmo que se presenta en este trabajo, idealmente siendo potenciado por alguna estrategia de contactabilidad de pacientes, la cual puede ser implementada a través del mismo establecimiento mediante llamados telefónicos o a través de sistemas más automatizados que envíen mensajes de texto o correos electrónicos. Siempre es necesario revisar estas estrategias y considerar diversas variables en su implementación, como son las vías mediante la cual se deben contactar a los pacientes o estrategias de recolección de datos teniendo en cuenta factores propios de los pacientes como son su edad, alfabetización digital, entre otros. Se debe considerar durante su implementación un mecanismo de control sobre los pacientes que serán contactados, de tal forma de que la estrategia no afecte el desempeño del modelo. Para esto se propone utilizar de forma inicial un grupo de control sobre el cual implementar la estrategia, ya sea mediante un caso de uso

específico como podría ser una unidad o especialidad determinada. Luego de esto puede ser incluida dentro del modelo una variable denominada “Contactabilidad” o similar que refleje los esfuerzos realizados en esa área y que permitan al modelo aprender también de la implementación de esas estrategias sobre el fenómeno del NSP.

Es importante hacer notar que gran parte de los análisis fueron complejos debido a la pobre calidad de los datos (datos incompletos, con errores en registros), lo cual puede deberse a múltiples factores como son los sistemas de información o la digitación por parte de los usuarios finales. Esto generó problemas a la hora de interpretar los modelos y pudo afectar la robustez de estos. Adicionalmente, por la gran magnitud de datos y el número pruebas generadas existe una complejidad asociada a la interpretabilidad de los modelos que dificulta utilizar herramientas más explicativas como son los *Shap-values*, la cual en caso de utilizarse generaría un resultado por cada predicción requiriendo un tiempo de estudio mayor. En cuanto a recursos tecnológicos es necesario contar con el equipamiento necesario en caso de implementaciones, utilizando más datos o al momento de escalar la implementación de modelos a nivel nacional. Respecto al último punto sería interesante el poder realizar este mismo trabajo pero con set de datos de otros establecimientos o servicios de salud, que permitieran validar si los resultados obtenidos son replicables en otros contextos.

Adicionalmente, el poder conseguir un set de datos que considerara además de las variables ya trabajadas, algunos datos relacionados a variables sociodemográficas o clínicas como podría ser el diagnóstico, tiempo de espera de Listas de espera, si los pacientes son GES, nivel socioeconómico y educacional, y analizar el impacto de éstas en los modelos, cosa de generar un análisis más profundo y tener resultados que permitan capturar a las poblaciones más vulnerables o con consideraciones de mayor urgencia o gravedad desde el punto de vista sanitario. Esto podría ser posible si se contara con una mayor implementación de registro clínico electrónico en la red del servicio sur, por lo que es recomendable redoblar los esfuerzos en este aspecto. Se debería tener en cuenta que en el caso de utilizar este tipo de variables se debiera hacer los ajustes pertinentes en términos de consentimiento informado a los pacientes participantes del estudio.

Por último, y respecto a las teleconsultas, para generar modelos específicos o análisis más detallados sería interesante como se podría desempeñar en un set de datos con mayor cantidad de estas citas, como podría ser en el periodo 2021-2022, donde las estrategias de telemedicina se encuentran más avanzadas. Adicionalmente, el poder contar con una mejor tipificación de este tipo de citas dentro de los sistemas de información, ya que como se evidenció en esta tesis existía un universo mayor de citas sólo consideradas como “distancia” sin especificar la vía por la cual fueron realizadas, lo que impidió extrapolar mejor los análisis.

8. CONCLUSIONES

Mediante este trabajo se corrobora que es posible obtener un modelo que permite identificar a aquellos pacientes con mayor riesgo de no presentarse a su cita ambulatoria y determinar los factores más influyentes en este proceso. Esto fue posible gracias a un modelo tipo *boosting* que considera como variables de entrenamiento la información de los 12 meses previos a cada cita para realizar las predicciones. Adicionalmente, se realizó un flujo de entrenamiento utilizando solo los datos de la pandemia obteniendo peores métricas, lo que nos indica que siempre es preferible contar con información histórica a la hora de entrenar modelos de aprendizaje de máquinas. Por último, se implementaron modelos con una metodología de selección de variables lo que generó un incremento en las métricas del modelo y por consiguiente fue posible realizar análisis más detallados al tener un conjunto reducido de variables.

Es innegable que la pandemia ha tenido un impacto en la forma de vida de las personas, lo cual se refleja en su comportamiento y, en este caso, en su relación con los servicios sanitarios. El que el NSP disminuya durante los meses de pandemia, desde un 13% promedio a un 6% en el mes de abril 2020, habla de que, si bien las citas si disminuyeron de forma importante, los pacientes asistían más a sus citas en este contexto que en épocas normales. Esto puede significar que las citas que no fueron canceladas fueron aquellas que tenían una mayor importancia o urgencia en su realización, por lo que los pacientes de todas formas y pese a las restricciones se encargaban de asistir a su cita.

Los modelos entrenados entregan una lista de importancia de variables que permitieron analizar en detalle cuáles fueron las variables más importantes a la hora de predecir. Si bien la importancia de cada variable no se encuentra directamente relacionada con el fenómeno, si podemos saber que cada variable tiene importancia bajo el concepto que fue definido, que para el caso de este trabajo corresponde a la ganancia. Entre las variables relevantes se destacan el NSP previo de 365 días, los días de diferencia entre el agendamiento y la cita, algunos tipos de profesional como son psicólogo, nutricionista, odontólogo y médico y la especialidad de hematología. Esto nos permite encontrar subgrupos de población que podrían resultar interesantes de tratar de manera particular o entender el por qué ocurre esté fenómeno, por ejemplo: ¿Existe efectivamente un cambio en NSP de este grupo o se debe a una mala digitación, disponibilidad de servicios, protocolos internos, entre otros?

Si nos enfocamos en las variables que predicen el fenómeno en los periodos de pre pandemia y pandemia nos encontramos con que éstas presentan algunas modificaciones, aunque en su gran mayoría se mantienen estables y se tratan de las variables Diferencias entre el día del agendamiento y la cita y el NSP previo de los 365 días. De las variables que destacan por un cambio en su comportamiento debido a la pandemia nos encontramos con el NSP histórico de 30 días, lo cual nos podría indicar que comienza a tomar relevancia la información más reciente del paciente. Otras variables que aparecen en época de pandemia son las variables Atención en Pandemia, que marca a las atenciones ocurridas en este periodo, y la variable Atención a Distancia, lo que nos da a entender que existe una influencia, si bien menor, de las teleconsultas en la predicción del NSP. Finalmente, la comuna del establecimiento, en concreto San Miguel cuenta con 4 apariciones del total de 9 de meses de pandemia, lo que nos puede indicar que hay un fenómeno asociado con los establecimientos que pertenecen o no pertenecen a estas comunas por sobre las demás comunas.

Con los resultados del modelo creemos que es posible generar un impacto importante en el fenómeno de NSP. Si consideramos, por ejemplo, el modelo específico de asistente social (*F1-Score:* 0,4 y Precisión 0,36) sería posible identificar en dos años a aproximadamente 2.000 pacientes que no se presentan a sus citas y que, de ser intervenidos exitosamente mediante acciones de

confirmación de citas, se podría evitar la pérdida de más de 4 millones considerando el coste monetario promedio de una atención de este tipo según FONASA(44). Si extrapolamos el mismo análisis monetario al modelo global (*F1-Score*: 0,32 y Precisión de 0,22) se evitaría la pérdida de al menos $1.200 millones utilizando el valor de una consulta médica sin especialidad del Arancel FONASA en Modalidad de Atención Institucional (MAI), considerando que todos los pacientes son intervenidos exitosamente. Notar que la pérdida puede estar subestimada dado que parte del conjunto de datos eran consultas de especialidad con un valor mayor al utilizado el cálculo. Adicionalmente, es necesario notar que para cualquier intervención de los pacientes es necesario contar con un modelo de contactabilidad que nos asegure que las acciones preventivas tengan el mayor alcance posible.

Desde el punto de vista de los modelos es importante destacar que el desempeño de estos presenta una disminución una vez llegada la pandemia, lo que se evidencia en el descenso del promedio del *F1-Score* desde un 0,33 en marzo a un 0,2 en abril 2020. Esto refuerza la idea de que el fenómeno se ve afectado por la llegada de la pandemia y que los modelos deben adaptarse a este nuevo escenario, mismo fenómeno se puede apreciar en los análisis descriptivos contra los grupos temporales de control, por ejemplo, la diferencia de días entre el agendamiento y la cita disminuye en época de pandemia.

Parte del trabajo corresponde a la identificación de las teleconsultas como parte importante en el periodo de pandemia y ser capaz de establecer si tenían importancia en la predicción de NSP. Se evidenció que en la mayoría de los modelos la ganancia era menor frente a otras variables, aunque la variable Atención a Distancia aparece como relevante en una medida menor en la época de pandemia. Sin embargo, en el análisis descriptivo si fue posible observar que existe un cambio en el comportamiento de los pacientes, ejemplificado por el alto porcentaje de NSP en las citas por video llamada (26%). La hipótesis es que al ser un subconjunto reducido de citas que tienen estos atributos no obtienen la importancia versus otras variables.

Por último, establecer que la pandemia tuvo un efecto tanto en el comportamiento del NSP como en los modelos utilizados para su predicción. En el NSP se evidencia la disminución de la cantidad de citas y del porcentaje de NSP

del periodo, mientras que en los modelos de predicción se destacan las disminuciones en el poder predictivos de los modelos, tanto aquellos entrenados con datos exclusivos de la pandemia (que tuvieron peores resultados) como en aquellos entrenados con datos históricos, demostrando una disminución en todas las métricas definidas para su evaluación.

Se puede establecer que los algoritmos de predicción son una opción real a la hora de optimizar procesos clínico-administrativos y con este trabajo se busca ser un aporte en la dirección correcta, permitiendo entregar herramientas digitales a los establecimientos de salud que permitan optimizar sus procesos y, finalmente, aportar en la salud pública.

9. BIBLIOGRAFÍA

1. Becerril-Montekio Víctor RJ de DMA. Sistema de salud de Chile. Salud pública Méx [Internet]. 2011 Ene [citado 2022 Feb 4];53. Disponible en: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0036-36342011000800009
2. Subsecretaría de Redes Asistenciales. Diseño del proceso clínico asistencial en la red pública de salud en Chile. Proceso de atención ambulatoria en red: Consulta ambulatoria en red. Serie Cuadernos de Redes N°28 [Internet]. 2018 [citado 2022 Feb 4]; Disponible en: http://www.bibliotecaminsal.cl/wp/wp-content/uploads/2016/03/28.pdf
3. Ayala M. Salud Pública: Concepto Y Descripción De La Red Asistencial Chilena [Internet]. Síntesis. 2017 [citado 2022 Mar 18]. Disponible en: https://sintesis.med.uchile.cl/index.php/profesionales/informacion-para-profesionales/medicina/condiciones-clinicas2/otorrinolaringologia/1344-7-01-3-039
4. Goldstein E. El sistema de salud en Chile y la Atención Primaria de Salud municipal: Marco para un debate sobre desmunicipalización [Internet]. 2018 Nov [citado 2022 Mar 18]. Disponible en: https://obtienearchivo.bcn.cl/obtienearchivo?id=repositorio/10221/26811/2/BCN_Gobernanza_salud_y_demunicip_para_reposit_final.pdf
5. FONASA. Tramos 2022 [Internet]. 2022 [citado 2022 Abr 2]. Disponible en: https://www.fonasa.cl/sites/fonasa/tramos
6. Marbouh D, Khaleel I, al Shanqiti K, al Tamimi M, Simsekler MCE, Ellahham S, et al. Evaluating the Impact of Patient No-Shows on Service Quality [Internet]. Risk Management and Healthcare Policy. 2020 [citado 2021 Jun 25]. Disponible en: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7280239/
7. Dantas LF, Fleck JL, Cyrino Oliveira FL, Hamacher S. No-shows in appointment scheduling – a systematic literature review [Internet]. Vol. 122, Health Policy. Elsevier Ireland Ltd; 2018 [citado 2021 Ene 8]. p. 412–21. Disponible en: https://www.sciencedirect.com/science/article/abs/pii/S0168851018300459?via%3Dihub

8. Salinas Rebolledo EA, de la Cruz Medías R, Bastías Silva G. Nonattendance to medical specialists' appointments and its relation to regional environmental and socioeconomic indicators in the Chilean public health system. Medwave [Internet]. 2014 Oct 15 [citado 2021 Jun 2];14(09). Disponible en: https://repositorio.uc.cl/xmlui/bitstream/handle/11534/46862/Nonattendance%20to%20medical%20specialists%C2%BF%20appointments%20and%20its%20relation%20to%20regional%20environmental%20and%20socioeconomic%20indicators%20in%20the%20Chilean%20public%20health%20system.pdf?sequence=1
9. Carreras-García D, Delgado-Gómez D, Llorente-Fernández F, Arribas-Gil A. Patient no-show prediction: A systematic literature review [Internet]. Vol. 22, Entropy. MDPI AG; 2020 [citado 2021 Ene 8]. Disponible en: https://www.mdpi.com/1099-4300/22/6/675/pdf
10. Kheirkhah P, Feng Q, Travis LM, Tavakoli-Tabasi S, Sharafkhaneh A. Prevalence, predictors and economic consequences of no-shows. BMC Health Services Research [Internet]. 2015 Dic 14 [citado 2021 Ago 27];16(1). Disponible en: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4714455/
11. Machado AT, Werneck MAF, Lucas SD, Abreu MHNG. Who did not appear? First dental visit absences in secondary care in a major Brazilian city: a cross-sectional study. Ciência & Saúde Coletiva [Internet]. 2015 Ene [citado 2021 Ago 27];20(1). Disponible en: https://www.scielo.br/j/csc/a/6dkKNp5nTLm8N5Chx4Ds9Nh/?lang=en
12. Torres O, Rothberg MB, Garb J, Ogunneye O, Onyema J, Higgins T. Risk Factor Model to Predict a Missed Clinic Appointment in an Urban, Academic, and Underserved Setting. Population Health Management [Internet]. 2015 Abr [citado 2021 Ago 27];18(2). Disponible en: https://www.liebertpub.com/doi/full/10.1089/pop.2014.0047
13. Kempny A, Diller GP, Dimopoulos K, Alonso-Gonzalez R, Uebing A, Li W, et al. Determinants of outpatient clinic attendance amongst adults with congenital heart disease and outcome. International Journal of Cardiology [Internet]. 2016 Ene [citado 2021 Ago 27];203. Disponible en: https://www.sciencedirect.com/science/article/abs/pii/S0167527315306471
14. Daggy J, Lawley M, Willis D, Thayer D, Suelzer C, DeLaurentis PC, et al. Using no-show modeling to improve clinic performance. Health Informatics

Journal [Internet]. 2010 Dic 7 [citado 2021 Ago 27];16(4). Disponible en: https://www.researchgate.net/publication/49738856_Using_no-show_modeling_to_improve_clinic_performance

15. Miller AJ, Chae E, Peterson E, Ko AB. Predictors of repeated “no-showing” to clinic appointments. American Journal of Otolaryngology [Internet]. 2015 May [citado 2021 Ago 27];36(3). Disponible en: https://www.sciencedirect.com/science/article/abs/pii/S019607091500037X
16. Dove HG, Schneider KC. The Usefulness of Patients?? Individual Characteristics in Predicting No-Shows in Outpatient Clinics. Medical Care [Internet]. 1981 Jul [citado 2021 Ago 27];19(7). Disponible en: https://pubmed.ncbi.nlm.nih.gov/7266121/
17. Huang Y, Hanauer DA. Patient No-Show Predictive Model Development using Multiple Data Sources for an Effective Overbooking Approach. Applied Clinical Informatics [Internet]. 2014 Dic 19 [citado 2021 Ago 27];05(03). Disponible en: https://www.thieme-connect.de/products/ejournals/pdf/10.4338/ACI-2014-04-RA-0026.pdf
18. Whiting PS, Greenberg SE, Thakore R v., Alamanda VK, Ehrenfeld JM, Obremskey WT, et al. What factors influence follow-up in orthopedic trauma surgery? Archives of Orthopaedic and Trauma Surgery [Internet]. 2015 Mar 24 [citado 2021 Ago 27];135(3). Disponible en: https://link.springer.com/article/10.1007/s00402-015-2151-8#citeas
19. Mugavero MJ, Lin HY, Allison JJ, Willig JH, Chang PW, Marler M, et al. Failure to Establish HIV Care: Characterizing the “No Show” Phenomenon. Clinical Infectious Diseases [Internet]. 2007 Jul 1 [citado 2021 Ago 27];45(1). Disponible en: https://academic.oup.com/cid/article/45/1/127/479588
20. Ministerio de Salud - Subsecretaría de Redes Asistenciales - División de Gestión de la Red Asistencial. Programa Nacional de Telesalud [Internet]. 2018 [citado 2021 Ene 12]. Disponible en: http://biblioteca.digital.gob.cl/bitstream/handle/123456789/3635/Programa%20Nacional%20de%20Telesalud.pdf?sequence=1&isAllowed=y
21. Ministerio de Salud. ¿Qué es Hospital Digital? [Internet]. 2021 [citado 2021 Jul 5]. Disponible en: https://www.hospitaldigital.gob.cl/hospital-digital/que-es-hospital-digital

22. Chavéz M. Hospital Digital enfrenta difícil puesta en marcha con menos consultas de las esperadas. La Segunda [Internet]. 2019 May 29 [citado 2021 Jul 1];4-undefined. Disponible en: http://cache-elastic.emol.com/2019/05/29/A/8V3JT9M8/all
23. el Naqa I, Murphy MJ. What Is Machine Learning? In: Machine Learning in Radiation Oncology [Internet]. Cham: Springer International Publishing; 2015 [citado 2021 Jul 5]. Disponible en: https://link.springer.com/chapter/10.1007/978-3-319-18305-3_1#citeas
24. Burkov A. The Hundred-Page Machine Learning Book. Ilustrada. Burkov A, editor. 2019.
25. World Health Organization. Listings of WHO's response to COVID-19 [Internet]. 2020 [citado 2021 Jun 16]. Disponible en: https://www.who.int/news/item/29-06-2020-covidtimeline
26. Veloso L. Ministerio de Salud confirma primer caso de Coronavirus en Chile: se trata de un médico. Radio Bío Bío [Internet]. 2020 [citado 2021 Jun 2]; Disponible en: https://www.biobiochile.cl/noticias/nacional/region-del-maule/2020/03/03/confirman-primer-caso-de-coronavirus-en-chile.shtml
27. Ministerio de Salud. Casos Confirmados en Chile COVID 19 [Internet]. 2021 [citado 2021 Ago 24]. Disponible en: https://www.minsal.cl/nuevo-coronavirus-2019-ncov/casos-confirmados-en-chile-covid-19/
28. Santos-Sánchez NF, Salas-Coronado R. Origin, structural characteristics, prevention measures, diagnosis and potential drugs to prevent and COVID-19. Medwave [Internet]. 2020 Sep 30 [citado 2021 Jun 30];20(08). Disponible en: https://www.medwave.cl/link.cgi/Medwave/Revisiones/RevisionClinica/8037.act
29. Organización Mnundial de la Salud. Preguntas y respuestas sobre la transmisión de la COVID-19 [Internet]. 2021 [citado 2022 Mar 26]. Disponible en: https://www.who.int/es/news-room/questions-and-answers/item/coronavirus-disease-covid-19-how-is-it-transmitted
30. Organización Mundial de la Salud. Actualización de la estrategia frente a la COVID-19 [Internet]. 14 de abril 2020. 2020 [citado 2021 Jul 5]. Disponible en: https://www.who.int/docs/default-source/coronaviruse/covid-strategy-update-14april2020_es.pdf?sfvrsn=86c0929d_10

31. Ministerio de Salud S de SP. Decreto 11 - Suspende garantía de oportunidad de las garantias explícitas en salud en los problemas de salud que indica [Internet]. 2020 [citado 2021 Jul 5]. Disponible en: https://www.bcn.cl/leychile/navegar?idNorma=1144149&idParte=0
32. Said C. Atenciones médicas caen 62% en junio: la mayor baja se registra en pediatría. 2020 Jun 22 [citado 2021 Ago 26]; Disponible en: https://www.latercera.com/nacional/noticia/atenciones-medicas-caen-62-en-junio-la-mayor-baja-se-registra-en-pediatria/Y2YGCB2WNFFLFEGAUHMIC6GUXE/
33. Kemp MT, Liesman DR, Brown CS, Williams AM, Biesterveld BE, Wakam GK, et al. Factors Associated with Increased Risk of Patient No-Show in Telehealth and Traditional Surgery Clinics. J Am Coll Surg [Internet]. 2020 Dic 1 [citado 2021 May 20];231(6):695–702. Disponible en: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7470818/
34. Schober P, Boer C, Schwarte LA. Correlation Coefficients. Anesthesia & Analgesia [Internet]. 2018 May [citado 2021 Dic 7];126(5):1763–8. Disponible en: https://journals.lww.com/anesthesia-analgesia/fulltext/2018/05000/correlation_coefficients__appropriate_use_and.50.aspx
35. Ramirez H, Villena F, Dunstan J, Riquelme V, Hoyos JP, Madariaga J, et al. Predicting no-show appointments in a pediatric hospital in Chile using machine learning. 2020.
36. Chen J, Goldstein IH, Lin WC, Chiang MF, Hribar MR. Application of Machine Learning to Predict Patient No-Shows in an Academic Pediatric Ophthalmology Clinic. AMIA Annu Symp Proc [Internet]. 2020 [citado 2021 Dic 19];2020:293–302. Disponible en: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8075453/
37. Sperandei S. Understanding logistic regression analysis. Biochemia Medica [Internet]. 2014 [citado 2022 Mar 4];12–8. Disponible en: https://hrcak.srce.hr/115732
38. Breiman L. Random Forests. Machine Learning [Internet]. 2001 [citado 2022 Mar 4];45(1):5–32. Disponible en: https://link.springer.com/article/10.1023/A:1010933404324

39. Liu Y, Wang Y, Zhang J. New Machine Learning Algorithm: Random Forest. In 2012 [citado 2022 Mar 4]. p. 246–52. Disponible en: https://link.springer.com/chapter/10.1007/978-3-642-34062-8_32
40. Chen T, Guestrin C. XGBoost. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Internet]. New York, NY, USA: ACM; 2016 [citado 2022 Mar 4]. p. 785–94. Disponible en: https://arxiv.org/abs/1603.02754
41. Natekin A, Knoll A. Gradient boosting machines, a tutorial. Frontiers in Neurorobotics. 2013;7.
42. Potdar K, S. T, D. C. A Comparative Study of Categorical Variable Encoding Techniques for Neural Network Classifiers. International Journal of Computer Applications [Internet]. 2017 Oct 17 [citado 2021 Dic 18];175(4):7–9. Disponible en: https://www.researchgate.net/publication/320465713_A_Comparative_Study_of_Categorical_Variable_Encoding_Techniques_for_Neural_Network_Classifiers
43. Denis G, Hermosilla M, Aracena C, Sanchez R, Gonzales N, Pombo C. Uso responsable de IA para política pública: manual de formulación de proyectos [Internet]. Washington; 2021 Sep [citado 2022 Mar 4]. Disponible en: https://publications.iadb.org/publications/spanish/document/Uso-responsable-de-IA-para-politica-publica-manual-de-formulacion-de-proyectos.pdf
44. FONASA. Normativa y Aranceles Modalidad de Atención Institucional (MAI) [Internet]. 2022 [citado 2022 Mar 12]. Disponible en: https://www.fonasa.cl/sites/Satellite;jsessionid=ufxK0PR14nDblws9jb1JP6Ql712d7TAMjZ6y8DD02gGBt41DOfSw!730046559!444791042:hg4VikvuMH5wNz93rMfjSGDmV-x7YSSo?c=Page&cid=1520002044318&pagename=Fonasa2019%2FPage%2FF2_ContenidoDerecha

10. ANEXOS

10.1. Detalle análisis variables numéricas por periodos de tiempo

Variable	Periodo	Mes	N° Filas	Comp	Prom	DS	Mín	Máx	Perc 25	Perc 50	Perc 75	Perc 99	VMC
Edad	Control	11	125.499	100%	45,51	26,32	2	116	20	49	68	91	3
		12	107.328	100%	47,17	25,93	2	105	24	51	69	91	3
		1	127.022	100%	47,29	25,89	2	106	24	51	69	91	3
		2	99.028	100%	47,66	25,64	2	106	26	52	69	91	3
		3	126.119	100%	46,82	26,02	2	106	24	51	69	91	2
		4	126.444	100%	46,75	25,88	0	107	24	51	68	91	2
		5	125.200	100%	46,76	25,56	2	106	24	51	68	91	2
		6	121.770	100%	46,59	25,73	2	106	24	50	68	91	2
	Pandemia	11	112.770	100%	46,55	25,74	1	106	24	51	68	91	2
		12	108.630	100%	46,04	25,90	1	106	22	50	68	91	2
		1	125.259	100%	46,59	25,64	1	121	24	51	68	91	2
		2	100.409	100%	46,92	25,48	1	121	26	51	68	91	2
		3	100.805	100%	46,52	25,50	1	121	25	51	68	91	60
		4	44.761	100%	47,97	24,33	0	107	28	52	68	91	60
		5	39.772	100%	45,68	24,35	1	105	26	48	65	90	1
		6	41.604	100%	44,55	23,99	1	110	26	46	64	90	1

Variable	Periodo	Mes	N° Filas	Comp	Prom	DS	Mín	Máx	Perc 25	Perc 50	Perc 75	Perc 99	VMC
Citas previas 30 días	Control	11	125.499	100%	2,67	3,81	0	45	0	1	3	18	0
		12	107.328	100%	1,14	2,20	0	51	0	0	1	10	0
		1	127.022	100%	2,60	3,85	0	53	0	1	3	17	0
		2	99.028	100%	2,77	3,91	0	58	0	1	4	18	0
		3	126.119	100%	2,76	4,10	0	59	0	1	4	18	0
		4	126.444	100%	2,67	3,92	0	49	0	1	3	19	0
		5	125.200	100%	2,86	4,11	0	51	0	1	4	20	0
		6	121.770	100%	2,86	4,21	0	50	0	1	4	20	0
	Pandemia	11	112.770	100%	2,71	3,69	0	38	0	1	4	17	0
		12	108.630	100%	1,29	2,50	0	39	0	0	2	12	0
		1	125.259	100%	2,71	3,96	0	47	0	1	4	18	0
		2	100.409	100%	2,69	3,90	0	61	0	1	4	18	0
		3	100.805	100%	1,60	2,83	0	58	0	1	2	13	0
		4	44.761	100%	2,07	3,29	0	32	0	1	3	15	0
		5	39.772	100%	2,34	3,47	0	34	0	1	3	16	0
		6	41.604	100%	2,43	3,50	0	30	0	1	3	16	0

Variable	Periodo	Mes	N° Filas	Comp	Prom	DS	Mín	Máx	Perc 25	Perc 50	Perc 75	Perc 99	VMC
Citas Previas 60 días	Control	11	125.499	100%	3,42	4,88	0	52	0	2	4	23	0
		12	107.328	100%	1,14	2,20	0	51	0	0	1	10	0
		1	127.022	100%	4,79	6,60	0	91	1	3	6	31	0
		2	99.028	100%	5,14	7,04	0	114	1	3	7	33	0
		3	126.119	100%	4,95	7,11	0	106	1	3	6	34	0
		4	126.444	100%	5,05	7,06	0	73	1	3	6	35	0
		5	125.200	100%	5,30	7,39	0	78	1	3	7	36	0
		6	121.770	100%	5,24	7,45	0	76	1	3	7	37	0
	Pandemia	11	112.770	100%	3,64	5,04	0	55	0	2	5	24	0
		12	108.630	100%	1,29	2,50	0	39	0	0	2	12	0
		1	125.259	100%	4,87	6,78	0	107	1	2	6	33	0
		2	100.409	100%	4,04	5,92	0	110	0	2	5	28	0
		3	100.805	100%	2,67	4,69	0	67	0	1	3	24	0
		4	44.761	100%	3,83	5,93	0	57	0	2	5	28	0
		5	39.772	100%	4,33	6,16	0	52	0	2	6	29	0
		6	41.604	100%	4,74	6,67	0	63	0	2	6	32	0

Variable	Periodo	Mes	N° Filas	Comp	Prom	DS	Mín	Máx	Perc 25	Perc 50	Perc 75	Perc 99	VMC
Citas previas 90 días	Control	11	125.499	100%	3,42	4,88	0	52	0	2	4	23	0
		12	107.328	100%	1,14	2,20	0	51	0	0	1	10	0
		1	127.022	100%	6,85	9,25	0	117	1	4	9	44	0
		2	99.028	100%	7,12	9,73	0	142	1	4	9	47	0
		3	126.119	100%	6,99	9,87	0	132	1	4	9	49	0
		4	126.444	100%	7,12	9,85	0	87	1	4	9	50	0
		5	125.200	100%	7,41	10,26	0	98	1	4	9	51	0
		6	121.770	100%	7,32	10,30	0	103	1	4	9	51	0
	Pandemia	11	112.770	100%	3,64	5,04	0	55	0	2	5	24	0
		12	108.630	100%	1,29	2,50	0	39	0	0	2	12	0
		1	125.259	100%	6,01	8,40	0	134	1	3	8	42	0
		2	100.409	100%	5,00	7,55	0	117	1	2	6	38	0
		3	100.805	100%	3,63	6,55	0	72	0	1	4	34	0
		4	44.761	100%	5,42	8,26	0	81	0	2	7	39	0
		5	39.772	100%	6,36	9,09	0	81	1	3	8	45	0
		6	41.604	100%	6,94	9,82	0	87	1	3	9	49	0

Variable	Periodo	Mes	N° Filas	Comp	Prom	DS	Mín	Máx	Perc 25	Perc 50	Perc 75	Perc 99	VMC
Citas previas 120 días	Control	11	125.499	100%	3,42	4,88	0	52	0	2	4	23	0
		12	107.328	100%	1,14	2,20	0	51	0	0	1	10	0
		1	127.022	100%	8,65	11,64	0	137	2	5	11	56	0
		2	99.028	100%	9,09	12,36	0	158	2	5	11	60	0
		3	126.119	100%	8,89	12,51	0	145	1	5	11	63	0
		4	126.444	100%	9,05	12,49	0	113	2	5	11	63	0
		5	125.200	100%	9,33	12,91	0	115	2	5	11	64	0
		6	121.770	100%	9,08	12,68	0	130	2	5	11	63	0
	Pandemia	11	112.770	100%	3,64	5,04	0	55	0	2	5	24	0
		12	108.630	100%	1,29	2,50	0	39	0	0	2	12	0
		1	125.259	100%	6,84	9,88	0	134	1	3	9	49	0
		2	100.409	100%	5,89	9,24	0	117	1	3	7	47	0
		3	100.805	100%	4,55	8,26	0	94	0	2	5	43	0
		4	44.761	100%	7,15	10,83	0	97	1	3	9	53	0
		5	39.772	100%	8,43	12,13	0	99	1	4	11	62	0
		6	41.604	100%	9,01	12,76	0	116	1	4	12	62	0

Variable	Periodo	Mes	N° Filas	Comp	Prom	DS	Mín	Máx	Perc 25	Perc 50	Perc 75	Perc 99	VMC
Citas Previas 365 días	Control	11	125.499	100%	3,42	4,88	0	52	0	2	4	23	0
		12	107.328	100%	1,14	2,20	0	51	0	0	1	10	0
		1	127.022	100%	19,04	25,74	0	333	3	10	24	125	0
		2	99.028	100%	18,27	25,21	0	304	3	10	23	123	0
		3	126.119	100%	16,65	23,68	0	282	3	9	21	119	0
		4	126.444	100%	15,64	21,93	0	251	3	8	19	108	0
		5	125.200	100%	14,56	20,24	0	215	3	8	18	102	0
		6	121.770	100%	12,89	17,92	0	202	2	7	16	89	0
	Pandemia	11	112.770	100%	3,64	5,04	0	55	0	2	5	24	0
		12	108.630	100%	1,29	2,50	0	39	0	0	2	12	0
		1	125.259	100%	13,66	21,38	0	247	2	6	16	108	0
		2	100.409	100%	12,54	20,39	0	230	1	6	14	102	0
		3	100.805	100%	10,67	18,33	0	218	1	4	12	91	0
		4	44.761	100%	14,42	21,28	0	206	2	7	17	109,4	0
		5	39.772	100%	14,68	20,87	0	190	2	7	18	106	0
		6	41.604	100%	13,76	19,32	0	179	2	7	17	96	0

Variable	Periodo	Mes	N° Filas	Comp	Prom	DS	Mín	Máx	Perc 25	Perc 50	Perc 75	Perc 99	VMC
NSP 30 días	Control	11	125.499	67,5%	0,14	0,28	0	1	0	0	0,17	1	0
		12	107.328	42,1%	0,13	0,29	0	1	0	0	0	1	0
		1	127.022	67,2%	0,13	0,26	0	1	0	0	0,11	1	0
		2	99.028	68,9%	0,13	0,27	0	1	0	0	0,13	1	0
		3	126.119	67,8%	0,13	0,27	0	1	0	0	0,13	1	0
		4	126.444	67,2%	0,14	0,27	0	1	0	0	0,17	1	0
		5	125.200	68,8%	0,14	0,28	0	1	0	0	0,17	1	0
		6	121.770	68,8%	0,14	0,27	0	1	0	0	0,17	1	0
	Pandemia	11	112.770	68,5%	0,16	0,29	0	1	0	0	0,20	1	0
		12	108.630	43,8%	0,15	0,30	0	1	0	0	0,11	1	0
		1	125.259	66,7%	0,13	0,27	0	1	0	0	0,14	1	0
		2	100.409	67,5%	0,14	0,27	0	1	0	0	0,17	1	0
		3	100.805	51,6%	0,10	0,24	0	1	0	0	0	1	0
		4	44.761	56,0%	0,08	0,21	0	1	0	0	0	1	0
		5	39.772	60,6%	0,09	0,23	0	1	0	0	0	1	0
		6	41.604	62,1%	0,08	0,21	0	1	0	0	0	1	0

Variable	Periodo	Mes	N° Filas	Comp	Prom	DS	Mín	Máx	Perc 25	Perc 50	Perc 75	Perc 99	VMC
NSP 60 días	Control	11	125.499	72,2%	0,14	0,27	0	1	0	0	0,20	1	0
		12	107.328	42,1%	0,13	0,29	0	1	0	0	0	1	0
		1	127.022	79,1%	0,13	0,25	0	1	0	0	0,17	1	0
		2	99.028	80,2%	0,13	0,25	0	1	0	0	0,19	1	0
		3	126.119	78,9%	0,14	0,25	0	1	0	0	0,19	1	0
		4	126.444	79,6%	0,14	0,25	0	1	0	0	0,20	1	0
		5	125.200	80,1%	0,15	0,25	0	1	0	0	0,20	1	0
		6	121.770	79,5%	0,14	0,25	0	1	0	0	0,20	1	0
	Pandemia	11	112.770	73,9%	0,16	0,28	0	1	0	0	0,22	1	0
		12	108.630	43,8%	0,15	0,30	0	1	0	0	0,11	1	0
		1	125.259	78,8%	0,14	0,25	0	1	0	0	0,20	1	0
		2	100.409	74,7%	0,12	0,25	0	1	0	0	0,15	1	0
		3	100.805	60,1%	0,09	0,22	0	1	0	0	0	1	0
		4	44.761	67,3%	0,08	0,20	0	1	0	0	0	1	0
		5	39.772	72,3%	0,08	0,21	0	1	0	0	0	1	0
		6	41.604	74,3%	0,08	0,19	0	1	0	0	0,04	1	0

Variable	Periodo	Mes	N° Filas	Comp	Prom	DS	Mín	Máx	Perc 25	Perc 50	Perc 75	Perc 99	VMC
NSP 90 días	Control	11	125.499	72,2%	0,14	0,27	0	1	0	0	0,20	1	0
		12	107.328	42,1%	0,13	0,29	0	1	0	0	0	1	0
		1	127.022	83,8%	0,14	0,23	0	1	0	0	0,20	1	0
		2	99.028	83,9%	0,14	0,24	0	1	0	0	0,20	1	0
		3	126.119	83,1%	0,14	0,24	0	1	0	0	0,20	1	0
		4	126.444	83,8%	0,15	0,24	0	1	0	0	0,21	1	0
		5	125.200	84,1%	0,15	0,24	0	1	0	0	0,20	1	0
		6	121.770	83,6%	0,14	0,24	0	1	0	0	0,20	1	0
	Pandemia	11	112.770	73,9%	0,16	0,28	0	1	0	0	0,22	1	0
		12	108.630	43,8%	0,15	0,30	0	1	0	0	0,11	1	0
		1	125.259	81,4%	0,14	0,24	0	1	0	0	0,19	1	0
		2	100.409	76,9%	0,12	0,24	0	1	0	0	0,14	1	0
		3	100.805	63,7%	0,09	0,22	0	1	0	0	0,06	1	0
		4	44.761	72,6%	0,08	0,19	0	1	0	0	0,06	1	0
		5	39.772	77,3%	0,08	0,20	0	1	0	0	0,07	1	0
		6	41.604	79,8%	0,08	0,19	0	1	0	0	0,08	1	0

Variable	Periodo	Mes	N° Filas	Comp	Prom	DS	Mín	Máx	Perc 25	Perc 50	Perc 75	Perc 99	VMC
NSP 120 días	Control	11	125.499	72,2%	0,14	0,27	0	1	0	0	0,20	1	0
		12	107.328	42,1%	0,13	0,29	0	1	0	0	0	1	0
		1	127.022	86,2%	0,14	0,23	0	1	0	0	0,20	1	0
		2	99.028	86,4%	0,14	0,23	0	1	0	0	0,20	1	0
		3	126.119	85,7%	0,14	0,23	0	1	0	0	0,20	1	0
		4	126.444	86,2%	0,15	0,23	0	1	0	0	0,21	1	0
		5	125.200	86,6%	0,15	0,23	0	1	0	0,01	0,21	1	0
		6	121.770	86,0%	0,15	0,23	0	1	0	0	0,20	1	0
	Pandemia	11	112.770	73,9%	0,16	0,28	0	1	0	0	0,22	1	0
		12	108.630	43,8%	0,15	0,30	0	1	0	0	0,11	1	0
		1	125.259	82,4%	0,13	0,23	0	1	0	0	0,17	1	0
		2	100.409	78,3%	0,12	0,23	0	1	0	0	0,14	1	0
		3	100.805	67,0%	0,09	0,21	0	1	0	0	0,07	1	0
		4	44.761	77,8%	0,08	0,19	0	1	0	0	0,07	1	0
		5	39.772	81,5%	0,09	0,19	0	1	0	0	0,09	1	0
		6	41.604	83,6%	0,08	0,18	0	1	0	0	0,09	1	0

Variable	Periodo	Mes	N° Filas	Comp	Prom	DS	Mín	Máx	Perc 25	Perc 50	Perc 75	Perc 99	VMC
NSP 365 días	Control	11	125.499	72,2%	0,14	0,27	0	1	0	0	0,20	1	0
		12	107.328	42,1%	0,13	0,29	0	1	0	0	0	1	0
		1	127.022	91,7%	0,15	0,21	0	1	0	0,07	0,21	1	0
		2	99.028	91,2%	0,15	0,21	0	1	0	0,07	0,22	1	0
		3	126.119	90,4%	0,15	0,21	0	1	0	0,07	0,22	1	0
		4	126.444	90,6%	0,16	0,22	0	1	0	0,07	0,23	1	0
		5	125.200	90,4%	0,16	0,22	0	1	0	0,07	0,24	1	0
		6	121.770	89,1%	0,16	0,23	0	1	0	0,06	0,23	1	0
	Pandemia	11	112.770	73,9%	0,16	0,28	0	1	0	0	0,22	1	0
		12	108.630	43,8%	0,15	0,30	0	1	0	0	0,11	1	0
		1	125.259	87,8%	0,12	0,21	0	1	0	0,01	0,17	1	0
		2	100.409	85,5%	0,11	0,21	0	1	0	0	0,14	1	0
		3	100.805	80,3%	0,10	0,20	0	1	0	0	0,13	1	0
		4	44.761	87,1%	0,09	0,18	0	1	0	0	0,13	1	0
		5	39.772	87,5%	0,10	0,18	0	1	0	0	0,13	1	0
		6	41.604	88,1%	0,10	0,18	0	1	0	0	0,13	1	0

Variable	Periodo	Mes	N° Filas	Comp	Prom	DS	Mín	Máx	PERC_25	PERC_50	PERC_75	PERC_99	VMC
Dif . Fecha Cita	Control	11	125.499	100%	-16,42	21,80	-210	93	-26	-9	-1	8	0
		12	107.328	100%	-15,95	21,68	-276	71	-23	-9	-1	9	0
		1	127.022	100%	-13,62	16,19	-179	336	-23	-8	-1	12	0
		2	99.028	100%	-14,34	17,30	-185	256	-23	-8	-1	6	0
		3	126.119	100%	-14,83	18,51	-175	235	-23	-8	-1	6	0
		4	126.444	100%	-14,19	16,13	-147	291	-23	-9	-1	6	0
		5	125.200	100%	-15,54	17,89	-182	336	-26	-9	-1	5	0
		6	121.770	100%	-15,87	19,92	-182	154	-25	-8	-1	5	0
	Pandemia	11	112.770	100%	-16,25	20,67	-210	417	-28	-8	-1	9	0
		12	108.630	100%	-15,23	20,86	-238	344	-22	-8	0	12	0
		1	125.259	100%	-12,80	16,09	-143	389	-21	-7	0	8	0
		2	100.409	100%	-13,54	16,56	-154	364	-22	-7	0	6	0
		3	100.805	100%	-16,43	20,25	-182	345	-28	-9	-1	8	0
		4	44.761	100%	-12,80	21,80	-144	169	-26	-3	0	12	0
		5	39.772	100%	-7,71	19,63	-175	381	-8	-1	0	25	0
		6	41.604	100%	-6,95	19,05	-210	224	-8	-1	0	46	0

10.2. Detalle análisis variables categóricas por periodos de tiempo

Variable	Periodo	Mes	Completitud	Cardinalidad	VMC	% VMC
Sexo	Control	11	1.0	2	Mujer	58%
		12	1.0	2	Mujer	58%
		1	1.0	3	Mujer	58%
		2	1.0	3	Mujer	58%
		3	1.0	2	Mujer	58%
		4	1.0	3	Mujer	58%
		5	1.0	3	Mujer	59%
		6	1.0	3	Mujer	58%
	Pandemia	11	1.0	3	Mujer	58%
		12	1.0	3	Mujer	57%
		1	1.0	3	Mujer	58%
		2	1.0	3	Mujer	58%
		3	1.0	3	Mujer	57%
		4	1.0	3	Mujer	57%
		5	1.0	3	Mujer	57%
		6	1.0	3	Mujer	57%

Variable	Periodo	Mes	Completitud	Cardinalidad	VMC	% VMC
Pueblo originario	Control	11	56%	14	Ninguna	79%
		12	57%	14	Ninguna	79%
		1	59%	14	Ninguna	79%
		2	60%	14	Ninguna	79%
		3	61%	14	Ninguna	80%
		4	66%	14	Ninguna	82%
		5	69%	14	Ninguna	83%
		6	72%	14	Ninguna	82%
	Pandemia	11	77%	14	Ninguna	80%
		12	77%	14	Ninguna	80%
		1	77%	14	Ninguna	79%
		2	78%	14	Ninguna	79%
		3	77%	14	Ninguna	79%
		4	76%	14	Ninguna	78%
		5	76%	14	Ninguna	77%
		6	75%	14	Ninguna	77%

Variable	Periodo	Mes	Completitud	Cardinalidad	VMC	% VMC
Nacionalidad	Control	11	100%	32	Chile	98%
		12	100%	29	Chile	97%
		1	100%	31	Chile	98%
		2	100%	30	Chile	97%
		3	100%	28	Chile	98%
		4	100%	28	Chile	98%
		5	100%	30	Chile	98%
		6	100%	29	Chile	98%
	Pandemia	11	100%	27	Chile	98%
		12	100%	26	Chile	98%
		1	100%	29	Chile	98%
		2	100%	25	Chile	97%
		3	100%	30	Chile	98%
		4	100%	26	Chile	97%
		5	100%	21	Chile	97%
		6	100%	22	Chile	97%

Variable	Periodo	Mes	Completitud	Cardinalidad	VMC	% VMC
Previsión	Control	11	100%	5	FONASA	94%
		12	100%	5	FONASA	94%
		1	100%	5	FONASA	94%
		2	100%	5	FONASA	94%
		3	100%	5	FONASA	94%
		4	100%	5	FONASA	95%
		5	100%	5	FONASA	94%
		6	100%	5	FONASA	94%
	Pandemia	11	100%	5	FONASA	94%
		12	100%	5	FONASA	94%
		1	100%	5	FONASA	94%
		2	100%	5	FONASA	94%
		3	100%	5	FONASA	94%
		4	100%	5	FONASA	94%
		5	100%	5	FONASA	93%
		6	100%	5	FONASA	93%

Variable	Periodo	Mes	Completitud	Cardinalidad	VMC	% VMC
Tipo Profesional	Control	11	83%	21	Médico	54%
		12	85%	21	Médico	54%
		1	86%	21	Médico	56%
		2	84%	21	Médico	54%
		3	84%	20	Médico	55%
		4	87%	20	Médico	54%
		5	87%	20	Médico	54%
		6	86%	21	Médico	55%
	Pandemia	11	93%	20	Médico	57%
		12	93%	20	Médico	56%
		1	93%	20	Médico	55%
		2	92%	21	Médico	56%
		3	93%	20	Médico	62%
		4	89%	19	Médico	66%
		5	89%	20	Médico	62%
		6	90%	20	Médico	62%

Variable	Periodo	Mes	Completitud	Cardinalidad	VMC*	% VMC
Prestación	Control	11	100%	268	Cons integral de esp	15%
		12	100%	255	Cons integral de esp	16%
		1	100%	285	Cons integral de esp	17%
		2	100%	280	Cons integral de esp	16%
		3	100%	293	Cons integral de esp	16%
		4	100%	254	Cons integral de esp	16%
		5	100%	183	Cons integral de esp	16%
		6	100%	177	Cons integral de esp	17%
	Pandemia	11	100%	184	Cons integral de esp	18%
		12	100%	191	Cons integral de esp	17%
		1	100%	179	Cons integral de esp	17%
		2	100%	182	Cons integral de esp	18%
		3	100%	212	Cons integral de esp	18%
		4	100%	149	Cons integral de esp	19%
		5	100%	137	Cons integral de esp	18%
		6	100%	131	Cons integral de esp	18%

*Cons integral de esp corresponde a la prestación “Consulta Integral de Especialidades en Medic Interna y Subesp, Oftalmo, Neurolo, Oncologia en CDT”

Variable	Periodo	Mes	Completitud	Cardinalidad	VMC	% VMC
Comuna	Control	11	100%	243	San Bernardo	21%
		12	100%	226	San Bernardo	20%
		1	100%	229	San Bernardo	21%
		2	100%	222	San Bernardo	21%
		3	100%	232	San Bernardo	21%
		4	100%	237	San Bernardo	21%
		5	100%	234	San Bernardo	21%
		6	100%	229	San Bernardo	20%
	Pandemia	11	100%	225	San Bernardo	21%
		12	100%	212	San Bernardo	22%
		1	100%	234	San Bernardo	21%
		2	100%	211	San Bernardo	21%
		3	100%	210	San Bernardo	22%
		4	100%	163	San Bernardo	24%
		5	100%	149	San Bernardo	24%
		6	100%	154	San Bernardo	24%

Variable	Periodo	Mes	Completitud	Cardinalidad	VMC	% VMC
Especialidad	Control	11	66%	76	Psiquiatría Adulto	8%
		12	68%	76	Med. Interna	8%
		1	69%	78	Med. Interna	10%
		2	67%	77	Med. Interna	9%
		3	67%	78	Med. Interna	8%
		4	69%	78	Med. Interna	8%
		5	68%	78	Med. Interna	8%
		6	68%	78	Med. Interna	8%
	Pandemia	11	74%	78	Med. Interna	12%
		12	73%	78	Med. Interna	11%
		1	73%	79	Med. Interna	10%
		2	72%	78	Med. Interna	11%
		3	76%	78	Med. Interna	12%
		4	78%	71	Med. Interna	14%
		5	75%	69	Psiquiatría Adulto	17%
		6	76%	70	Psiquiatría Adulto	19%

Variable	Periodo	Mes	Completitud	Cardinalidad	VMC	% VMC
Descripción Cita	Control	11	6%	62	Consulta Repetida	33%
		12	6%	60	Consulta Repetida	34%
		1	5%	60	Consulta Repetida	30%
		2	5%	63	Consulta Repetida	35%
		3	5%	59	Consulta Repetida	34%
		4	5%	64	Consulta Repetida	39%
		5	5%	58	Consulta Repetida	36%
		6	5%	64	Consulta Repetida	37%
	Pandemia	11	6%	59	Consulta Repetida	47%
		12	5%	57	Consulta Repetida	47%
		1	5%	57	Consulta Repetida	39%
		2	5%	55	Consulta Repetida	40%
		3	6%	43	Consulta Repetida	35%
		4	6%	23	Consulta Repetida	32%
		5	7%	22	Consulta Repetida	29%
		6	8%	20	Consulta de Urgencia	32%

Variable	Periodo	Mes	Completitud	Cardinalidad	VMC	% VMC
Establecimiento	Control	11	100%	4	H. Barros Luco Trudeau	53%
		12	100%	4	H. Barros Luco Trudeau	55%
		1	100%	4	H. Barros Luco Trudeau	55%
		2	100%	4	H. Barros Luco Trudeau	56%
		3	100%	4	H. Barros Luco Trudeau	54%
		4	100%	4	H. Barros Luco Trudeau	55%
		5	100%	4	H. Barros Luco Trudeau	57%
		6	100%	4	H. Barros Luco Trudeau	56%
	Pandemia	11	100%	4	H. Barros Luco Trudeau	52%
		12	100%	4	H. Barros Luco Trudeau	51%
		1	100%	4	H. Barros Luco Trudeau	54%
		2	100%	4	H. Barros Luco Trudeau	54%
		3	100%	4	H. Barros Luco Trudeau	51%
		4	100%	4	H. Barros Luco Trudeau	43%
		5	100%	4	H. Barros Luco Trudeau	44%
		6	100%	4	H.l Barros Luco Trudeau	46%

Variable	Periodo	Mes	Completitud	Cardinalidad	VMC	% VMC
Comuna establecimiento	Control	11	100%	3	San Miguel	72%
		12	100%	3	San Miguel	72%
		1	100%	3	San Miguel	71%
		2	100%	3	San Miguel	71%
		3	100%	3	San Miguel	70%
		4	100%	3	San Miguel	71%
		5	100%	3	San Miguel	72%
		6	100%	3	San Miguel	73%
	Pandemia	11	100%	3	San Miguel	68%
		12	100%	3	San Miguel	68%
		1	100%	3	San Miguel	69%
		2	100%	3	San Miguel	69%
		3	100%	3	San Miguel	67%
		4	100%	3	San Miguel	55%
		5	100%	3	San Miguel	58%
		6	100%	3	San Miguel	60%

10.3. Detalle correlaciones por periodos de tiempo

Variable	Periodo	Mes	Pearson	Spearman
Edad	Control	11	-0,090	-0,090
		12	-0,120	-0,116
		1	-0,100	-0,101
		2	-0,100	-0,099
		3	-0,100	-0,097
		4	-0,100	-0,103
		5	-0,090	-0,094
		6	-0,100	-0,102
	Pandemia	11	-0,120	-0,119
		12	-0,130	-0,125
		1	-0,100	-0,099
		2	-0,100	-0,101
		3	-0,070	-0,070
		4	-0,020	-0,023
		5	-0,030	-0,029
		6	-0,050	-0,046

Variable	Periodo	Mes	Pearson	Spearman
Citas Previas 30 días	Control	11	0,000	-0,069
		12	-0,052	-0,051
		1	-0,054	-0,068
		2	-0,065	-0,082
		3	-0,048	-0,060
		4	-0,046	-0,059
		5	-0,050	-0,066
		6	-0,045	-0,064
	Pandemia	11	-0,084	-0,103
		12	-0,038	-0,044
		1	-0,061	-0,082
		2	-0,057	-0,078
		3	-0,048	-0,059
		4	-0,022	-0,011
		5	-0,065	-0,065
		6	-0,058	-0,066

Variable	Periodo	Mes	Pearson	Spearman
Citas Previas 60 días	Control	11	-0,067	-0,078
		12	-0,052	-0,051
		1	-0,059	-0,082
		2	-0,061	-0,085
		3	-0,050	-0,073
		4	-0,050	-0,071
		5	-0,053	-0,078
		6	-0,046	-0,073
	Pandemia	11	-0,080	-0,108
		12	-0,038	-0,044
		1	-0,062	-0,091
		2	-0,058	-0,088
		3	-0,050	-0,072
		4	-0,025	-0,024
		5	-0,067	-0,082
		6	-0,061	-0,078

Variable	Periodo	Mes	Pearson	Spearman
Citas Previas 90 días	Control	11	-0,067	-0,078
		12	-0,052	-0,051
		1	-0,058	-0,085
		2	-0,061	-0,089
		3	-0,051	-0,078
		4	-0,051	-0,076
		5	-0,052	-0,081
		6	-0,045	-0,077
	Pandemia	11	-0,080	-0,108
		12	-0,038	-0,044
		1	-0,062	-0,095
		2	-0,055	-0,090
		3	-0,046	-0,077
		4	-0,025	-0,030
		5	-0,069	-0,087
		6	-0,061	-0,083

Variable	Periodo	Mes	Pearson	Spearman
Citas Previas 120 días	Control	11	-0,067	-0,078
		12	-0,052	-0,051
		1	-0,058	-0,090
		2	-0,062	-0,094
		3	-0,051	-0,082
		4	-0,050	-0,079
		5	-0,051	-0,083
		6	-0,045	-0,079
	Pandemia	11	-0,080	-0,108
		12	-0,038	-0,044
		1	-0,059	-0,096
		2	-0,050	-0,090
		3	-0,046	-0,083
		4	-0,026	-0,038
		5	-0,067	-0,092
		6	-0,060	-0,086

Variable	Periodo	Mes	Pearson	Spearman
Citas Previas 365 días	Control	11	-0,067	-0,078
		12	-0,052	-0,051
		1	-0,056	-0,097
		2	-0,059	-0,100
		3	-0,049	-0,086
		4	-0,048	-0,082
		5	-0,051	-0,087
		6	-0,046	-0,083
	Pandemia	11	-0,080	-0,108
		12	-0,038	-0,044
		1	-0,043	-0,092
		2	-0,040	-0,088
		3	-0,041	-0,088
		4	-0,029	-0,055
		5	-0,063	-0,093
		6	-0,059	-0,091

Variable	Periodo	Mes	Pearson	Spearman
NSP 30 días	Control	11	0,175	0,158
		12	0,168	0,152
		1	0,163	0,146
		2	0,158	0,138
		3	0,155	0,137
		4	0,161	0,142
		5	0,181	0,159
		6	0,178	0,157
	Pandemia	11	0,182	0,158
		12	0,175	0,161
		1	0,155	0,141
		2	0,173	0,151
		3	0,093	0,088
		4	0,148	0,127
		5	0,177	0,140
		6	0,154	0,126

Variable	Periodo	Mes	Pearson	Spearman
NSP 60 días	Control	11	0,182	0,161
		12	0,168	0,152
		1	0,179	0,152
		2	0,169	0,140
		3	0,165	0,142
		4	0,173	0,150
		5	0,195	0,167
		6	0,187	0,161
	Pandemia	11	0,185	0,159
		12	0,175	0,161
		1	0,168	0,148
		2	0,165	0,141
		3	0,091	0,084
		4	0,144	0,122
		5	0,181	0,140
		6	0,156	0,126

Variable	Periodo	Mes	Pearson	Spearman
NSP 90 días	Control	11	0,182	0,161
		12	0,168	0,152
		1	0,186	0,154
		2	0,176	0,145
		3	0,176	0,148
		4	0,180	0,155
		5	0,199	0,168
		6	0,189	0,163
	Pandemia	11	0,185	0,159
		12	0,175	0,161
		1	0,165	0,144
		2	0,163	0,139
		3	0,093	0,086
		4	0,137	0,120
		5	0,181	0,134
		6	0,160	0,121

Variable	Periodo	Mes	Pearson	Spearman
NSP 120 días	Control	11	0,182	0,161
		12	0,168	0,152
		1	0,190	0,156
		2	0,182	0,150
		3	0,179	0,150
		4	0,186	0,159
		5	0,206	0,175
		6	0,194	0,169
	Pandemia	11	0,185	0,159
		12	0,175	0,161
		1	0,162	0,141
		2	0,161	0,138
		3	0,091	0,085
		4	0,140	0,119
		5	0,183	0,136
		6	0,155	0,112

Variable	Periodo	Mes	Pearson	Spearman
NSP 365 días	Control	11	0,182	0,161
		12	0,168	0,152
		1	0,202	0,166
		2	0,197	0,164
		3	0,193	0,165
		4	0,192	0,165
		5	0,215	0,186
		6	0,204	0,178
	Pandemia	11	0,185	0,159
		12	0,175	0,161
		1	0,163	0,141
		2	0,168	0,144
		3	0,119	0,100
		4	0,135	0,107
		5	0,186	0,132
		6	0,153	0,107

Variable	Periodo	Mes	Pearson	Spearman
Dif. Fecha Cita	Control	11	-0,128	-0,175
		12	-0,127	-0,172
		1	-0,118	-0,151
		2	-0,122	-0,163
		3	-0,114	-0,162
		4	-0,093	-0,134
		5	-0,102	-0,143
		6	-0,097	-0,136
	Pandemia	11	-0,151	-0,206
		12	-0,132	-0,177
		1	-0,143	-0,185
		2	-0,136	-0,183
		3	-0,102	-0,155
		4	-0,072	-0,137
		5	-0,128	-0,214
		6	-0,122	-0,187

10.4. Resultado mensual entrenamiento Regresión Logística

Modelo	Año	Mes	PT	% NSP	Positivos ($\hat{p} > 0.5$)	*F1-Score*	ROC AUC	Precisión	Sensibilidad	GM	MCC
Regresión Logística	**Total**	**-**	**-**	12%	**111.5028**	**0,26**	**0,63**	**0,17**	**0,61**	**0,57**	**0,13**
	2019	Ene	1	12%	54.954	0,27	0,65	0,17	0,62	0,60	0,14
		Feb	2	13%	36.898	0,29	0,65	0,20	0,57	0,61	0,15
		Mar	3	13%	63.641	0,27	0,63	0,17	0,68	0,60	0,13
		Abr	4	13%	50.109	0,27	0,63	0,18	0,56	0,59	0,13
		May	5	14%	67.748	0,28	0,62	0,18	0,70	0,58	0,13
		Jun	6	14%	81.274	0,26	0,61	0,16	0,79	0,53	0,10
		Jul	7	13%	54.405	0,28	0,64	0,18	0,59	0,60	0,14
		Ago	8	13%	77.748	0,26	0,62	0,16	0,74	0,57	0,12
		Sept	9	14%	60.683	0,28	0,63	0,18	0,75	0,58	0,13
		Oct	10	17%	76.801	0,33	0,65	0,22	0,80	0,57	0,16
		Nov	11	17%	85.406	0,32	0,64	0,20	0,89	0,49	0,14
		Dic	12	15%	70.115	0,30	0,65	0,19	0,82	0,56	0,15
	2020	Ene	13	13%	44.941	0,29	0,66	0,20	0,55	0,61	0,16
		Feb	14	13%	31.457	0,29	0,66	0,21	0,49	0,59	0,15
		Inicio de la Pandemia									
		Mar	15	13%	52.747	0,27	0,62	0,17	0,68	0,59	0,12
		Abr	16	6%	19.623	0,15	0,62	0,09	0,62	0,60	0,09
		May	17	9%	10.056	0,21	0,65	0,14	0,42	0,56	0,12
		Jun	18	8%	6.189	0,21	0,62	0,17	0,31	0,52	0,13
		Jul	19	8%	9.454	0,17	0,59	0,12	0,27	0,48	0,07
		Ago	20	9%	15.860	0,21	0,63	0,14	0,41	0,56	0,11
		Sept	21	10%	20.915	0,21	0,62	0,14	0,43	0,56	0,10
		Oct	22	10%	39.930	0,22	0,63	0,13	0,65	0,59	0,11
		Nov	23	11%	43.617	0,24	0,63	0,15	0,66	0,60	0,12
		Dic	24	12%	40.457	0,27	0,65	0,17	0,70	0,61	0,15

10.5. Resultado mensual entrenamiento *Random Forests*

Modelo	Año	Mes	PT	% NSP	Positivos ($\hat{p}$ > 0.5)	*F1-Score*	ROC AUC	Precisión	Sensibilidad	GM	MCC
Random Forests	**Total**	**-**	**-**	12%	**10.079**	**0.02**	**0,72**	**0,38**	**0,01**	**0,09**	**0,05**
	2019	Ene	1	12%	1.027	0,07	0,75	0,57	0,04	0,19	0,12
		Feb	2	13%	1.131	0,09	0,77	0,60	0,05	0,23	0,15
		Mar	3	13%	180	0	0,72	0,33	0,00	0,06	0,02
		Abr	4	13%	1.307	0,07	0,76	0,53	0,04	0,21	0,12
		May	5	14%	745	0,03	0,72	0,37	0,02	0,13	0,05
		Jun	6	14%	703	0,03	0,71	0,54	0,02	0,15	0,09
		Jul	7	13%	580	0,04	0,76	0,66	0,02	0,15	0,10
		Ago	8	13%	690	0,04	0,71	0,30	0,01	0,11	0,04
		Sept	9	14%	552	0,02	0,71	0,29	0,01	0,11	0,03
		Oct	10	17%	499	0,02	0,70	0,36	0,01	0,09	0,03
		Nov	11	17%	158	0,04	0,71	0,28	0,00	0,05	0,01
		Dic	12	15%	16	0,01	0,69	0,63	0,00	0,02	0,02
	2020	Ene	13	13%	598	0,03	0,75	0,57	0,02	0,14	0,09
		Feb	14	13%	882	0,06	0,74	0,50	0,03	0,18	0,10
		Inicio de la Pandemia									
		Mar	15	13%	731	0,02	0,67	0,26	0,01	0,12	0,03
		Abr	16	6%	4	0	0,73	0,00	0,00	0,00	0,00
		May	17	9%	5	0	0,77	0,80	0,00	0,03	0,03
		Jun	18	8%	0	0	0,74	0,00	0,00	0,00	0,00
		Jul	19	8%	0	0	0,75	0,00	0,00	0,00	0,00
		Ago	20	9%	7	0	0,74	0,14	0,00	0,01	0,00
		Sept	21	10%	5	0	0,71	1,00	0,00	0,03	0,03
		Oct	22	10%	259	0,02	0,72	0,38	0,01	0,11	0,05
		Nov	23	11%	0	0	0,69	0,00	0,00	0,00	0,00
		Dic	24	12%	0	0	0,68	0,00	0,00	0,00	0,00