



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

**ANÁLISIS Y PRONÓSTICO DE VENTA DE PRODUCTOS PARA LA
GESTIÓN DE INVENTARIO DE LA FARMACIA COMUNAL SAN MIGUEL**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL INDUSTRIAL

FLAVIA GIGLIOLA VARNERO RAIN

PROFESOR GUÍA:
CAROLINA SEGOVIA RIQUELME

MIEMBROS DE LA COMISIÓN:
ALEJADRA PUENTE CHANDÍA
FELIPE VILDOSO CASTILLO

Este trabajo ha sido parcialmente financiado por:
FarmaBot

SANTIAGO DE CHILE
2021

ANÁLISIS Y PRONÓSTICO DE VENTA DE PRODUCTOS PARA LA GESTIÓN DE INVENTARIO DE LA FARMACIA COMUNAL SAN MIGUEL

La presente memoria aborda el análisis y pronóstico de venta de productos de la Farmacia Comunal de San Miguel con foco en generar un aporte a la gestión de inventario de esta institución.

El trabajo se centra en la aplicación metodológica del CRISP-DM y el uso de técnicas de modelamiento variadas que aportarán a resolver la problemática de hallar el mejor método de predicción de valores de venta en periodos cercanos, desde el mes actual y hasta 5 meses hacia adelante. Se tuvo acceso directo a la base de datos transaccional de la farmacia, permitiendo contar con la información necesaria para generar las series de tiempo que permitieron indagar, analizar y corregir datos de series de tiempo.

Se modelaron más de 350 series de tiempo, elegidas por su viabilidad analítica, con foco en la aplicación y análisis de las técnicas de *medias móviles simples*, *ARIMA*, *Prophet* y su variación *Prophet con estacionalidad anual forzada*, pero también se replicó la metodología que se utiliza actualmente en la farmacia a fin de comparar las nuevas opciones con la situación actual.

Adicionalmente, se aplicó un clasificador basado en el método de aprendizaje *Random Forest* capaz de elegir, en un 58 % de los casos, la mejor técnica de predicción para cualquier serie de tiempo solamente en base a sus características y luego de ser entrenada con un set de datos. A partir del modelamiento se realizó el análisis iterativo de resultados para la generación de conclusiones que aportasen a la resolución del problema.

Las principales conclusiones del trabajo son, en primer lugar, que cada una de las series de tiempo responde de manera distinta a cada una de las series de tiempo basados en sus propias características y, por lo tanto, se obtendrán mejores desempeños al utilizar etiquetas que asignen a cada serie de tiempo la técnica que debe usarse con ella específicamente. En segundo lugar, se determinó que la utilización del clasificador no fue suficientemente eficaz en su labor respecto de la inversión de tiempo y recursos computacionales que implica.

A lo anterior se suman la entrega de recomendaciones de utilización del *output* del trabajo de memoria, así como también sugerencias y lineamientos a seguir para dar continuidad a la labor ya realizada. Por último, se destaca la recomendación de aplicación de prácticas de negocio sencillas que aporten a la mejora del sistema FarmaBot donde, sin la necesidad de grandes inversiones, se generen impactos positivos en el servicio que la farmacia presta a la comunidad.

*Dedico este trabajo a mi familia,
quienes siempre han creído en mí pese a todo,
demostrando su apoyo y afecto
durante todo el proceso y hasta el final.*

Les amo

Agradecimientos

A toda mi amada familia, quienes siempre creyeron en mi. En especial a mi madre, padre, hermanos, mi tía Ester y Carlito, por apoyarme y motivarme en todo momento, por estar siempre a mi lado y darme las palabras de apoyo que tanto necesitaba aunque no las pidiera. Otra mención especial a mi hermana, por ayudarme sin ningún reparo cada vez que se lo pedí.

A mis amigas y amigos, las y los de la universidad, del liceo y de la vida, por siempre dar palabras de aliento e inspirarme cada día con sus inacabables éxitos, por todas esas juntas preciosas y por todo su cariño. Doy especial mención a Simón Sepúlveda, quien mantuvo la calma para ayudarme cuando tuve problemas técnicos.

A mis profesores, entre ellos a Carolina Segovia, por no perder la fe en mi trabajo; y a Felipe Vildoso, quien confió en mi para desarrollar este tema de memoria.

Por último, pero no menos importante, a mis perritos Bianca, Papelucha y Kofke, por ellos lo dí todo.

¡Gracias!

Tabla de Contenido

| | |
|---|-----------|
| 1. Introducción | 1 |
| 1.1. Antecedentes de la empresa cliente | 1 |
| 1.1.1. Farmacia Comunal de San Miguel | 1 |
| 1.1.1.1. Estructura organizacional | 2 |
| 1.1.1.2. Funcionamiento de la farmacia | 2 |
| 1.1.2. Programación de pedidos | 4 |
| 1.2. Planteamiento del problema y justificación | 5 |
| 1.2.1. Identificación del problema | 5 |
| 1.2.2. Identificación de la hipótesis y posibles alternativas de resolución | 6 |
| 1.2.3. Propuesta de valor | 7 |
| 1.3. Objetivos | 8 |
| 1.3.1. Objetivo general | 8 |
| 1.3.2. Objetivos específicos | 8 |
| 1.4. Alcances | 8 |
| 1.4.1. Entregables | 8 |
| 1.5. Marco conceptual | 9 |
| 1.5.1. Técnicas de predicción | 9 |
| 1.5.1.1. Modelo de proyección actual | 9 |
| 1.5.1.2. Medias móviles simples | 10 |
| 1.5.1.3. ARIMA | 11 |
| 1.5.1.4. Prophet | 12 |
| 1.5.2. <i>Machine learning</i> : clasificación de técnicas | 13 |
| 1.5.2.1. <i>Resampling</i> en desbalance de datos | 13 |
| 1.5.3. Desempeño de técnicas y del clasificador | 14 |
| 1.5.3.1. MAE (<i>Mean Absolute Error</i>) | 14 |
| 1.5.3.2. MAPE (<i>Mean Absolute Percentage Error</i>) | 14 |
| 1.5.3.3. Métricas para el Clasificador | 15 |
| 1.5.3.3.1. <i>Accuracy</i> | 15 |
| 1.5.3.3.2. <i>Precision</i> | 15 |
| 1.5.3.3.3. <i>Recall</i> | 15 |
| 1.5.3.3.4. <i>F1</i> | 15 |
| 2. Metodología y desarrollo metodológico | 16 |
| 2.1. Entendimiento del negocio | 17 |
| 2.2. Entendimiento y preparación de los datos | 17 |
| 2.2.1. Estructura de los datos | 17 |
| 2.2.2. Hallazgos y modificación de datos | 19 |

| | | |
|------------|---|-----------|
| 2.2.2.1. | Identificación de <i>outlier</i> | 21 |
| 2.2.3. | Visualización de series | 21 |
| 2.2.4. | Preparación de los datos | 24 |
| 2.3. | Modelamiento de técnicas y evaluación | 25 |
| 2.3.1. | Modelamiento de técnicas | 26 |
| 2.3.1.1. | Resultados de cada técnica | 26 |
| 2.3.1.2. | Resultados utilizando mejores técnicas | 29 |
| 2.3.1.3. | Resultados de medicamentos relevantes | 30 |
| 2.3.1.3.1. | Análisis <i>LOSARTAN - 50 MG - 30 - COMPRIMIDO</i> | 31 |
| 2.3.1.3.2. | Análisis <i>METFORMINA - 850 MG - 60 - COMPRIMIDO</i> | 32 |
| 2.3.1.3.3. | Análisis <i>QUETIAPINA - 50 MG - 30 - COMPRIMIDO</i> | 33 |
| 2.3.2. | Modelamiento de clasificador | 34 |
| 2.3.2.1. | <i>Resampling</i> : sobremuestreo | 36 |
| 2.3.2.2. | Resultados clasificador de modelos | 38 |
| 3. | Conclusiones | 43 |
| 3.1. | Recomendaciones de implementación | 44 |
| 3.2. | Recomendaciones de trabajo futuros | 45 |
| | Bibliografía | 47 |
| | Anexo A. Resultados de la implementación de técnicas | 48 |
| | Anexo B. Rendimientos de la implementación de técnicas | 50 |

Índice de Tablas

| | | |
|-------|--|----|
| 2.1. | Resumen de los datos. | 19 |
| 2.2. | Principales modificaciones a la base de datos. | 20 |
| 2.3. | Resumen de los datos después de las modificaciones. | 20 |
| 2.4. | Resultados generales. | 27 |
| 2.5. | Ranking resultados por grupos en base a MAPE. | 28 |
| 2.6. | Comparación resultados <i>SMA5</i> versus selección de mejores técnicas. | 30 |
| 2.7. | Resumen métricas de resultados para <i>LOSARTAN - 50 MG - 30 - COMPRIMIDO</i> | 31 |
| 2.8. | Resumen métricas de resultados para <i>METFORMINA - 850 MG - 60 - COMPRIMIDO</i> | 33 |
| 2.9. | Resumen métricas de resultados para <i>QUETIAPINA - 50 MG - 30 - COMPRIMIDO</i> | 34 |
| 2.10. | Resumen métricas de rendimiento del clasificador. | 37 |
| 2.11. | Resumen distribución de clasificaciones originales. | 38 |
| 2.12. | Resumen métricas de rendimiento del clasificador post entrenamiento con <i>oversampling</i> | 39 |
| 2.13. | Resumen distribución de clasificaciones post entrenamiento con <i>oversampling</i> , solo datos reales. | 40 |
| 2.14. | Resumen de variación de rendimientos MAPE en base al uso del clasificador. | 41 |
| 2.15. | Resumen de variación de rendimientos MAE en base al uso del clasificador. | 42 |
| A.1. | Ejemplos de resultados de la implementación de las técnicas de predicción. T1: Técnica Actual, T2: SMA5, T3: Arima, T4: Prophet, T5: Prophet_year. | 49 |
| B.1. | Ejemplos de métricas de rendimiento MAE y MAPE a partir de la implementación de las técnicas de predicción. | 50 |

Índice de Ilustraciones

| | | |
|-------|---|----|
| 1.1. | Organigrama de la Corporación Municipal | 2 |
| 1.2. | Unidades de medicamentos adquiridas por proveedores entre años 2019 y 2021 | 3 |
| 1.3. | Inscripción de pacientes entre años 2019 y 2021 | 3 |
| 1.4. | Mecánica de abastecimiento a través de CENABAST. | 4 |
| 1.5. | Evolución de egresos monetarios por concepto de merma. | 7 |
| 1.6. | Ejemplo de serie ajustada por actual metodología. | 10 |
| 1.7. | Ejemplo de serie ajustada por <i>Simple Moving Average</i> de 5 periodos móviles. | 11 |
| 1.8. | Ejemplo de serie ajustada por <i>ARIMA(2,1,0)</i> | 12 |
| 1.9. | Ejemplo de serie ajustada por <i>Prophet</i> | 12 |
| 1.10. | Ejemplo de serie ajustada por <i>Prophet</i> con estacionalidad anual forzada. | 13 |
| 2.1. | Ciclo de vida de minería de datos | 16 |
| 2.2. | Modelo relacional del sistema transaccional de Farmacia de San Miguel | 18 |
| 2.3. | Distribución de promedios de venta para cada medicamento (serie de tiempo). | 22 |
| 2.4. | Venta de mejores productos agregados por principio activo. | 23 |
| 2.5. | Venta de medicamentos de interés analítico por grupo asignado. | 23 |
| 2.6. | Estructura de la tabla resumen. | 24 |
| 2.7. | Proporción de datos de testeo. | 25 |
| 2.8. | Resultados generales por grupo. | 27 |
| 2.9. | Resultados generales por grupo sin <i>Prophet con estacionalidad anual forzada</i> | 28 |
| 2.10. | Detalle resultados generales por grupo sin <i>Prophet con estacionalidad anual forzada</i> | 28 |
| 2.11. | Resultados de mejores técnicas. | 29 |
| 2.12. | Resultados de mejores técnicas por grupo. | 30 |
| 2.13. | Ajustes de distintas técnicas a la serie de tiempo asociada a ventas de <i>LOSARTAN - 50 MG - 30 - COMPRIMIDO</i> | 31 |
| 2.14. | Ajuste de técnica <i>SMA5</i> para la serie de tiempo de <i>LOSARTAN - 50 MG - 30 - COMPRIMIDO</i> | 32 |
| 2.15. | Ajustes de distintas técnicas a la serie de tiempo <i>METFORMINA - 850 MG - 60 - COMPRIMIDO</i> | 32 |
| 2.16. | Ajuste de técnica <i>ARIMA(0,0,0)</i> para la serie de tiempo de <i>METFORMINA - 850 MG - 60 - COMPRIMIDO</i> | 33 |
| 2.17. | Ajuste de técnica <i>Modelo Actual</i> para la serie de tiempo de <i>QUETIAPINA - 50 MG - 30 - COMPRIMIDO</i> | 34 |
| 2.18. | Distribución de valores de variables <i>X_ACF1</i> y <i>entropía</i> entre las distintas técnicas asignadas. | 35 |
| 2.19. | Correlograma detallado de variables con mayores coeficientes de correlación | 36 |
| 2.20. | Matriz de confusión | 37 |
| 2.21. | Curvas de validación de parámetros para clasificador <i>Random Forest</i> | 38 |

| | | |
|-------|--|----|
| 2.22. | Matriz de confusión post entrenamiento con <i>oversampling</i> | 39 |
| 2.23. | Matriz de confusión post entrenamiento con <i>oversampling</i> , solo datos reales. . | 40 |
| 2.24. | Ejemplo de medicamento asignado a <i>Prophet</i> con <i>estacionalidad anual forzada</i> . . | 41 |

Capítulo 1

Introducción

Es de sentido común que toda entidad que brinde un servicio deba ofrecer certezas a sus usuarios o clientes. No solo por una lógica de responsabilidad, sino que también por un sentido práctico que permita la permanencia de la entidad en el tiempo.

Esta premisa adquiere un tenor mucho más profundo cuando se trata de iniciativas sociales, donde el compromiso y continuidad son una piedra angular, en tanto suelen constituirse como la única fuente que tienen muchos de sus beneficiados de acceder al servicio en cuestión. Y qué decir de aquellos espacios vinculados a la salud, como las farmacias municipales o populares, donde los números, así como las decisiones que se toman a partir de estos, pueden ser cruciales para la vida de los pacientes.

Su motivación social implica necesariamente responder, tanto a nivel logístico como a nivel humano, a un público que cuenta con comprar en ellas medicamentos que, ya sea por su alto costo y/o por la frecuencia de uso, pueden llegar a ser casi imposibles de adquirir por otras vías. Si eso no ocurre, no se arriesga la comodidad de simples clientes: se pone en jaque la salud de las personas.

Teniendo esa inquietud como móvil, se escogió como caso de análisis el funcionamiento de la Farmacia Comunal de San Miguel donde para acceder a los datos necesarios se contó con la colaboración de la empresa Paga Rápido S.A. Dicha entidad provee a la mencionada farmacia el servicio FarmaBot (ChabotChile, 2020), aplicación web de gestión de inventario y chatbot especializado en farmacias, que a través de diversas herramientas facilita la administración de las mismas, siendo su principal propósito brindar un servicio integral a sus clientes de manera flexible e innovadora.

1.1. Antecedentes de la empresa cliente

1.1.1. Farmacia Comunal de San Miguel

La Farmacia Comunal de San Miguel, también llamada Farmacia “Popular” de San Miguel, es una iniciativa de la alcaldía y del Concejo Municipal, dependiente directamente de la droguería de San Miguel, cuyo objeto es facilitar a todos los vecinos de la comuna el acceso al tratamiento farmacológico para patologías crónicas a un costo menor en comparación con el mercado (SanMiguel, 2019).

1.1.1.1. Estructura organizacional

En cuanto a la estructura organizacional (figura 1.1), se pueden establecer 4 niveles de jerarquía descritos a continuación.

En el primer nivel, se encuentra el **director de la corporación comunal**, quien está a cargo tanto del proyecto de la droguería como el de la farmacia comunal. Dentro de sus funciones está el tomar decisiones respecto a la administración de dichas iniciativas, principalmente en cuanto a presupuesto y recursos adicionales asignados a estas.

En segundo nivel, está el **director técnico de la droguería** cuya principal tarea es administrar dicha entidad procurando su correcto funcionamiento; esto implica, entre otras cosas, el aseguramiento del stock suficiente para satisfacer las necesidades de los centros dependientes de la droguería y la supervisión de los procesos de recepción y entrega de productos, garantizando el uso óptimo de los recursos. Cerciorarse del adecuado funcionamiento de la farmacia comunal también es parte de sus competencias.

En un tercer nivel se encuentra el **director técnico y administrador de la farmacia comunal**. Este cumple labores similares al director de la droguería, pero aterrizadas al funcionamiento particular de la farmacia y sus operaciones internas. Es con esta entidad con quien se trabajará en cuanto a la entrega de información y requerimientos específicos que tiene el trabajo de memoria.

En cuarto nivel, se encuentran el **personal de atención al público y el personal de aseo**, encargados de mantener en óptimas condiciones el establecimiento.



Figura 1.1: Organigrama de la Corporación Municipal

1.1.1.2. Funcionamiento de la farmacia

Actualmente la farmacia comunal de San Miguel se dedica a vender principalmente medicamentos, y en una menor proporción insumos médicos. Estos productos son adquiridos a través de distintos mecanismos, pudiendo proveerlos la droguería de la comuna, los laboratorios, o también intermediarios. Entre estos últimos se encuentra CENABAST¹, uno de los

¹ Central de Abastecimiento del Sistema Nacional de Servicios de Salud, institución pública descentralizada dependiente del Ministerio de Salud, que tiene como objetivo la disponibilización de medicamentos, ali-

principales medios de abastecimiento con más del 23% del suministro total de medicamentos adquiridos por la farmacia, y con particular estabilidad en el tiempo en comparación a otros proveedores.

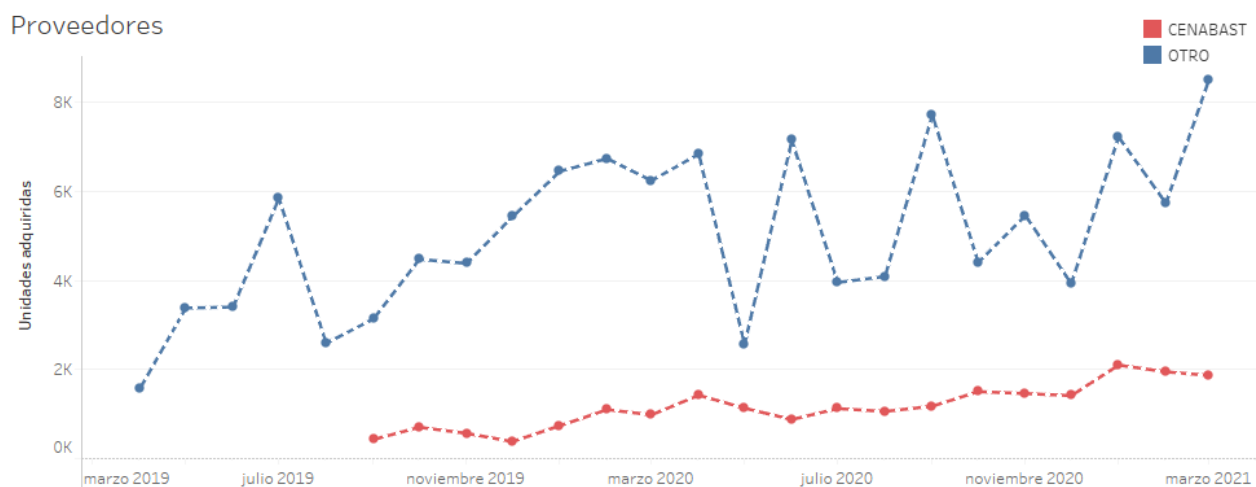


Figura 1.2: Unidades de medicamentos adquiridas por proveedores entre años 2019 y 2021

Por otro lado, la organización se limita a vender sus productos sólo a vecinos de la comuna, teniendo un alcance aproximado del 10,48% del total de la población comunal actual – 10.315 pacientes de un total de 107.954 habitantes aproximadamente – quienes para acceder al servicio deben rellenar un formulario y presentar documentos que acrediten su residencia en la comuna, junto con receta médica que certifique la necesidad de medicamentos.

La farmacia se encuentra actualmente en un proceso de crecimiento de demanda de fármacos generada por una gran cantidad de inscripciones en los años recientes, con un promedio de 200 pacientes nuevos inscritos mensualmente durante el último año.

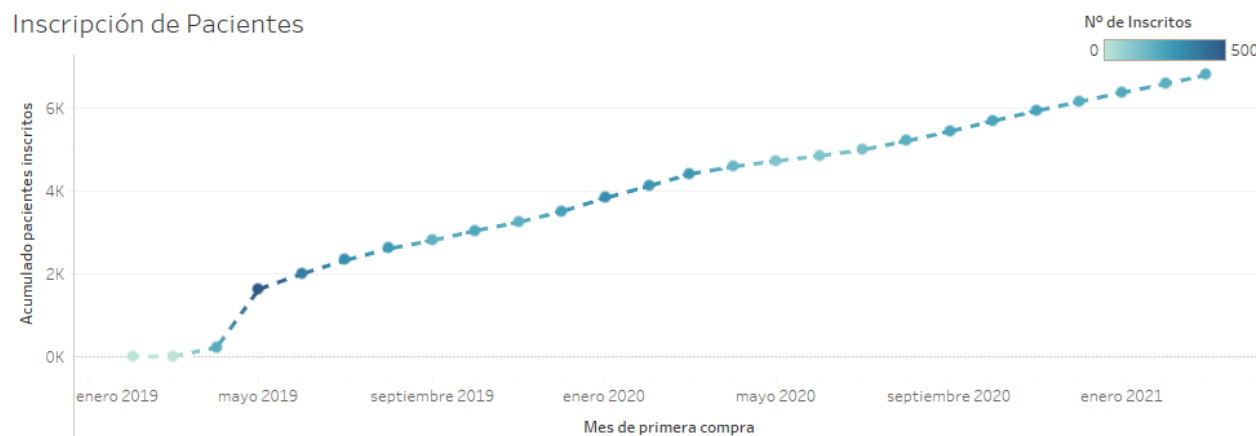


Figura 1.3: Inscripción de pacientes entre años 2019 y 2021

En particular, los medicamentos comercializados se ofrecen a los pacientes de la comuna al mismo precio de adquisición por parte de la farmacia, por lo cual no se generan márgenes

mentos, insumos y equipamiento a la Red de Salud, mediante la consolidación de demanda y la gestión de un servicio de abastecimiento. (CENABAST, 2017)

de ganancia por el ejercicio de venta. Los gastos operacionales son asumidos como inversión por parte de la corporación municipal y por tanto no se transfieren a los pacientes.

1.1.2. Programación de pedidos

Como se menciona anteriormente, gran parte de las adquisiciones se realizan mediante CENABAST respondiendo, por lo tanto, a los plazos asignados por este proveedor para la planificación de sus operaciones de abastecimiento.

Además, la farmacia se abastece a través de otros intermediarios y mediante compra directa a laboratorios. Estos mecanismos resultan más inmediatos y son utilizados para la compra de medicamentos específicos y para compensar la demanda no prevista ni considerada al momento de programar los pedidos vía CENABAST. Sin embargo, los valores de los medicamentos mediante estas alternativas suelen ser más elevados, transfiriendo así este valor al consumidor final en su totalidad.

Actualmente la farmacia comercializa 876 distintos medicamentos, los cuales se resumen en 523 distintos Principios Activos adquiridos en distintas dosificaciones y formatos². El pedido de gran parte de dichos productos es realizado a la CENABAST a través de una proyección de demanda anual de los medicamentos. El cálculo se realiza el año anterior al periodo a cubrir, disponibilizando adicionalmente “ventanas de reprogramación” de pedidos 3 meses antes de cada mes a modificar. Dichas ventanas permiten la alteración de hasta un 30 %, al alza o a la baja, de la programación inicial³. La programación se resume en la figura 1.4.

En particular, en esta memoria se trabajó con foco específico en las ventanas de reprogramación debido a la importancia que tiene este hito ante la fluctuación de demanda y stock en la farmacia.

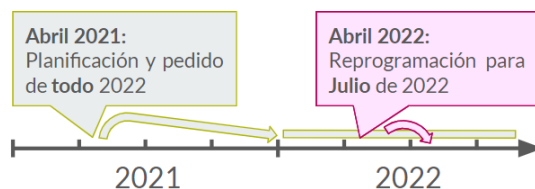


Figura 1.4: Mecánica de abastecimiento a través de CENABAST.

Pese a que no todos los productos se adquieren necesariamente con este régimen, la farmacia se ha acomodado a este sistema de análisis de ventas y programación de pedidos y, por ende, también la administración de inventario asociada al abastecimiento directo desde los laboratorios.

² Un principio activo puede tener distintos formatos, los cuales se contabilizan como medicamentos distintos. Por ejemplo, Paracetamol 500 miligramos y Paracetamol 1 gramo son 2 medicamentos para un mismo principio activo.

³ Por ejemplo, para el abastecimiento del año 2022, en abril de 2021 se hace la proyección y pedido mensual para todo ese año (2022). Así mismo, el mes de febrero (mes X) de 2022 se disponibiliza una ventana de modificación para el pedido del mes de mayo (mes X+3) del 2022 en máximo 30 % al alza o a la baja del pedido realizado originalmente en la proyección inicial.

1.2. Planteamiento del problema y justificación

Actualmente la farmacia maneja un gran stock de productos farmacéuticos para su funcionamiento anual. Si bien en este momento se cuenta con un sistema que permite visualizar y administrar en tiempo real las existencias del inventario, esto no ha sido suficiente para una correcta prestación del servicio a la comunidad.

1.2.1. Identificación del problema

Debido al régimen de pedidos expuesto en la sección 1.1.2, se generan de manera habitual dificultades asociadas a la gestión y almacenamiento de stock. En tal contexto el principal problema está referido a que los pronósticos de demanda de los fármacos actualmente no responden a la real demanda de estos por parte de la población.

En particular, el problema detectado consiste en que **el pronóstico de demanda actualmente utilizado no responde a la real demanda de productos**. Esto genera consecuencias tales como:

- **Quiebres de stock** que llevan a la farmacia a no prestar un servicio adecuado para el que, por cierto, fue creada específicamente, generando externalidades negativas en la comunidad que debe atender. Los quiebres también generan la adquisición de medicamentos a mayor precio a distribuidores y laboratorios con un servicio más inmediato, sin embargo esto eleva significativamente y directamente el valor de los productos ante la población.
- **Sobre stock** de medicamentos, lo que desemboca en la ocupación de espacio físico en bodegas de espacio limitado. Adicionalmente, se generan mermas las cuales representan pérdidas directas para la administración de la farmacia.

Vale subrayar que, aun si la generación de margen monetario no es el foco de las operaciones de la farmacia, las pérdidas de dinero invertido a las que se ve enfrentada la entidad constituyen un problema relevante, siendo dichas pérdidas provocadas principalmente por la generación de mermas asociadas al sobre stock.

Otro aspecto a mencionar es la restricción de espacio para almacenamiento, limitación que se ve acentuada por el sobre stock señalado, ya que complica la adquisición de otros fármacos en pedidos posteriores y acrecienta la gravedad de la situación de los medicamentos con quiebre de stock. Finalmente, se destaca que las consecuencias del problema planteado dificultan el logro de los objetivos con los que fue concebida la farmacia.

Como algunas de las raíces del proceso se han identificado:

- **Ausencia de respaldo analítico** para la toma de decisiones de abastecimiento, principalmente debido a que no se cuenta con herramientas que faciliten estas actividades específicamente, por lo tanto los cálculos y “proyecciones” son realizados de manera manual.

- **Estimación arbitraria de demanda proyectada.** La estimación de demanda se hace a través del uso del promedio de meses anteriores y, en ocasiones, aumentando un porcentaje a este dependiendo de si la encargada “cree” que aumentará la demanda del producto. Si bien este método parece funcionar para algunos de los productos, la decisión sobre si se aumenta dicho porcentaje, o en cuanto se hace, se realiza arbitrariamente, y es una práctica arraigada en dicha farmacia debido a la ausencia de cultura analítica.

1.2.2. Identificación de la hipótesis y posibles alternativas de resolución

A partir de la problemática expuesta en la sección 1.2, se plantea como hipótesis que la falta de procesos analíticos para el pronóstico de demanda y consecuente toma de decisiones en cuanto a los pedidos de abastecimiento, ocasionan desequilibrios recurrentes en el inventario, expresándose en sobre stock y quiebres de stock para distintos productos.

Actualmente el “método” utilizado para las proyecciones de demanda es aumentar entre 5 % a 30 % la demanda del año anterior para cada producto según criterio del encargado del pedido (Cathalifaud, 2020). Se declara que en los últimos meses se ha utilizado la técnica de promedio simple de los últimos 5 meses para repetir estos en los periodos que se desean “proyectar”. A pesar de los esfuerzos de la encargada de realizar los pedidos en la farmacia, los quiebres y sobre stock recurrentes son prueba de que el proceso actual no es óptimo.

Existen variados factores que afectan, o tienen el potencial de afectar la demanda de productos en una farmacia. Entre ellos se pueden considerar el (de)crecimiento demográfico de la comuna, la visibilidad del servicio, el acceso físico al local, la incidencia de enfermedades en la población, la existencia de otros servicios similares en las cercanías, la disponibilidad del fármaco en consultorios, etc. Estos elementos a su vez varían en el tiempo, sumado a una presumible componente de aleatoriedad o impredecibilidad en el comportamiento⁴. Todo lo anterior se encuentra actualmente remitido a la variación de un porcentaje arbitrario en el aumento del pedido que actualmente se realiza en la farmacia periodo tras periodo.

A partir de los datos de la farmacia, se evidencia un gran crecimiento en la inscripción de pacientes con recetas (asociados a uno o más fármacos), con casi 115 % más de personas usando los beneficios de la institución en poco más de un año⁵.

Atendiendo a la estocasticidad de las variables y a la imposibilidad de considerar todas estas en el actual proceso de proyección de demanda, se propone como solución la utilización de modelamiento de demanda a través de un proceso que proporcione sustento analítico y que permita disminuir la brecha entre lo proyectado y lo demandado, permitiendo mejoras en la gestión de inventario para brindar un mejor servicio a la comunidad y el cumplimiento de los objetivos de la farmacia.

⁴ Por ejemplo, en el año 2002, el Instituto Nacional de Estadísticas (INE), proyectó que el decrecimiento poblacional se acentuaría para la comuna de San Miguel, llegando a tasas de crecimiento anual de -1,3 desde el 2002 al año 2020. Sin embargo, se ha evidenciado un explosivo crecimiento poblacional de la comuna entre los años 2002 y 2017 de 36,87%, traducido a una tasa anual de más de 2,1 % anual, contrario a las iniciales proyecciones. (SanMiguel, 2018)

⁵ En septiembre de 2019 se encontraban inscritos 4.932 pacientes. En diciembre de 2020 la cifra asciende a casi 10.600 y a marzo de 2021 más de 11.300 pacientes.

1.2.3. Propuesta de valor

El análisis preliminar sobre los medicamentos permite evidenciar que en general casi un 20 % de los principios activos que maneja la farmacia suelen sufrir quiebres de stock, **los que ante una mejora en el sistema de predicción se podrían evitar en una gran proporción**, con los consiguientes beneficios que ello conllevaría para la población que demanda los medicamentos involucrados.

Adicionalmente, existen más de 70 principios activos que cuentan con un gran stock, el que no se justifica si se consideran sus niveles de demanda. Una predicción de inventario más precisa permitiría a la farmacia contar con menor nivel de stock, acorde a las necesidades reales del producto, **reduciendo así situaciones como la merma por vencimiento y la sobreutilización de espacio de bodega que impide pedidos más grandes de los fármacos de mayor demanda.**

Cabe destacar que el valor de las mermas generadas en el último año asciende a más de 11 millones de pesos⁶, con un promedio de 1,8 % del valor de la venta total⁷.

Merma en el Tiempo

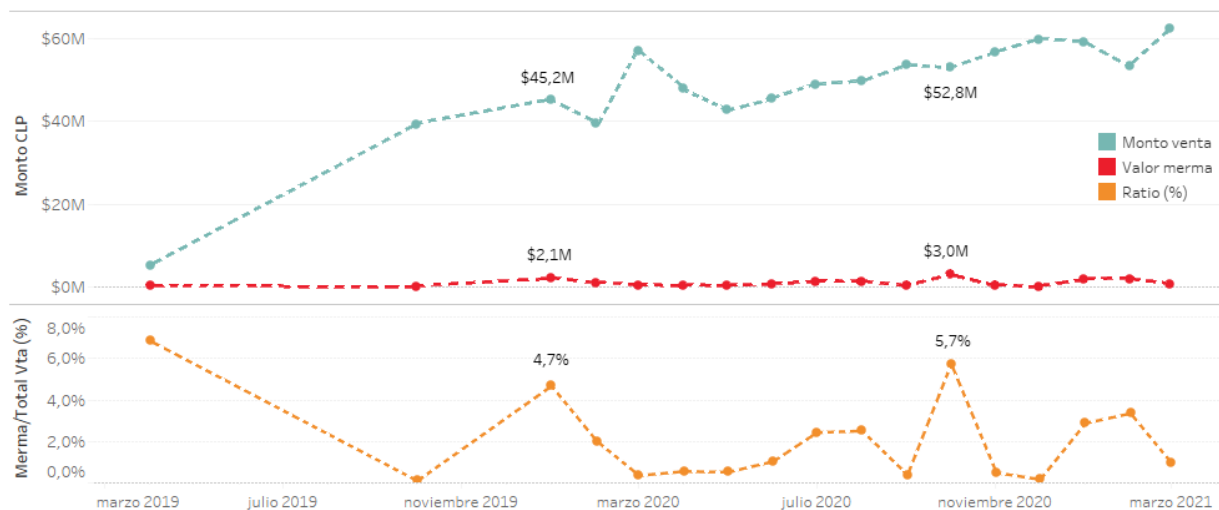


Figura 1.5: Evolución de egresos monetarios por concepto de merma.

En la figura 1.5, se muestra en rojo el egreso en *Millones de Pesos* de la merma entre años 2019 y 2021, contando con dos *peaks* destacables, primero en Enero de 2019, llegando a un monto de \$2,1 millones (4,7 % de la venta del aquel mes), y en octubre de 2020, con un monto ascendente a los \$3 millones (5,7 % del total de la venta del mes).

Con el resultado del trabajo realizado, se busca **disminuir las brechas entre las proyecciones que desde la farmacia se hacen y la demanda de los pacientes**, ya sea a través de la disminución de quiebres de stock, así como de la disminución de mermas en el tiempo.

⁶ Entre abril de 2020 y marzo de 2021.

⁷ Es importante indicar que las mermas promedios evaluadas en Chile en 2014 para el sector de farmacias son de 1,22 % sobre las ventas ((Logitec, 2014), por lo cual la Farmacia de San Miguel se encuentra muy por sobre dicho promedio.

1.3. Objetivos

Frente al problema presentado, se plantean los siguientes objetivos a alcanzar con la realización del trabajo de memoria.

1.3.1. Objetivo general

Generar alternativas predictivas de la demanda de medicamentos para su uso en la planificación del stock en la Farmacia Comunal de San Miguel.

1.3.2. Objetivos específicos

Para la concreción del punto anterior, se llevaron a cabo los siguientes objetivos particulares:

- Generar series de tiempo ordenadas y distinguibles entre si en base a la data original.
- Aplicar variadas técnicas de pronóstico de demanda a los productos para distintos horizontes de tiempo.
- Generar métricas de rendimiento para cada técnica aplicada y seleccionar la mejor para cada serie de tiempo.
- Clasificar series de tiempo mediante el método de clasificación *machine learning*, asignando la mejor técnica a cada una de ellas en base a sus características.
- Generar recomendaciones de negocio en base a resultados obtenidos.

1.4. Alcances

Considerando que hasta ahora la farmacia no cuenta con sistemas elaborados de predicción, se plantea buscar la solución que mejor responda a las simulaciones de demanda para los distintos horizontes de tiempo basados en la actual programación de pedidos, incluso considerando el actual método. Se utilizan distintos técnicas matemáticas de predicción para contar con alternativas que permitan **evaluar tanto efectividad como eficiencia en pos de una solución óptima.**

Los datos utilizados para el modelamiento de técnicas comprenden los años 2019 y 2020, escogiéndose como periodos de entrenamiento los meses entre marzo de 2019 hasta septiembre de 2020. En tanto, desde noviembre de 2020 y los primeros tres meses de 2021 se configuran como datos de prueba.

Se seleccionan estratégicamente productos representativos para efectos de muestra de análisis y resultados del proceso de modelamiento.

1.4.1. Entregables

Se comprometió como entregable un set de técnicas aplicables a los productos actualmente comercializados en la Farmacia Comunal de San Miguel, esto a través de un archivo de código

ordenado en Python, así como los códigos asociados a la evaluación de resultados. Las técnicas ejecutables en dichos códigos serán capaces de generar datos predictivos del comportamiento de demanda de los productos en la farmacia a los horizontes de tiempo que la organización maneja, específicamente:

- Mensual con 3, 4 y 5 meses de anticipación para ejecución mensual⁸.

Estos horizontes de tiempo se definen como entregables en base a la cantidad de historia disponible, por parte del sistema transaccional de la farmacia y como parte de las necesidades de la administración de esta para cubrir con mayor precisión la demanda de productos que se adquieren de forma más inmediata a través de proveedores alternativos a la CENABAST (sobre del 70 % de los productos).

El lenguaje Python fue seleccionado para ser utilizado debido a su capacidad de interacción con el actual software utilizado por FarmaBot. Esto facilitará la implementación de las técnicas en las operaciones de la Farmacia.

1.5. Marco conceptual

Se propuso y desarrolló la indagación, prueba y selección entre distintas técnicas de modelamiento de demanda, con foco en considerar distintos niveles de complejidad para su comparación en términos tanto de efectividad como de eficiencia y finalmente recomendar la integración de los resultados con la plataforma actual utilizada en la gestión del negocio, FarmaBot.

1.5.1. Técnicas de predicción

Entendiendo que todos los medicamentos pueden tener comportamientos distintos entre sí, se consideraron diversas técnicas apuntando a cubrir los múltiples escenarios.

Las técnicas descritas a continuación se probaron con la intención de abarcar una variedad de comportamientos asociados a los productos a analizar. Los comportamientos y peculiaridades son, por ejemplo, características auto regresivas, tendencias, estacionalidades, componentes aleatorios, suavizamientos, etc.

1.5.1.1. Modelo de proyección actual

En los últimos meses, en la farmacia se ha implementado la proyección de demanda en base al promedio de los últimos cinco meses, y repitiendo este valor para todos los periodos que se deseen hacia adelante de la siguiente manera:

$$f = \frac{1}{5} \sum_t y_t \quad (1.1)$$

Donde t corresponde a los meses anteriores al mes actual (cuando se hace la estimación) y con valores entre $-5, -4, -3, -2, -1$; y donde y_t corresponde a los valores de venta reales de

⁸ Incluyendo también las proyecciones de los meses 1, 2 y del mes en curso, llamado mes 0.

los meses correspondientes a t . Por ejemplo, y_{-1} corresponde al valor de venta real del mes pasado.

Se emplea esta técnica de predicción a fin de obtener los parámetros de rendimiento que permitan compararlo con el resto de las técnicas implementadas.

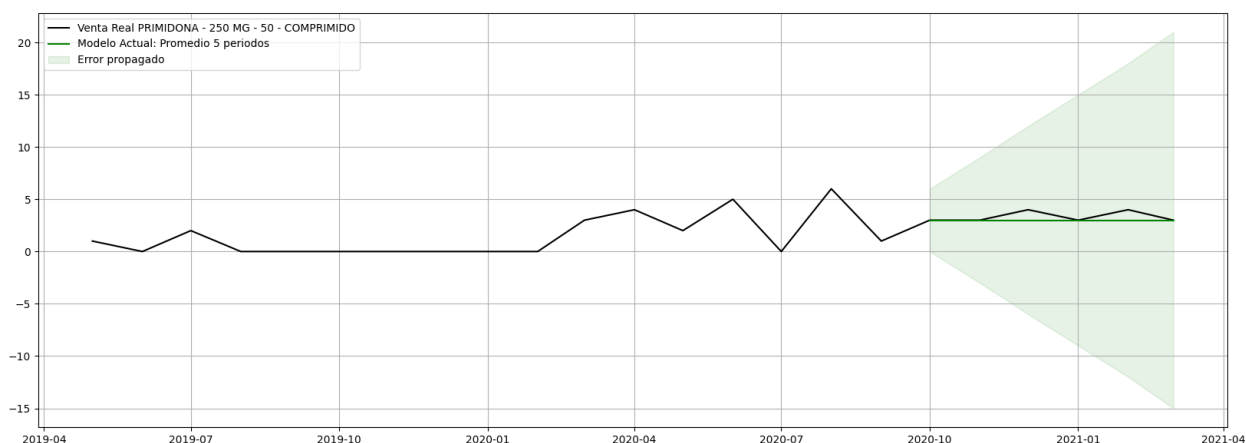


Figura 1.6: Ejemplo de serie ajustada por actual metodología.

1.5.1.2. Medias móviles simples

Las medias móviles simples o *SMA* (por sus siglas en inglés), son secuencias provenientes de la media aritmética de k elementos anteriores, en este caso, se aplicó el promedio simple de cinco periodos de las ventas anteriores por ser el de mejor rendimiento entre los casos $k = 3, 4, 5, 6$ probados. Esta técnica funciona bien en casos con baja variabilidad de ventas, sin embargo, falla en productos con mayor variación, reflejando de manera tardía los choques de demanda o las bajas abruptas de esta.

La técnica se construyó de la siguiente manera, tomando en consideración de que para proyecciones de más de un periodo hacia adelante, se toman como referencia los valores de periodos más cercanos ya pronosticados.

$$f_i = \frac{1}{5} \sum_{t=i-5}^{i-1} y_t \quad (1.2)$$

Donde i corresponde a meses con valores entre 0, 1, 2, 3, 4, 5; y donde f_i corresponde a los valores a predecir para cada mes, por ejemplo, f_0 es la estimación del mes actual y f_1 es la estimación del próximo mes.

Además, t corresponde a los meses anteriores al mes a calcular, donde y_t corresponde a los valores de venta reales de los meses correspondientes a t , por ejemplo, y_{-1} corresponde al valor real del mes pasado. Para el caso en que se busque predecir el valor, por ejemplo, f_1 , el valor de y_0 será igual al valor calculado correspondiente a f_0 .

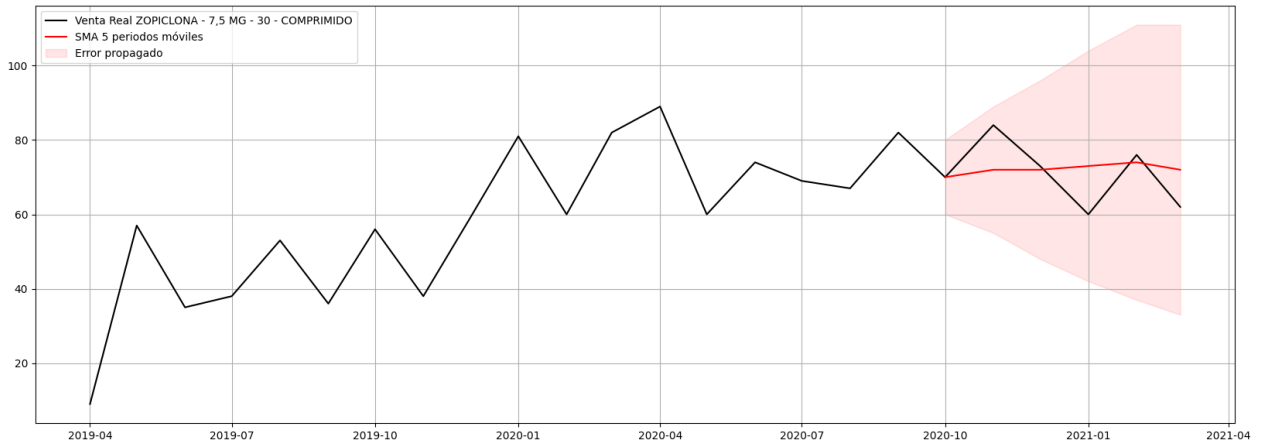


Figura 1.7: Ejemplo de serie ajustada por *Simple Moving Average* de 5 periodos móviles.

1.5.1.3. ARIMA

Técnica autorregresiva que estudia predicciones de series de tiempo, donde estas series pueden considerarse como la realización de procesos estocásticos que se observan secuencialmente en el tiempo (Chávez Quisbert, 1997).

Particularmente se implementó la función *auto_arima* que busca, a través de un proceso de optimización la búsqueda de los parámetro p , q y d óptimos que permitan el mejor ajuste a cada una de las series de tiempo.

Los parámetros mencionados se definen como:

- p es el número de términos autorregresivos, es decir, la cantidad de periodos inmediatamente anteriores que afectan el resultado un periodo determinado.
- d es el número de diferencias que se aplican a la serie de tiempo para que sea estacionaria.
- q es el número de periodos a considerar en el cálculo de medias móviles que realiza el proceso.

Se determinó que la función *auto_arima* considerase en cada iteración la búsqueda de términos que respondieran a la potencial estacionalidad de las series de tiempo; sin embargo, no se detectó para ninguna de las series de tiempo analizadas.

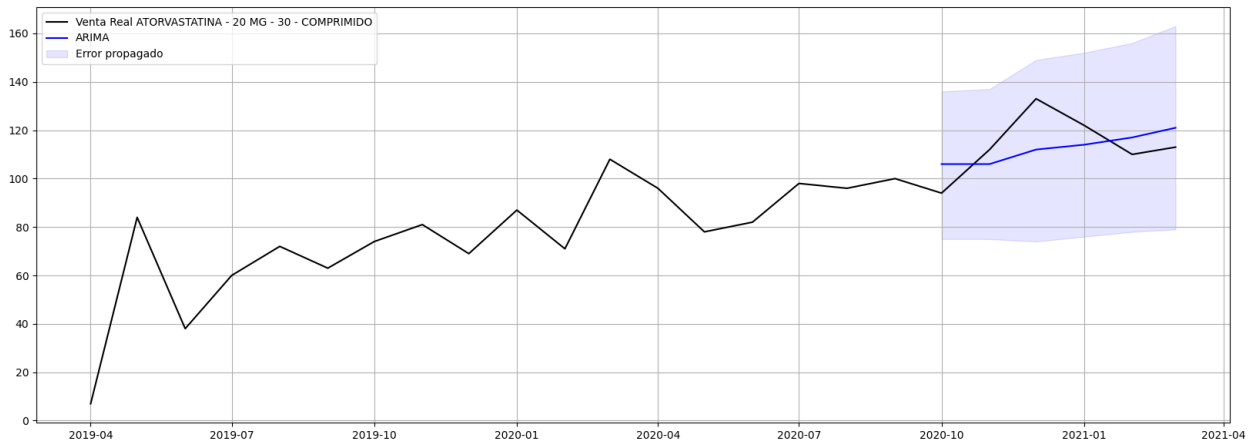


Figura 1.8: Ejemplo de serie ajustada por ARIMA(2,1,0).

1.5.1.4. Prophet

Prophet es un procedimiento para pronosticar datos de series de tiempo basado en una técnica aditiva en el que las tendencias no lineales se ajustan a la estacionalidad anual, semanal y diaria, además de los efectos de las vacaciones (Taylor y Letham, 2017). Esta técnica funciona mejor con series de tiempo que tienen fuertes efectos estacionales, por lo cual fue seleccionada para cubrir la brecha que las técnicas anteriores no lograron cubrir.

A continuación se presenta una serie de tiempo ajustada por *Prophet* para los meses entre octubre de 2019 y Marzo de 2020.

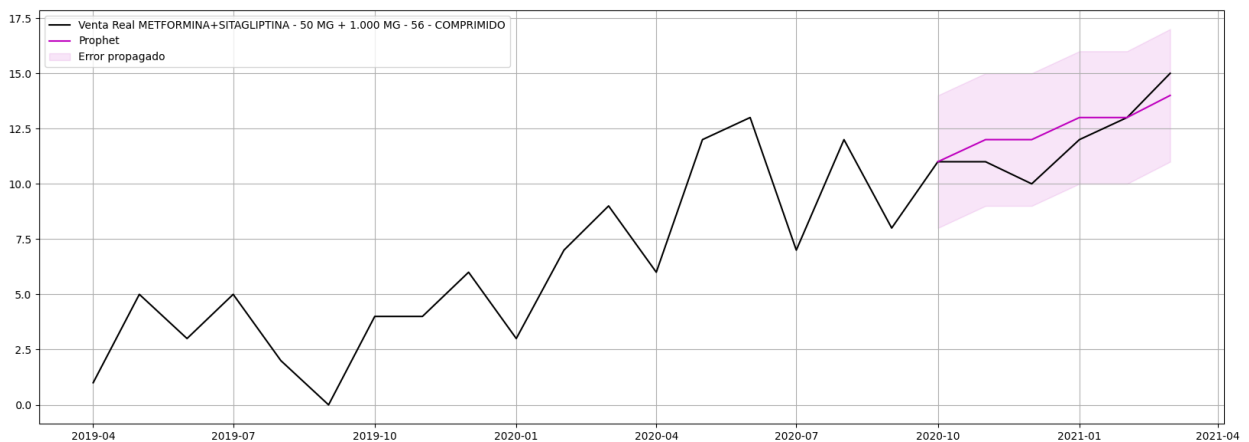


Figura 1.9: Ejemplo de serie ajustada por *Prophet*.

Adicionalmente a la técnica simple de Prophet, se forzó la consideración de estacionalidad anual sobre la predicción de valores, la cual ajustó mejor que las técnicas anteriores solo en algunas series de tiempo.

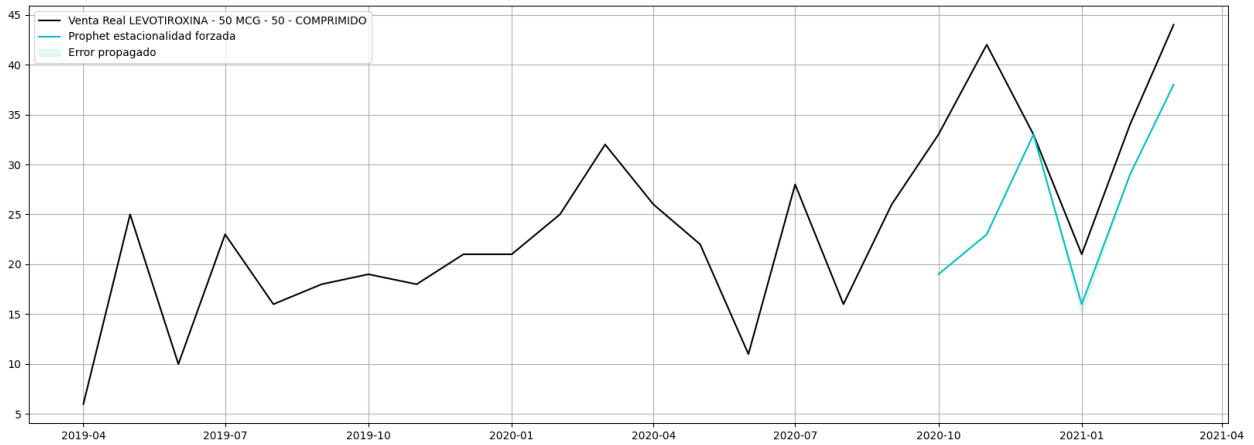


Figura 1.10: Ejemplo de serie ajustada por Prophet con estacionalidad anual forzada.

1.5.2. *Machine learning*: clasificación de técnicas

Considerando que se cuenta con los datos de ajuste de las técnicas anteriores, queda preguntarse qué hacer si se cuenta con una nueva serie de tiempo asociada a un nuevo medicamento. Evidentemente es posible aplicar las técnicas disponibles y seleccionar la que genere mejores resultados; no obstante, también es posible caracterizar esta serie de tiempo e identificar cuál sería la mejor técnica a aplicar en base a dichas características.

Ante esto, se propuso la utilización de *machine learning* a través del método de clasificación *random forest*, algoritmo de aprendizaje supervisado que debido a su capacidad de considerar múltiples árboles de decisión genera resultados robustos en base a las técnicas disponibles y las características de las series de tiempo (Breiman, 2001). Dichas características son determinadas a través del paquete `tsfeatures`⁹.

1.5.2.1. *Resampling* en desbalance de datos

En los problemas de clasificación en donde se debe etiquetar entre múltiples categorías se suele encontrar datos de entrenamiento entre las cuales existe una base minoritaria, es decir, para la cual se tienen de manera importante menor cantidad de muestras. Esta situación generará potencialmente problemas de desbalanceo y errores en la clasificación de datos de prueba, subestimando dichas categorías minoritarias.

Ante este tipo de problema, existen diferentes estrategias de abordaje. En particular, y con base en los resultados que se obtuvieron en la sección 2.3.1, se optó por implementar el muestreo artificial de datos u *oversampling* el cual consiste en crear muestras sintéticas (no idénticas) utilizando diversos algoritmos que intentan seguir la tendencia del grupo (o grupos) minoritarios.

En función de lo anterior es posible mejorar el proceso de clasificación que se utilice; sin embargo, debe tenerse en consideración el riesgo de alterar la distribución “natural” de las

⁹ Implementación en *Python* del paquete `tsfeatures` de R.

clases sobre muestreadas y confundir al modelo en su clasificación.

1.5.3. Desempeño de técnicas y del clasificador

Para la evaluación de la calidad y precisión de los técnicas y a la vez lograr compararlos de manera justa entre si, se utilizaron las métricas MAE y MAPE. La primera provee la noción de error expresada en la diferencia de unidades entre el valor real versus el valor predicho. En tanto, la segunda expresa el error como porcentaje del valor real por lo cual sirve para comparar resultados para distintas magnitudes evaluadas.

Además de MAE y MAPE, se utilizan las métricas específicas para medir clasificadores. En este caso la *Precision*, *Recall*, *Accuracy* y *F1*.

1.5.3.1. MAE (*Mean Absolute Error*)

Consiste en el promedio del valor absoluto de los errores, lo cuales están definidos como la diferencia entre el pronóstico y el valor real, como se muestra en la ecuación (1.3).

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - f_t| \quad (1.3)$$

Donde y_t representa el valor real en el periodo t, y f_t representa el valor pronosticado para el periodo t.

Al utilizar el MAE, debe considerarse que este arroja un número en las mismas unidades que los valores a evaluar, por lo tanto se puede interpretar fácilmente al analizar un solo producto. No obstante, esta métrica toma sentido en función de la magnitud de ventas del producto analizado, y por tanto no es comparable con otros que no correspondan a la misma magnitud. Por ejemplo, no es lo mismo tener un $MAE = 10$ ante un producto de venta promedio 100 que ante un producto con venta promedio 30.

1.5.3.2. MAPE (*Mean Absolute Percentage Error*)

Consiste en el promedio del valor absoluto de los errores porcentuales, lo cuales están definidos como la diferencia entre el pronóstico y el valor real, sobre el valor real como se muestra en la ecuación (1.4).

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - f_t}{y_t} \right| \quad (1.4)$$

Donde y_t representa el valor real en el periodo t, y f_t representa el valor pronosticado para el periodo t.

A diferencia del MAE, el valor de MAPE es un porcentaje y, por lo tanto, sirve para comparar rendimientos independientemente de las magnitudes de venta de los productos.

1.5.3.3. Métricas para el Clasificador

Para medir el comportamiento del clasificador se utilizaron distintas métricas que en conjunto generan información útil para tomar decisiones respecto al desarrollo de modelamiento y uso del clasificador *Random Forest*.

1.5.3.3.1. *Accuracy*

La **exactitud** (*accuracy*) mide el porcentaje de casos que el modelo ha acertado a nivel general. Esta métrica, si bien resume el comportamiento del modelo, no se usará por sí sola puesto que no es capaz por sí misma de representar el rendimiento de clases minoritarias en el caso de datos desbalanceados.

El cálculo se realiza sumando todos los casos bien clasificados y dividiendo por el total de datos testeados.

1.5.3.3.2. *Precision*

Con la métrica **precisión** se puede medir la calidad del modelo de *machine learning* en tareas de clasificación respecto a una clase en particular. Es decir, responde a “*De los datos clasificados en una clase ¿Cuántos realmente pertenecen a dicha clase?*”.

El cálculo se hace dividiendo el número de datos correctamente asignados a una clase entre el total de datos clasificados en esa misma clase.

1.5.3.3.3. *Recall*

La métrica de **exhaustividad** (*recall*) nos va a informar sobre la cantidad que el modelo de *machine learning* es capaz de identificar de una clase en particular. Es decir, responde a “*¿Qué cantidad de datos de una clase en particular se es capaz de clasificar correctamente?*”.

El cálculo se hace dividiendo el número de datos correctamente asignados a una clase entre el total de datos clasificados en esa misma clase.

1.5.3.3.4. *F1*

El valor F1 se utiliza para combinar las medidas de *precision* y *recall* en un sólo valor. Esto es práctico porque hace más fácil el poder comparar el rendimiento combinado de la precisión y la exhaustividad entre varias soluciones.

Este valor se calcula a través de la siguiente ecuación.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (1.5)$$

Capítulo 2

Metodología y desarrollo metodológico

En respuesta al planteamiento del problema, se utilizó como guía metódica el CRISP-DM¹⁰, cubriendo de este modo todos sus pasos y destacando el carácter de ciclo que esta metodología tiene, así como la importancia de iterar cada uno de sus componentes tantas veces sea necesario y en base al ciclo de vida de la minería de datos. Además, se destaca su naturaleza sencilla y flexible.

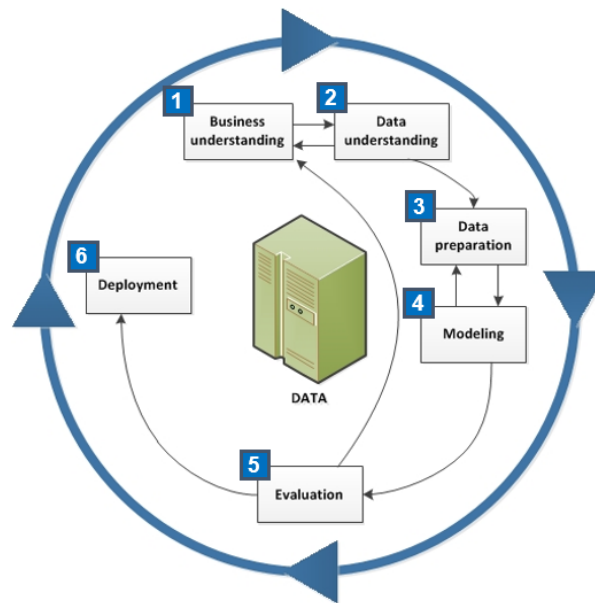


Figura 2.1: Ciclo de vida de minería de datos

En base al ciclo de vida de minería de datos de la metodología *CRISP-DM*, se plantearon como referente y se desarrollaron las siguientes etapas como plan de trabajo: entendimiento del negocio, entendimiento de los datos, preparación de los datos, modelamiento, evaluación y recomendación.

¹⁰ *Cross-Industry Standard Process for Data Mining.*

2.1. Entendimiento del negocio

Antes de trabajar con datos, se realizaron reuniones con las partes involucradas para lograr un entendimiento del negocio con principal foco en el proceso a intervenir, tanto de la programación como de la realización de compra de fármacos. A partir de este proceso es que se logra complementar los antecedentes de la empresa, expuestos en la sección 1.1 y la comprensión del problema planteado, expuesto en la sección 1.2.

Así mismo, a través de la indagación y el entendimiento de los datos tratado en la sección 2.2, a continuación, es posible complementar a través de números la información presentada en los antecedentes.

2.2. Entendimiento y preparación de los datos

Para determinar cómo se trabajará con las distintas técnicas de modelamiento, es necesario estudiar en profundidad los datos, comprender cómo se componen, establecer el nivel de completitud de de las series de tiempo, etc. Para ello se trabajó con lenguaje *SQL* y con la herramienta de exploración y análisis de datos *Tableau Desktop*.

Esta etapa se iteró de manera conjunta con la anterior para lograr un buen entendimiento del contexto del negocio y de los datos.

A medida que se fue explorando la base de datos, se fueron ejecutando modificaciones a modo de preparación para la posterior etapa de modelamiento. Dichos cambios también se fueron realizando iterativamente en conjunto con la siguiente etapa.

2.2.1. Estructura de los datos

Se tuvo acceso a la base de datos operacional de la Farmacia a través de la plataforma FarmaBot, de la cual es cliente. 103.692 transacciones fueron extraídas de una base de datos relacional la cual se inspeccionó en primera instancia con lenguaje *SQL*.

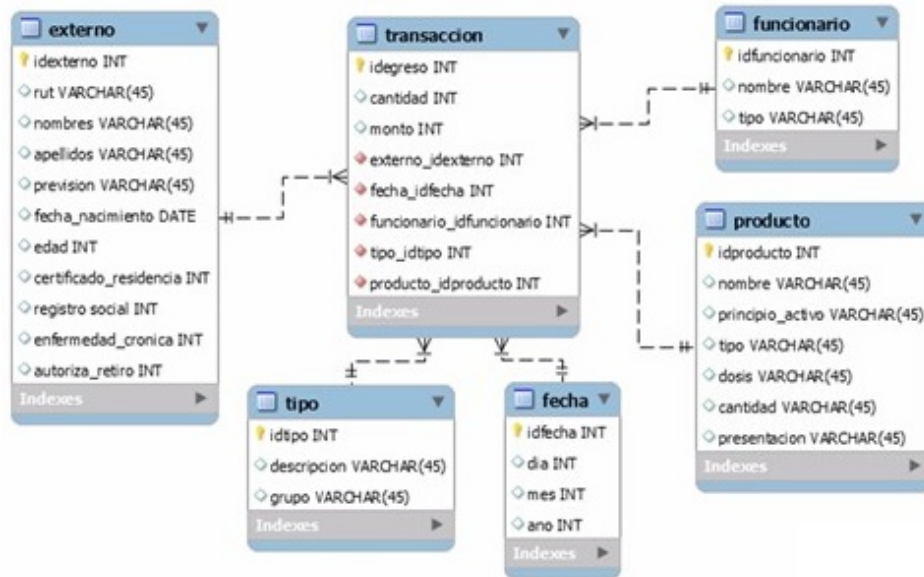


Figura 2.2: Modelo relacional del sistema transaccional de Farmacia de San Miguel

En función de las características del estudio de la memoria, se consideró que los atributos de compra/venta relevantes para el análisis son:

- Tabla **transaccion**:
 - **cantidad**: para determinar las unidades transadas
 - **monto**: para determinar el valor asociado a la transacción.
- Tabla **producto**: caracterizará cada medicamento para generar las series de tiempo. Por ejemplo, *PREGABALINA - 150 MG - 30 - CÁPSULAS*.
 - **principio activo**: en el ejemplo, *PREGABALINA*. Determina el componente activo del medicamento.
 - **dosis**: en el ejemplo, *150 MG*. Cantidad exacta de principio activo que contiene cada dosis/unidad del medicamento.
 - **presentacion**: en el ejemplo, *CÁPSULAS*. Forma farmacéutica en que se dispone el medicamento.
 - **cantidad**: en el ejemplo, *30*. Indica el número asociado a la presentación del medicamento.
- Tabla **fecha**: permiten situar la transacción en el tiempo con el uso de los atributos **dia**, **mes** y **ano**.
- Tabla **tipo**: hace referencia al tipo de transacción, compra de medicamentos (ingreso) por parte de la farmacia o venta (egreso) a clientes a través del atributo **grupo**.
- Tabla **externo**: hace referencia a la identidad de la contraparte en la transacción, ya sea cliente o proveedor.

- **nombres**: permite identificar cuando una transacción de venta corresponde a una merma (campo toma el valor “*EEspecial*”) o para identificar al proveedor cuando la transacción corresponde a un ingreso.

Si bien existe el campo **nombre** en la tabla **producto** (figura 2.2), el cual hace referencia al nombre comercial, este no se utiliza. Esto pues para un mismo medicamento (definido por su nombre) se puede tener distintas dosis, presentaciones o cantidades, no captando así las diferencias entre los ellos¹¹. En paralelo, ocurre que para un mismo principio activo (definido por su dosis, cantidad y presentación) se puede tener distintos nombres¹², por lo cual se separarían en distintas series de tiempo siendo medicamentos idénticos.

2.2.2. Hallazgos y modificación de datos

Para un análisis más profundo se complementó la indagación de los datos de compra y venta de medicamentos a través de la herramienta *Tableau Desktop*.

De las 103.692 transacciones presentes en la base de datos, 96.439 (93 %) corresponden a ventas, 6.830 (6,6 %) a ingresos a la farmacia y 423 (0,4 %) a egresos por merma.

Como se mencionó en la sección anterior, el análisis de esta memoria se basa en el trabajo sobre series de tiempos de medicamentos definidos por su **principio activo**, **dosis**, **cantidad** y **presentacion**, siendo estos los más relevantes para ser analizados, obteniendo los siguientes resultados.

| | |
|-------------------------------|-------------|
| Principio activo | 523 |
| Dosis | 472 |
| Cantidad | 92 |
| <i>Presentacion</i> | 41 |
| Total combinaciones | 1285 |
| Serie de tiempo ventas | 876 |

Tabla 2.1: Resumen de los datos.

Como se ve en la tabla 2.1, existen efectivamente 1.285 combinaciones entre los campos **principio activo**, **dosis**, **cantidad** y **presentacion**, implicando por lo tanto la existencia de esta cantidad de series de tiempo. No obstante, de estas 1.285 sólo 876 contienen ventas.

Es importante señalar la presencia de distintos errores e incongruencias en la base de datos, entre los que se encuentra la inconsistencia en el registro de ingresos de productos con respecto al stock de los productos¹³. Ante esto se tomó la decisión de trabajar con los datos de **venta** y no identificarlos como demanda pues no se tiene registro de las ventas perdidas por quiebre de stock, por ejemplo.

¹¹ Por ejemplo, una persona que debe comprar *PARACETAMOL - 125 mg - 6 - SUPOSITARIO*, no necesariamente le sirve el medicamento *PARACETAMOL - 500 mg - 16 - comprimido*, debido a que estos difieren en sus usos específicos distintos, aunque ambos tengan el mismo nombre comercial *Paracetamol*.

¹² Por ejemplo, para la combinación *PARACETAMOL - 1 G - 20 - COMPRIMIDO* existen dos nombres comerciales, *GESIDOL* y *XUMADOL COMPRIMIDOS*.

¹³ Por ejemplo, se identifico que para más casi un 50 % de los productos, el stock de productos (como resultado del ingreso de producto menos la venta) tomaba valores negativos en algún momento del tiempo.

Los errores que si fueron intervenidos, se abordaron tal como se muestra a continuación.

| Hallazgo | Modificación |
|--|--|
| Inconsistencia en separador de miles y decimales, algunos usando “,” (coma) y otros “.” (punto) de manera indistinta en el campo <i>dosis</i> . | Se determinó el uso de “.” (punto) como separador de miles y “,” (coma) como separador de decimales. Estos cambios fueron aplicados a la base de datos original a través de SQL. |
| Existencia de espaciado doble o ausencia de espaciado en los campos <i>dosis</i> , <i>cantidad</i> y <i>presentacion</i> . | A través de SQL se corrigieron estas particularidades directamente en la base de datos original reemplazando por espaciado simple en cada caso. |
| En el campo <i>presentacion</i> se identifica faltas de ortografía, errores de tipeo y nombres distintos para referirse a un mismo concepto ^a . | Se identificaron los casos, y tras ser validados por la contraparte en la farmacia, fueron hechas las correcciones correspondientes en la base de datos. |
| Existencia de <i>NULLs</i> y “Desconocido” en el campo <i>principio activo</i> . | Para evitar ignorar cualquier medicamento que no tenga un principio activo identificado, se procedió a repetir el nombre del medicamento en el campo <i>principio activo</i> a fin de individualizar dichos productos. |

Tabla 2.2: Principales modificaciones a la base de datos.

^a Por ejemplo, para la forma farmacológica *COMPRIMIDO* se utilizaron las palabras *COMPRIMIDO*, *COMPRIMIDOS* y *GRAJEAS*, dejando finalmente solo el correspondiente en los 3 casos.

En base a las correcciones realizadas, la tabla 2.1 se modifica, obteniendo los siguientes resultados:

| | Antes | Después |
|------------------------------|-------------|------------|
| Principio activo | 523 | 461 |
| Dosis | 472 | 349 |
| Cantidad | 92 | 71 |
| Presentación | 41 | 14 |
| Total combinaciones | 1285 | 931 |
| Serie de tiempo venta | 876 | 865 |

Tabla 2.3: Resumen de los datos después de las modificaciones.

Con las modificaciones realizadas, se obtiene un total de 865, sin embargo, solo el 585 de ellas (68%) cuentan con un año o más de historia. Además, de estas 865 sólo 356 series de tiempo cuentan con datos los últimos seis meses (entre octubre de 2020 y marzo de 2021), lo cual es relevante para el posterior modelamiento. En este sentido, se recalca que **sólo se trabajará con estas 356 series** debido a que cumplen con los requerimientos técnicos de suficiencia de datos tanto para el entrenamiento como el testeo de los modelos a utilizar en la siguiente etapa de la metodología, en particular, que existan más de tres meses con valores mayores a cero para el entrenamiento y seis meses con valores mayores a cero para el testeo.

2.2.2.1. Identificación de *outlier*

En la indagación de datos se identificaron datos correspondientes a *outliers*¹⁴ en las distintas series de tiempo en granularidad diaria (registro transaccional de ventas). Sin embargo, estos no fueron excluidos debido a que al agregar mensualmente las ventas (formato que finalmente se utilizó en el desarrollo), dichos *outliers* no generaron cambios significativos en las series de tiempo.

En los casos de existencia de *outliers* en la presentación agregada mensual de las series, estos tampoco se descartaron, debido a que se supuso la pérdida de información. En muchos de los casos, estos representan potenciales estacionalidades en el largo plazo¹⁵, las cuales se detectaron y utilizaron posteriormente para generar predicciones a través de las técnicas ARIMA y *Prophet*. Adicionalmente, en este caso, se consideró también la limitada cantidad de meses disponibles para el desarrollo del trabajo en cada una de las series de granularidad mensual, por lo cual la eliminación de dichos datos podría significar el descarte de la serie completa a la que pertenece, lo cual habría disminuido aún más la cantidad de productos a analizar.

2.2.3. Visualización de series

Dentro de las series de tiempo que se utilizaron, existe una distribución amplia en el promedio de ventas, donde algunas de las series tienen promedio 1 venta por mes y la más grande, con 150.

En base a lo descrito, se presumió que esto podría tener un impacto en el comportamiento de las técnicas usadas. Por lo tanto, para efectos prácticos de análisis, cada serie de tiempo fue clasificada en un grupo definido de la siguiente manera:

- **Grupo 1:** considera 180 series de tiempo (50,6% del total) y abarca desde promedio 0,7 hasta 6,4 aproximadamente.
- **Grupo 2:** considera 140 series de tiempo (39,3% del total) y abarca desde promedio 6,4 hasta 31 aproximadamente.
- **Grupo 3:** considera 36 series de tiempo (10,1% del total) y abarca desde promedio 31 hasta 149,7 aproximadamente.

¹⁴ Un *outlier* (valor atípico) es una observación extrañamente grande o pequeña. Los valores atípicos pueden tener un efecto desproporcionado en los resultados estadísticos, como la media, lo que puede conducir a interpretaciones engañosas.

¹⁵ Esto se da en las series de tiempo que cuentan con menos de 20 meses, en las cuales no es posible aseverar si una variación abrupta en las ventas corresponde a un *outlier* o una expresión de estacionalidad.

Distribución de promedios

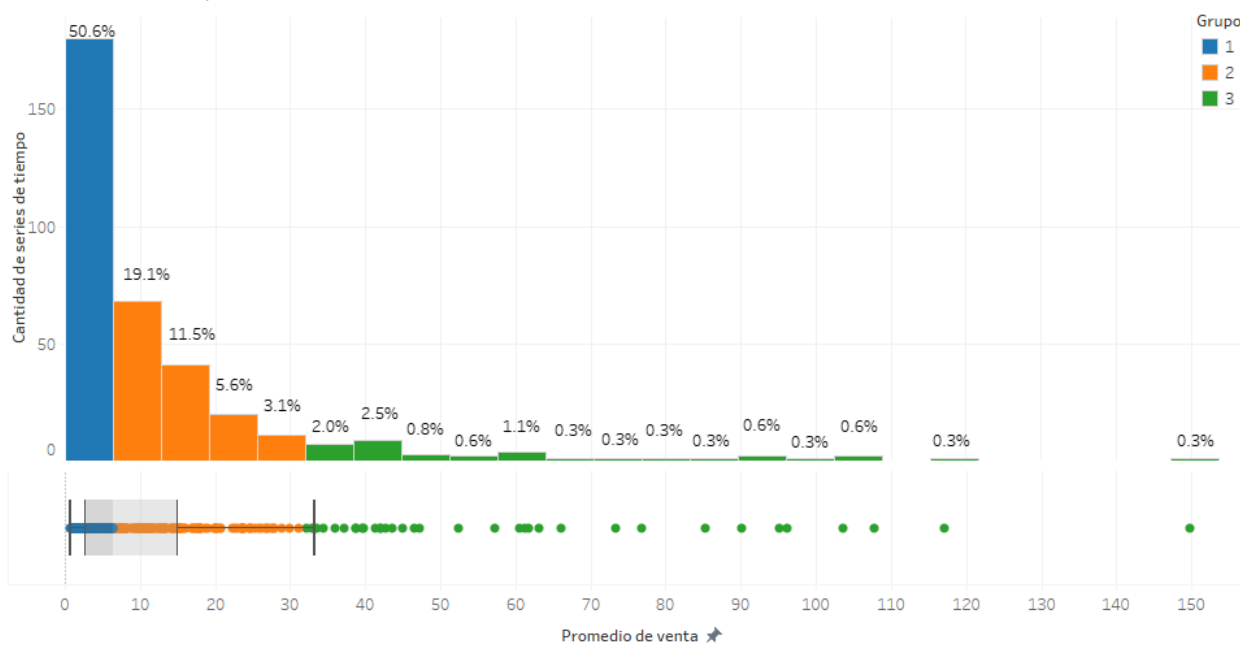


Figura 2.3: Distribución de promedios de venta para cada medicamento (serie de tiempo).

Para facilitar el agrupamiento se utilizó la herramienta *bins* de *Tableau* (Tableau, 2020), de manera que se encontró el tamaño óptimo para subgrupos de datos (en este caso el valor de corte es $6, 4^{16}$).

La agrupación final de series se hizo en base a este agrupamiento *bins* y a la distribución de los promedios de ventas de estas, donde el grupo 1 está definido por ser el grupo que contiene el primer intervalo *bins*, el cual además coincide con el valor de la mediana. Para el segundo grupo, se acotó superiormente por el valor máximo del *boxplot*¹⁷. Por último, el grupo 3 está conformado por los *outliers* como se puede observar en la figura 2.3.

En cuanto a la venta de productos, los principios activos más vendidos son los siguientes.

¹⁶ Es decir, se agrupan todos los elementos que se encuentren en los intervalos de $6, 4n$, con n la cantidad de grupos necesarios para abarcar todos los promedios de venta.

¹⁷ Un *boxplot* es una representación visual que describe varias características importantes, al mismo tiempo, sobre un conjunto de datos, tales como la dispersión, simetría, etc. Para su realización se representan los tres cuartiles y los valores mínimo y máximo de los datos sobre un rectángulo, alineado horizontal.

Venta agregada por principio activo

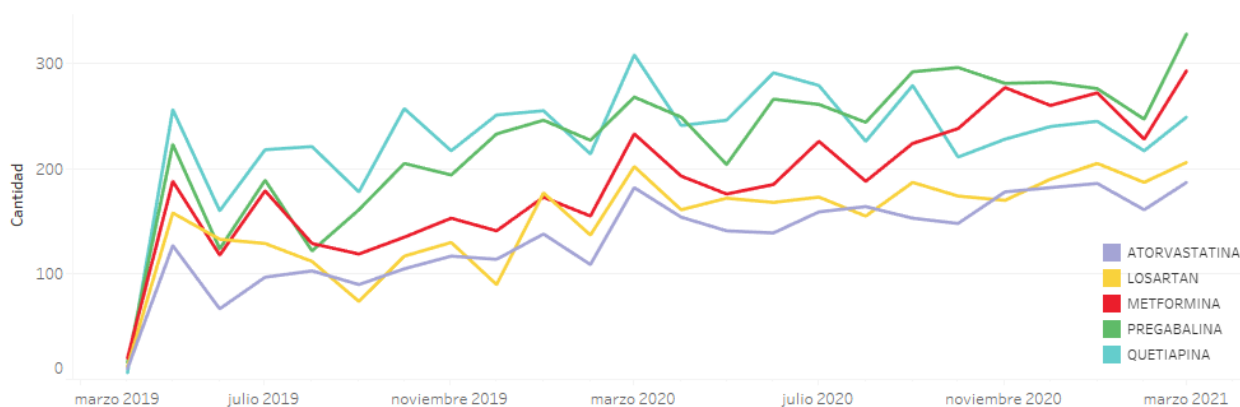


Figura 2.4: Venta de mejores productos agregados por principio activo.

Como se mencionó antes, las series de tiempo están definidas por las características del medicamento; sin embargo, a través del análisis agregado considerando solo el principio activo, es posible identificar cinco de estos especialmente demandados.

Los principios activos son *Atorvastatina*, *Losartan*, *Metformina*, *Pregabalina* y *Quetiapina*. Dichos principios activos están precisamente asociados a cinco de las principales enfermedades crónicas preponderantes en Chile (colesterol alto, hipertensión arterial, diabetes, dolores neuropáticos y trastornos mentales, respectivamente) (Zitko y Aceituno, 2019). Dichos medicamentos han incrementado su venta en el último año en aproximadamente un 50%, lo cual incrementa el interés de la dirección de la farmacia por anteponerse a la demanda para lograr satisfacerla oportunamente.

Se eligieron tres medicamentos de interés para ser mostrados en el análisis, en base a su importancia para la farmacia y por su comportamiento en los diferentes grupos asignados. Estos además serán analizados en la sección 2.3.1.3.

Venta medicamentos de interés analítico por grupo

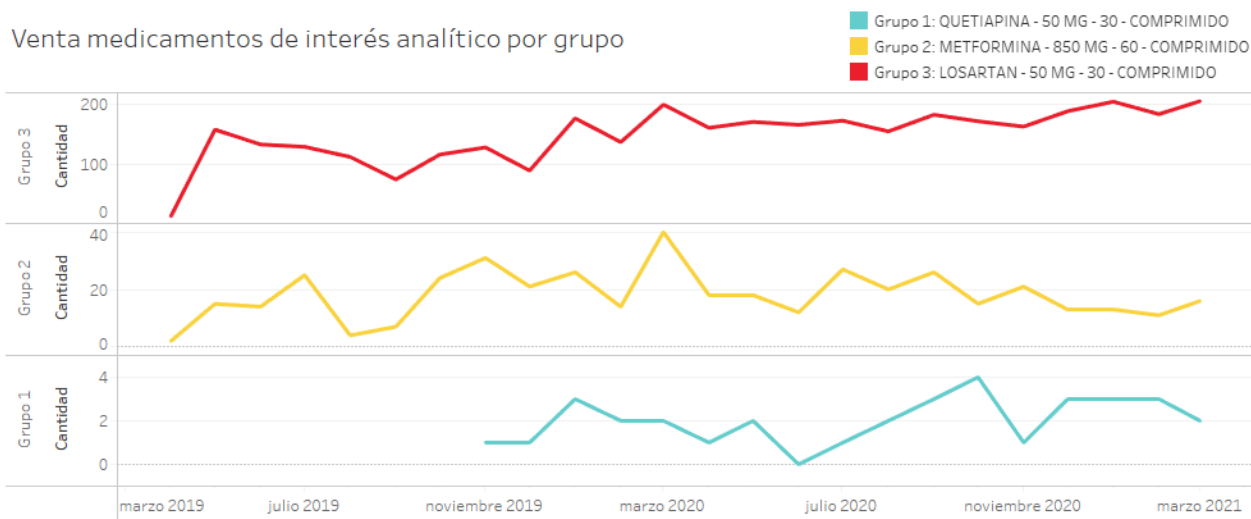


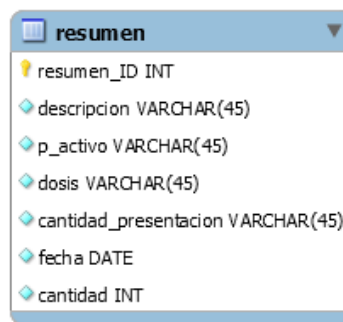
Figura 2.5: Venta de medicamentos de interés analítico por grupo asignado.

Tal como se ve en estas series de tiempo, en el resto de los medicamentos no se detectaron a primera vista estacionalidades, la cual se atribuye inicialmente a una limitada extensión de tiempo de las series (la más larga cuenta con 25 meses de ventas). No obstante, en algunas series se divisan grandes variaciones entre sus puntos máximos y mínimos, así como también otras se mantienen estables en el tiempo con tendencias evidentes, características que serán captadas e integradas por los distintas técnicas a usar.

2.2.4. Preparación de los datos

Profundizando las modificaciones anteriores, se prepararon los datos con un foco en el posterior modelamiento generando una gran tabla resumen que contuviese todas las series de tiempo en la misma base de datos a través de códigos *SQL*. Para ello se siguieron los pasos que a continuación se describen:

- a) Generación de tabla auxiliar de periodos, la cual contiene todas las combinaciones mes-año necesarias para el modelamiento, en este caso desde marzo de 2019 hasta marzo de 2021 (como el interés de la farmacia es obtener datos mensuales, se trabajó con esta granularidad, por lo tanto a cada mes se le identificó con el día 1).
- b) Generación de tabla auxiliar de medicamentos, la cual contiene todas las combinaciones “*principio activo - dosis - cantidad - presentacion*” que se utilizarán. Para ello se consulta todo medicamento que haya sido **egresado** en la tabla de transacciones, y que cumpla con no ser merma. Adicionalmente, se filtran todos los medicamentos que no cumplen con la cantidad de periodos con ventas como se señaló en el último párrafo de la sección 2.2.2.
- c) Armado de resumen auxiliar de todas las transacciones, debidamente identificadas con el nombre del medicamento y la fecha (granularidad mensual).
- d) Por último se creó y completó la **tabla resumen** a partir de las tablas auxiliares de los pasos *a)*, *b)* y *c)*, la cual contiene los campos señalados en la figura 2.6:



| Field Name | Field Type |
|-----------------------|-------------|
| resumen_ID | INT |
| descripcion | VARCHAR(45) |
| p_activo | VARCHAR(45) |
| dosis | VARCHAR(45) |
| cantidad_presentacion | VARCHAR(45) |
| fecha | DATE |
| cantidad | INT |

Figura 2.6: Estructura de la tabla resumen.

Es importante señalar que para cada serie de tiempo se utilizaron 6 periodos (meses) de evaluación independientemente del modelo a utilizar, por lo tanto, la proporción de testeo sobre el total de datos no será la misma en todos los casos.

Proporción de datos de testeo

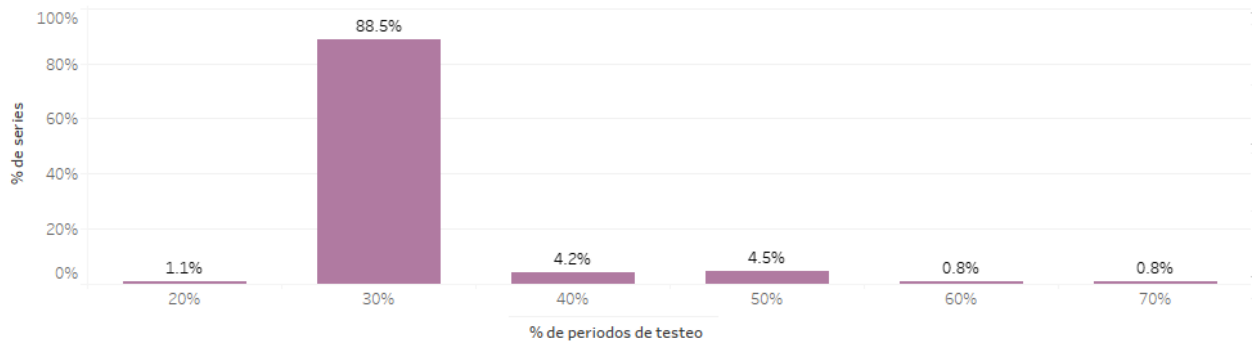


Figura 2.7: Proporción de datos de testeo.

A pesar de que en algunos casos la proporción de datos de testeo es alta respecto al total de la serie, casi un 90% de las series cuenta con más del 70% de los datos disponibles para el entrenamiento de las técnicas, es decir, tienen al menos 18 meses de historia.

2.3. Modelamiento de técnicas y evaluación

Con base en las etapas anteriores, se programó en Python los distintos comandos, con las librerías pertinentes, que permitieron generar los modelos de pronóstico, las correspondientes simulaciones y las evaluaciones de los resultados obtenidos, a fin de seleccionar modelos adecuados y con los menores errores posibles.

Como se indicó en secciones anteriores, se utilizaron 6 periodos (correspondientes a los meses octubre de 2020 a marzo de 2021) de testeo para todas las técnicas a utilizar y para todas las series de tiempo, independiente de su extensión. Esto para garantizar la justa comparación entre los resultados obtenidos por cada técnica, además de llevar a uso el escenario real que se busca implementar en la farmacia¹⁸.

En primer lugar, fue implementado el modelamiento de las técnicas, mencionadas en el capítulo 1.5.1, a cada una de las series de tiempo determinadas en la preparación de datos, obteniendo valores de desempeño MAE y MAPE. En base a dichos resultados se realizan diferentes análisis los cuales son: determinar los rendimientos de cada técnica individualmente, evidenciar la mejora de rendimientos utilizando las mejores técnicas para cada serie de tiempo, y el análisis de los medicamentos más relevantes. La finalidad de mostrar estas tres secciones, radica en evidenciar los distintos rendimientos que se obtienen en base al uso de las distintas técnicas para distintos tipos de serie de tiempo según su nivel de venta.

Posteriormente, se aplicó el modelo de clasificación *Random Forest* para determinar la asignación de la técnica más adecuada a cada serie de tiempo en base a las características de estas últimas y del rendimiento que la técnica consiga, realizando nuevamente un análisis que permita determinar el desempeño con dicha herramienta.

¹⁸ La farmacia necesita contar con 5 meses de predicción, por lo tanto se considerará la proyección de 6 (actual + 5 hacia adelante), debido a que se presume el desconocimiento de datos completos del mes en que se realiza la proyección.

Dentro de cada sección se evalúa cada predicción a través de las métricas MAE y MAPE, de modo de determinar por medio de los errores, los rendimientos de las técnicas para cada serie de tiempo, y establecer así cuáles son más adecuados en cada caso. Se procedió a hacer revisiones de los resultados y los alcances que se lograron obtener, iterando durante la realización del trabajo para enriquecer el proceso.

2.3.1. Modelamiento de técnicas

Para lograr aplicar las distintas técnicas, estas se modelaron en códigos independientes de manera que fuesen ejecutables y medibles por separado.

En primer lugar se conectó la base previamente arreglada con formato de series de tiempo, en donde 865 series de tiempo fueron dispuestas de la siguiente manera:

- Aplicar cada técnica por separado al total de series.
- Guardar resultados en una tabla de estructura: *Descripción_Producto*, *Fecha*, *Venta_real*, *Valor_Técnica_1*, *Valor_Técnica_2*, *Valor_Técnica_3*, (...); de manera de contar con todas las predicciones para cada serie de tiempo con cada una de las técnicas¹⁹.
- Fueron calculadas las métricas MAE y MAPE para cada una de las técnicas, almacenando dichos datos en una tabla de estructura: *Descripción_Producto*, *Técnica*, *MAE*, *MAPE*²⁰.

A partir de esto se logra desarrollar el análisis de la sección 2.3.1.1 explicado a continuación.

2.3.1.1. Resultados de cada técnica

Al observar los resultados generales (tabla 2.4), se puede apreciar que en términos de MAPE, la técnica *Modelo Actual* se comporta mejor para la totalidad de los productos en promedio. Por otro lado, en base a MAE, la que mejor rendimiento tiene es *SMA5* pero dicha métrica en este caso pierde relevancia debido a que se están comparando series con magnitudes distintas en promedio.

Lo anterior puede ser una señal de que, si bien en promedio la técnica *Modelo Actual* es mejor porcentualmente, la técnica *SMA5* tiene mejor rendimiento en series con valores más elevados.

En cuanto a la comparación entre tiempos de ejecución, es posible identificar que conforme la complejidad de la técnica implementada aumenta, el tiempo de ejecución también lo hace. En este caso, para la comparación entre las técnicas de mejor rendimiento, no es útil la consideración del tiempo de ejecución puesto que son parecidas dada la similar complejidad de estas.

¹⁹ Ver anexo A para conocer la estructura de tabla en que se guardan los resultados de las predicciones de las distintas técnicas utilizadas.

²⁰ Ver anexo B para conocer la estructura de la tabla en que se guarda los resultados de rendimientos.

| | MAPE | MAE | T ejecución ^a |
|-----------------|---------|------|--------------------------|
| Modelo Actual | 43,9 % | 5,6 | ~ 1 min |
| SMA5 | 46,5 % | 5,4 | ~ 1 min |
| ARIMA | 48,3 % | 6,0 | ~ 8 min |
| Prophet | 59,2 % | 5,8 | ~ 24 min |
| Prophet forzado | 300,1 % | 30,4 | ~ 124 min |

Tabla 2.4: Resultados generales.

^a Es de considerar que los tiempos de ejecución aquí señalados pueden variar en función de la capacidad de procesamiento del computador en que se realice. En este caso, todas las ejecuciones fueron realizadas en una misma máquina siendo, por lo tanto, comparables.

Al analizar los resultados de cada técnica utilizando las agrupaciones definidas en la sección 2.2.3, es posible detectar en primera instancia que los rendimientos de la técnica *Modelo Actual*, *SMA5*, *ARIMA* y *Prophet* son similares, en comparación con los grandes valores de MAPE y MAE que presenta la técnica *Prophet con estacionalidad anual forzada* (ver figura 2.8).

Resultados generales por grupo

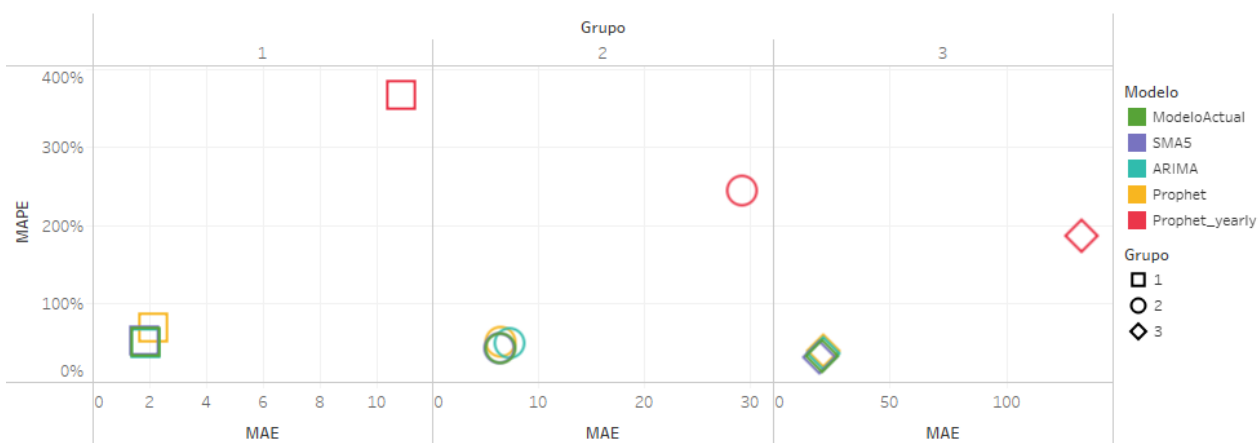


Figura 2.8: Resultados generales por grupo.

Al excluir los resultados de *Prophet con estacionalidad anual forzada* (figura 2.9), se advierten sutiles diferencias entre los rendimientos de las otras técnicas. Entrando más en detalle, a través de la figura 2.10, es posible observar que para el Grupo 1, basados en el promedio del MAPE, sería *ARIMA* la técnica que mejor da respuesta, sin embargo, este resultado se encuentra muy cerca a los rendimientos de *SMA5* y más aun de la técnica *Modelo Actual*, tanto en MAPE como en MAE.

Resultados generales por grupo

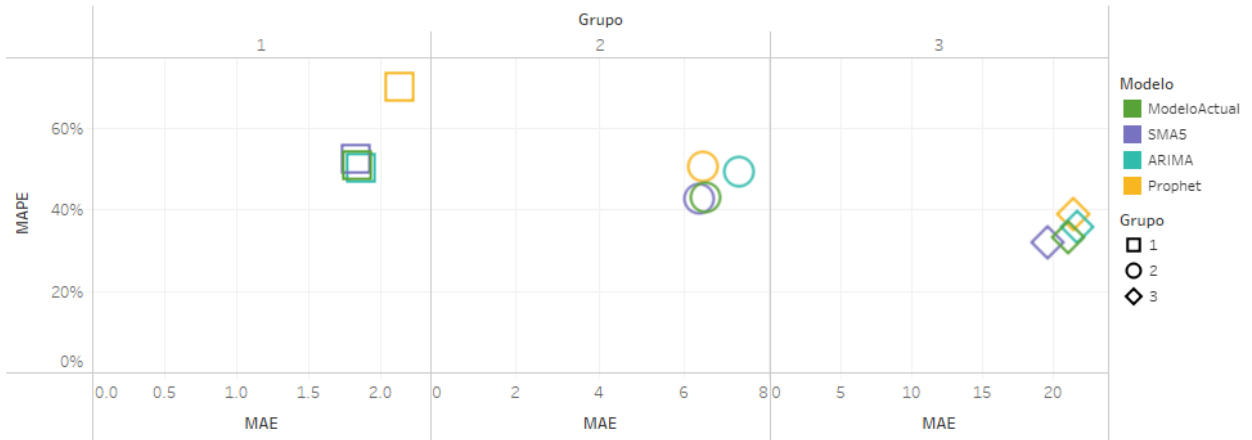


Figura 2.9: Resultados generales por grupo sin *Prophet con estacionalidad anual forzada*.

Por otro lado, para los grupos 2 y 3, por una leve diferencia contra la técnica *Modelo Actual*, el *SMA5* muestra mejores resultados tanto por MAPE como por MAE. Con todo, en el Grupo 3 la diferencia por MAE evidencia un poco más la ventaja de *SMA5* sobre la técnica *Modelo Actual*.

Detalle rendimientos generales

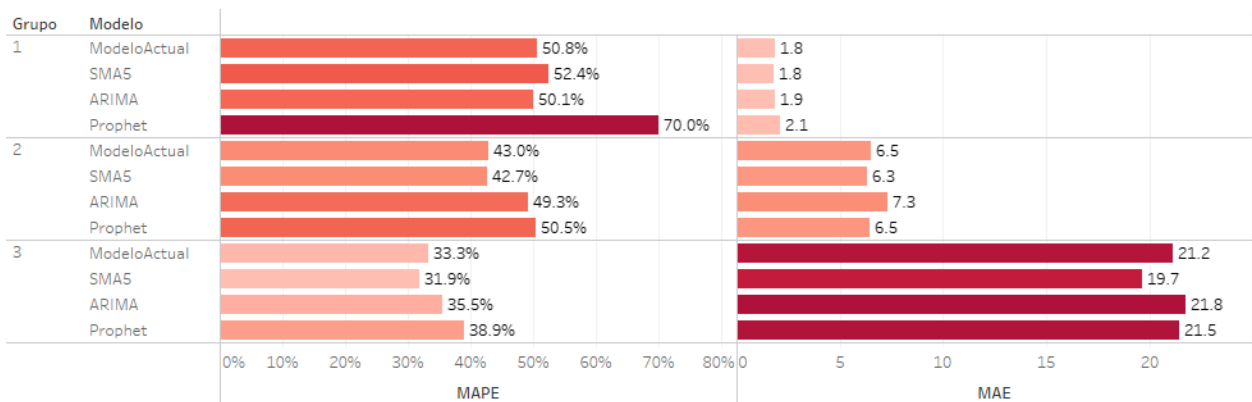


Figura 2.10: Detalle resultados generales por grupo sin *Prophet con estacionalidad anual forzada*.

En base a esto, el orden de precisión en base a MAPE sería:

| | 1° | 2° | 3° | 4° | 5° |
|----------------|-------|---------------|-------|---------|-----------------|
| Grupo 1 | ARIMA | Modelo Actual | SMA5 | Prophet | Prophet forzado |
| Grupo 2 | SMA5 | Modelo Actual | ARIMA | Prophet | Prophet forzado |
| Grupo 3 | SMA5 | Modelo Actual | ARIMA | Prophet | Prophet forzado |

Tabla 2.5: Ranking resultados por grupos en base a MAPE.

Pese a que es posible identificar que algunas de las técnicas son mejores que otras a nivel general, esto no permite generar una conclusión ni mucho menos entregar una solución al problema planteado, pues aún cuando es factible identificar mejores técnicas para cada

grupo, esto no resulta suficiente dado que deja afuera los resultados individuales de cada serie.

2.3.1.2. Resultados utilizando mejores técnicas

Considerando la tabla de métricas mencionada anteriormente se procedió a hacer la selección de mejores técnicas en una tabla de estructura: *Descripcion_Prod*, *Modelo_MAPE*, *MAE*, *MAPE*.

Se realizó la selección de las mejores técnicas en base a la métrica MAPE para cada una de las series de tiempo. A partir de esto se obtiene los siguientes resultados.

Resultados mejores modelos

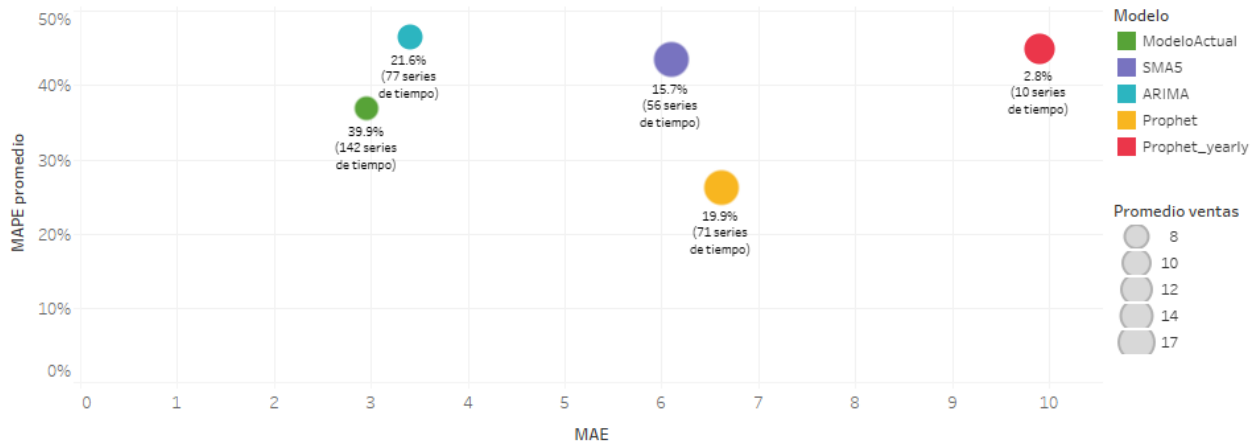


Figura 2.11: Resultados de mejores técnicas.

Se identifica que, a pesar de las diferencias en MAE, los errores porcentuales (MAPE) promedio son similares para la *SMA5* con 43%, *ARIMA* con (47%) y para *Prophet con estacionalidad anual forzada* con 45%. Las diferencias en MAE se presumen asociadas al promedio de ventas de las series a las que se ajustan, lo cual indica como las distintas técnicas responden mejor a distintos tipos de serie.

Con mejor rendimiento se encuentra a *Modelo Actual*, con un MAPE promedio de 37% y casi un 40% de las series ajustadas con esta como mejor técnica. Es relevante que el porcentaje de series ajustadas sea así de alto, debido a que implica que no es necesario utilizar grandes recursos (tiempo que requieren técnicas más complejas) para lograr obtener buenos resultados en la ejecución general del código.

Por último, se encuentra la técnica *Prophet*, con un 20% de series ajustadas y con un promedio MAPE de 26%, en donde además, como en los casos anteriores, el MAE aumenta dado el nivel promedio de ventas de los productos ajustados.

Resultados mejores modelos

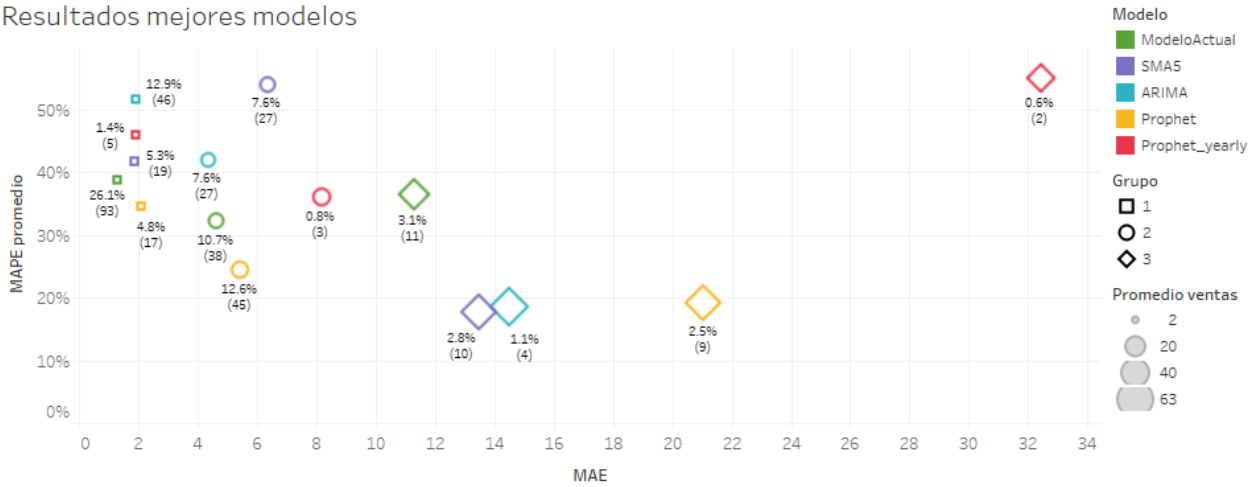


Figura 2.12: Resultados de mejores técnicas por grupo.

Al analizar los resultados en función de los grupos asignados, es posible confirmar el hallazgo de incidencia del promedio de ventas del grupo sobre el MAE promedio obtenido. Se rescata además, que tanto la distribución de MAE como de promedios de venta son distintas para cada técnica utilizada.

En particular, se pudo observar que para series estables (con baja varianza) se suele asignar técnicas más sencillas, generando menor MAE; en tanto, para series más complejas (de mayor varianza) suelen ajustar mejor técnicas más elaboradas como el *ARIMA*, *Prophet* y *Prophet con estacionalidad anual forzada*.

| | MAPE | MAE | T ejecución |
|------------------|--------|-----|-------------|
| SMA5 | 46,5 % | 5,4 | ~ 1 min |
| Mejores técnicas | 38,1 % | 4,5 | ~ 4 min |

Tabla 2.6: Comparación resultados *SMA5* versus selección de mejores técnicas.

Si se miran los resultados comparados con la técnica que en general tuvo mejor rendimiento en la sección anterior, es posible notar que hubo esperables mejoras tanto en el rendimiento de MAPE como de MAE; no obstante, es necesario recalcar que esto involucra el uso de mayor cantidad de tiempo de ejecución, pasando de 1 a 4 minutos aproximadamente.

2.3.1.3. Resultados de medicamentos relevantes

En base al sección 2.2.3, se analizan brevemente los medicamentos Losartán, Metformina y Quetiapina por su relevancia con respecto a la cantidad de ventas que registran en los 2 años de historia disponibles del sistema transaccional de la farmacia.

Para el análisis se consideran tanto los datos reales como las proyecciones arrojadas por las distintas técnicas.

Cabe mencionar que para efectos prácticos se omitieron las proyecciones generadas por la técnica de *Prophet con estacionalidad anual forzada* en las figuras siguientes, debido a

ajustes desproporcionados que no permiten apreciar correctamente el resto de los resultados (Ver anexo A para conocer la estructura de la tabla que guarda los resultados).

2.3.1.3.1. Análisis *LOSARTAN - 50 MG - 30 - COMPRIMIDO*

La serie de tiempo asociada a las ventas del producto *LOSARTAN - 50 MG - 30 - COMPRIMIDO* pertenece a la clasificación de Grupo 3, en particular con un promedio de 118 ventas por mes en toda su historia y con 177 unidades vendidas promedio el último año y, en general, presenta una tendencia de crecimiento positiva.

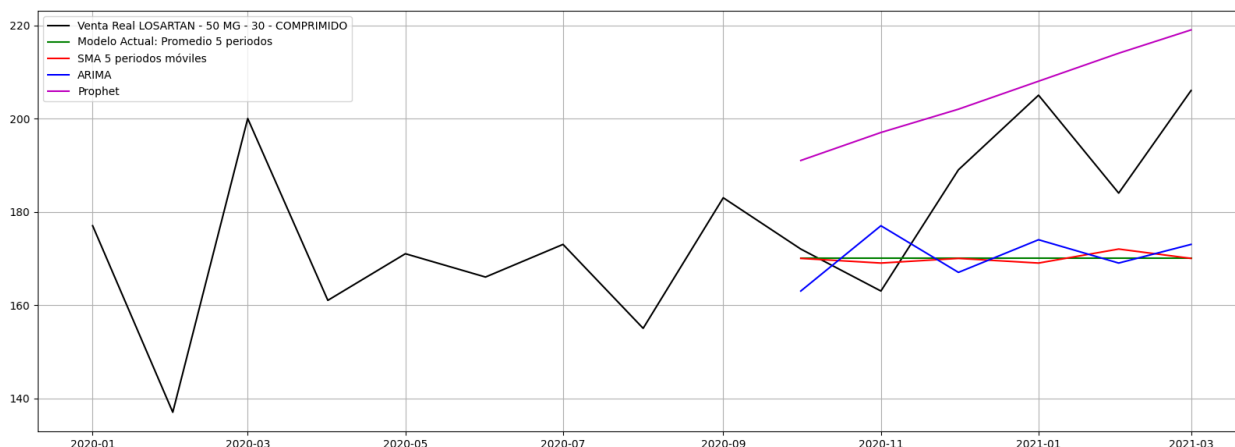


Figura 2.13: Ajustes de distintas técnicas a la serie de tiempo asociada a ventas de *LOSARTAN - 50 MG - 30 - COMPRIMIDO*

En comparación a otras series de tiempo del grupo al que pertenece, este medicamento tiene una elevada varianza. Dicha varianza dificulta un buen ajuste de las distintas técnicas de predicción (ver figura 2.13), obteniendo altos valores tanto en la métrica MAPE como en MAE (ver tabla 2.7).

| | MAPE | MAE |
|-----------------|---------|-------|
| Modelo Actual | ~ 10 % | 18,8 |
| SMA5 | ~ 9 % | 18,5 |
| ARIMA | ~ 11 % | 20,7 |
| Prophet | ~ 10 % | 18,7 |
| Prophet forzado | ~ 187 % | 334,3 |

Tabla 2.7: Resumen métricas de resultados para *LOSARTAN - 50 MG - 30 - COMPRIMIDO*.

En consideración de los resultados de métricas, es posible identificar que las técnicas *ARIMA*, *SMA5*, *Modelo Actual* y *Prophet*, tienen comportamientos similares; sin embargo, en base a la selección a través de MAPE, es posible asignar el que mejor ajuste logra dentro de lo posible, siendo *SMA5* el elegido.

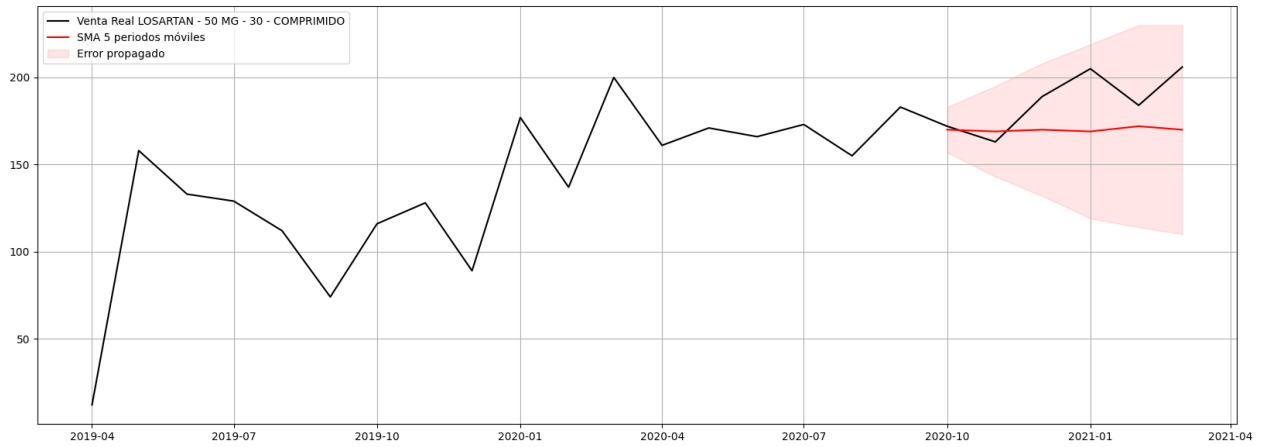


Figura 2.14: Ajuste de técnica *SMA5* para la serie de tiempo de *LOSARTAN - 50 MG - 30 - COMPRIMIDO*

En base a los valores de MAPE y MAE, es crucial tomar decisiones de abastecimiento que apunten al *sobreestockeo*, sobre todo por la gran importancia que tiene este medicamento en la población que lo consume y la gran demanda con que cuenta.

2.3.1.3.2. Análisis *METFORMINA - 850 MG - 60 - COMPRIMIDO*

La serie de tiempo asociada a las ventas del producto *METFORMINA - 850 MG - 60 - COMPRIMIDO* pertenece a la clasificación de Grupo 2, con un promedio de 16,4 ventas por mes en toda su historia y con 17,5 unidades vendidas promedio el último año y, en particular no se identifica una tendencia evidente.

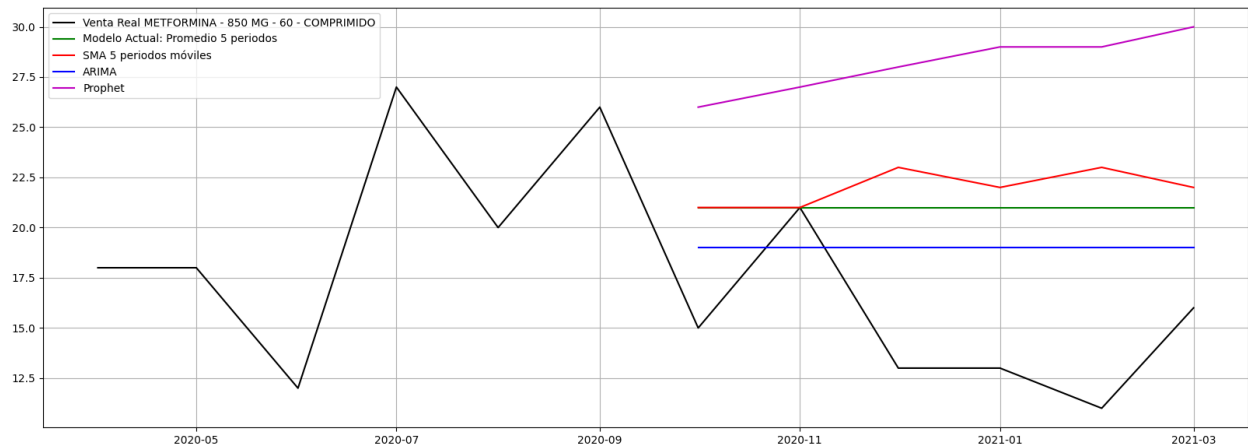


Figura 2.15: Ajustes de distintas técnicas a la serie de tiempo *METFORMINA - 850 MG - 60 - COMPRIMIDO*

En general, y en comparación a otras series de tiempo del grupo al que pertenece, este medicamento tiene una elevada varianza. Dicha varianza, dificulta un buen ajuste de las distintas técnicas de predicción (ver figura 2.15), obteniendo altos valores tanto en la métrica MAPE como en MAE (ver tabla 2.8).

| | MAPE | MAE |
|-----------------|---------|------|
| Modelo Actual | ~ 48 % | 6,2 |
| SMA5 | ~ 55 % | 7,2 |
| ARIMA | ~ 37 % | 4,8 |
| Prophet | ~ 99 % | 13,3 |
| Prophet forzado | ~ 103 % | 15,2 |

Tabla 2.8: Resumen métricas de resultados para *METFORMINA - 850 MG - 60 - COMPRIMIDO*.

En base a los resultados de métricas, es posible identificar que la técnica *SMA5* logra un mejor comportamiento con un MAPE de 37% aproximadamente. Aun siendo el mejor ajuste entre los disponibles, este continúa siendo alto. Sin embargo, la baja cantidad de ventas promedio de ventas de este medicamento genera que el error sea de muy pocas unidades ($MAE = 4,8$), y en donde se podría tomar la opción de *sobrestockear* a la farmacia debido al probable bajo impacto que pudiese tener la adquisición de 5 unidades extra del medicamento por mes.

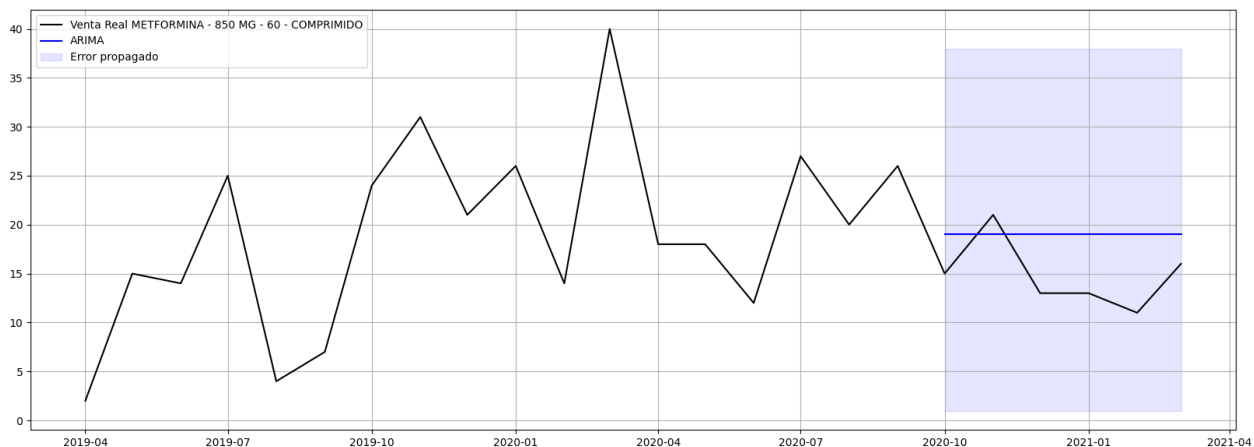


Figura 2.16: Ajuste de técnica *ARIMA(0,0,0)* para la serie de tiempo de *METFORMINA - 850 MG - 60 - COMPRIMIDO*

2.3.1.3.3. Análisis *QUETIAPINA - 50 MG - 30 - COMPRIMIDO*

La serie de tiempo asociada a las ventas del producto *LOSARTAN - 50 MG - 30 - COMPRIMIDO* pertenece a la clasificación de Grupo 3, en particular con un promedio de 0,9 ventas por mes en toda su historia y con 2,1 unidades vendidas promedio el último año.

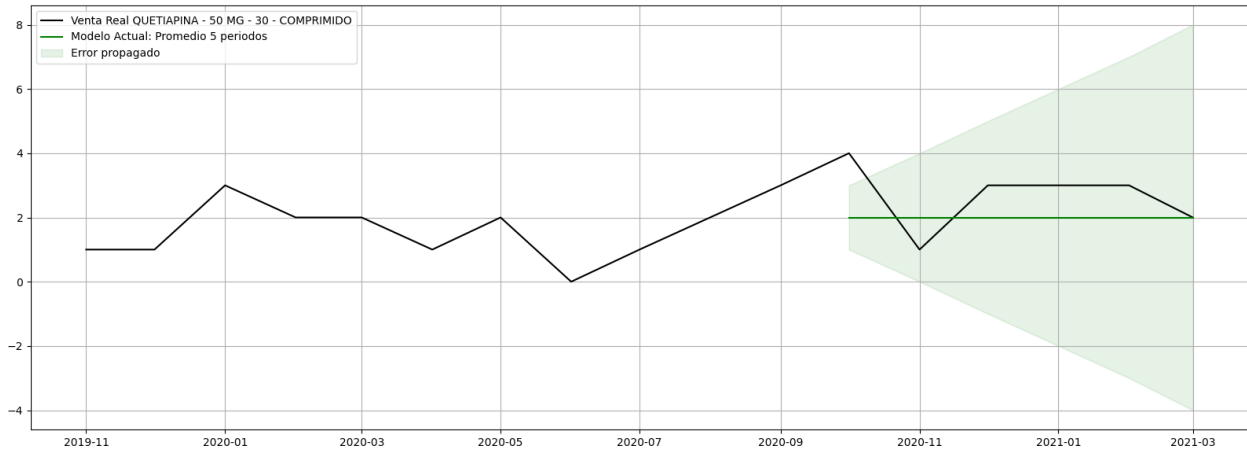


Figura 2.17: Ajuste de técnica *Modelo Actual* para la serie de tiempo de *QUETIAPINA - 50 MG - 30 - COMPRIMIDO*

Si bien la serie de tiempo no presenta una gran varianza, sí tiene fluctuaciones en el tiempo que dificultan un ajuste certero de las distintas técnicas aplicadas. Con todo, 4 de las 5 técnicas aplicadas coinciden en los mismos valores de predicción y, por lo tanto, con igual MAPE y MAE.

| | MAPE | MAE |
|-----------------|---------|-----|
| Modelo Actual | ~ 42 % | 1 |
| SMA5 | ~ 42 % | 1 |
| ARIMA | ~ 42 % | 1 |
| Prophet | ~ 42 % | 1 |
| Prophet forzado | ~ 328 % | 9,7 |

Tabla 2.9: Resumen métricas de resultados para *QUETIAPINA - 50 MG - 30 - COMPRIMIDO*.

Ante esta situación, y en consideración de que la técnica *Modelo Actual* es más sencilla y tiene un costo (en tiempo) menor de ejecución, se determina establecer dicha técnica como la mejor en este caso.

2.3.2. Modelamiento de clasificador

El procedimiento de esta sección nace a partir de la presunción de que variadas series de tiempo asignadas a una misma técnica, pudieran compartir características desde el punto de vista estadístico.

A partir de los resultados de la sección 2.3.1.2, se utiliza la selección de técnicas que mejor respondieron entre las usadas para cada producto. Adicionalmente, en base a las ventas, se calculan las características asociadas a cada serie de tiempo. Para ello se utilizó el paquete *tsfeatures* ampliamente usado para la caracterización estadística de series de tiempo.

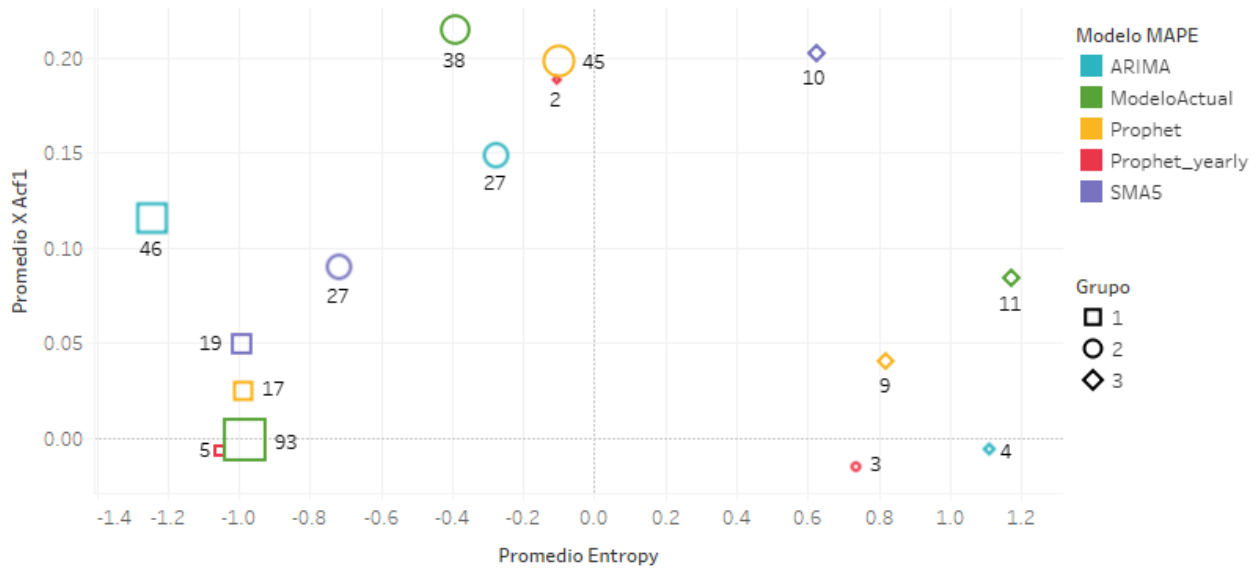


Figura 2.18: Distribución de valores de variables X_ACF1 y *entropía* entre las distintas técnicas asignadas.

Un ejemplo que apoya la idea antes señalada se puede observar en la figura 2.18, en donde se identifica que el promedio de la variable X_ACF1 ²¹ y de la *entropía*, se distribuyen de distinta manera según la técnica utilizada y el grupo al que pertenece cada serie (que hace referencia al nivel o promedio de venta de la misma).

A partir de la utilización de *tsfeatures*, se generaron un total de 43 variables para cada serie de tiempo. Sin embargo, de estas sólo son utilizables 30, debido a que muchas de ellas no entregaron información. En particular, la mayoría de los descartes se debieron a la existencia de valores invariables entre series²² y existencia de *NAs* en la mayoría de las series. Por ejemplo, para una misma característica, se detectaron un 98% de *NAs* para el total de los productos).

Además, existieron algunas series que contenían *NA's* aislados en alguna variable, esto principalmente asociado a que la posibilidad de cálculo de dichas características dependían del largo de la serie de tiempo. Para evitar disminuir dramáticamente el número de datos, estos *NA's* se reemplazaron por el valor 0, lo cual, dada la naturaleza de los *NA's*, no debiera presentar problema para el clasificador pues este valor explica precisamente una característica de la serie de tiempo.

Por otro lado, se intentó disminuir variables que hicieran referencia a una misma característica de la serie o en caso de que existiera correlaciones importantes. Sin embargo, como se aprecia en la figura 2.19, no se encontraron correlaciones relevantes que permitieran descartar variables bajo ese concepto.

²¹ Hace referencia a la función de autocorrelación del primer coeficiente para cada serie de tiempo.

²² Todas las series adquirieron el mismo valor, por lo tanto no representó información que aportara a los posteriores entrenamiento y clasificación.

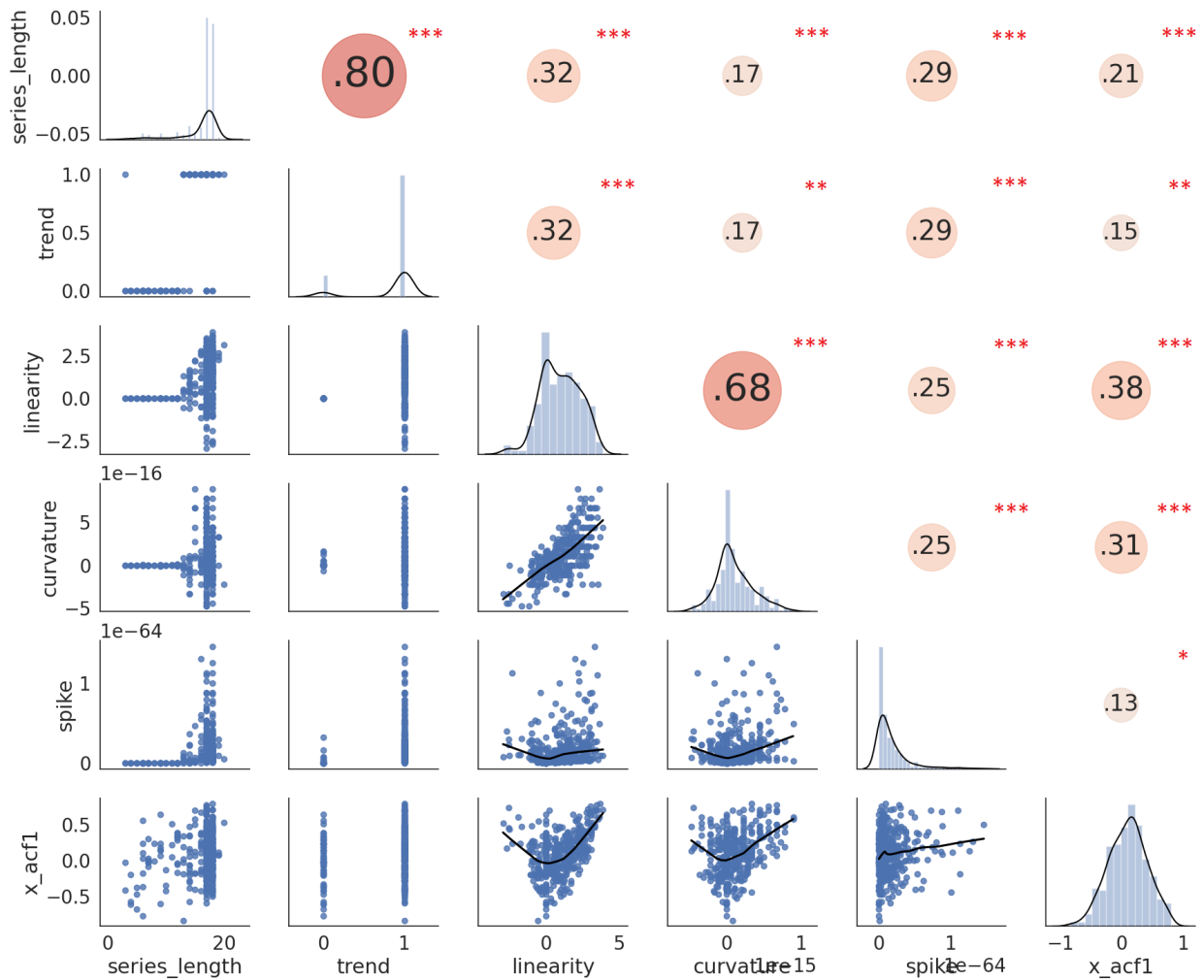


Figura 2.19: Correlograma detallado de variables con mayores coeficientes de correlación

2.3.2.1. *Resampling*: sobremuestreo

Una vez definidas las variables a utilizar en el clasificador, se procedió a generar un set aleatorio de entrenamiento y testeo con una proporción de 70% y 30%, respectivamente.

Con los datos de testeo seleccionados se procedió a buscar los parámetros óptimos para el ajuste del modelo *Random Forest*, usarlos para el entrenamiento del clasificador y, por último, testear la predicción de modelo a través de los datos seleccionados para ello, obteniendo los siguientes resultados.

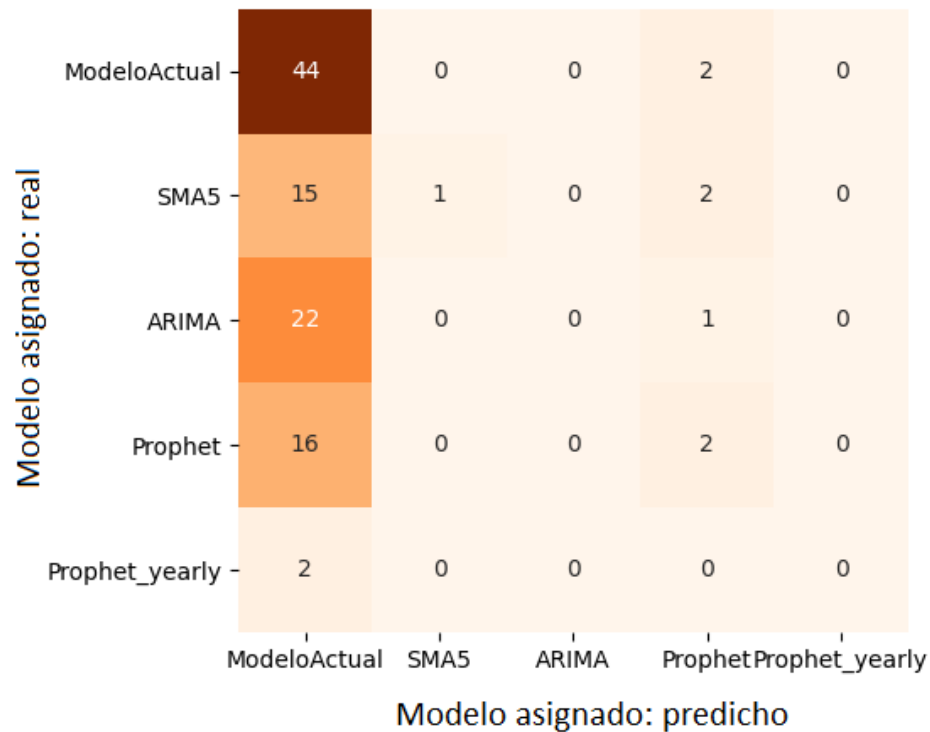


Figura 2.20: Matriz de confusión

La figura 2.20, además de evidenciar la clasificación errónea de 56 % (60) de las series de tiempo testada, hace evidente el desbalance de datos de testeo asociadas a clasificaciones minoritarias, generando las siguientes métricas.

| | Precision | Recall | F1 |
|-----------------|-----------|--------|------|
| Modelo Actual | 44 % | 96 % | 61 % |
| SMA5 | 100 % | 6 % | 11 % |
| ARIMA | - | 0 % | - |
| Prophet | 29 % | 11 % | 16 % |
| Prophet forzado | - | 0 % | - |
| Accuracy | 44 % | | |

Tabla 2.10: Resumen métricas de rendimiento del clasificador.

A partir de la tabla 2.10, se evidencian los bajos valores de métricas de rendimiento que genera el desbalance de clases. En particular, es incluso imposible calcular la precisión y F1 de las técnicas *ARIMA* y *Prophet con estacionalidad anual forzada* debido a la inexistencia de datos clasificados como tal por el modelo.

El desbalance inicial es evidente, considerando que la clase con mayor cantidad de datos (*Modelo Actual*) tiene casi el doble de datos que la segunda (77), y más de 14 veces el tamaño de la clase menor (*Prophet forzado*).

| | Cantidad |
|-----------------|----------|
| Modelo Actual | 142 |
| SMA5 | 56 |
| ARIMA | 77 |
| Prophet | 71 |
| Prophet forzado | 10 |

Tabla 2.11: Resumen distribución de clasificaciones originales.

Para abordar esta problemática, se selecciona la técnica de *resampling*, específicamente el sobremuestreo (explicado en la sección 1.5.2.1, logrando equilibrar los datos a fin de entrenar de mejor manera el clasificador y obteniendo, entre datos simulados y originales, 142 datos asociados a cada una de las técnicas y con un total de 710 (implicando 497 datos de prueba y 213 de testeo).

2.3.2.2. Resultados clasificador de modelos

Aplicada una solución al desbalance, se procedió nuevamente a buscar los parámetros óptimos de ajuste. Para esto se simuló el uso de más de 900 combinaciones para 4 parámetros (ver figura 2.21, obteniendo en aproximadamente 80 minutos, los valores que se utilizarían para entrenar el clasificador.

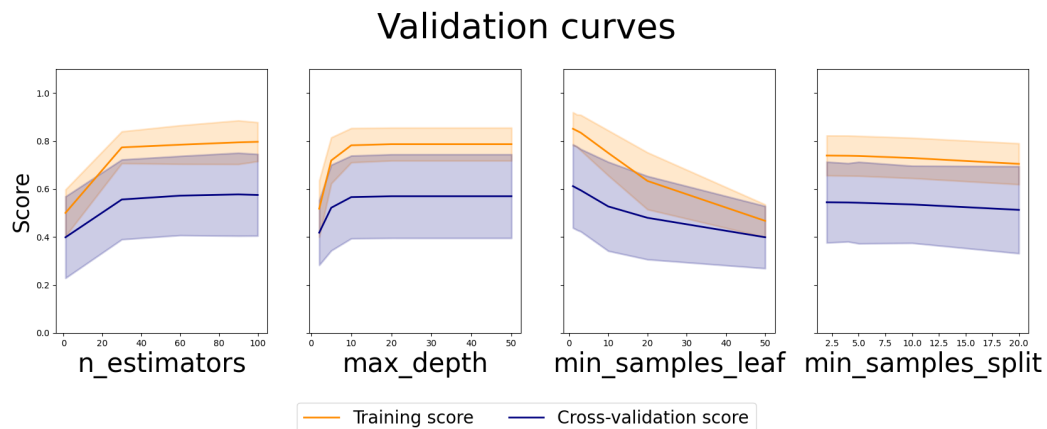


Figura 2.21: Curvas de validación de parámetros para clasificador *Random Forest*.

Finalmente, los parámetros óptimos a utilizar en el entrenamiento del modelo fueron: $n_estimators = 90$, $max_depth = 10$, $min_samples_leaf = 1$ and $min_samples_split = 2$. A partir de dicho entrenamiento, se obtiene los resultados de testeo siguientes.

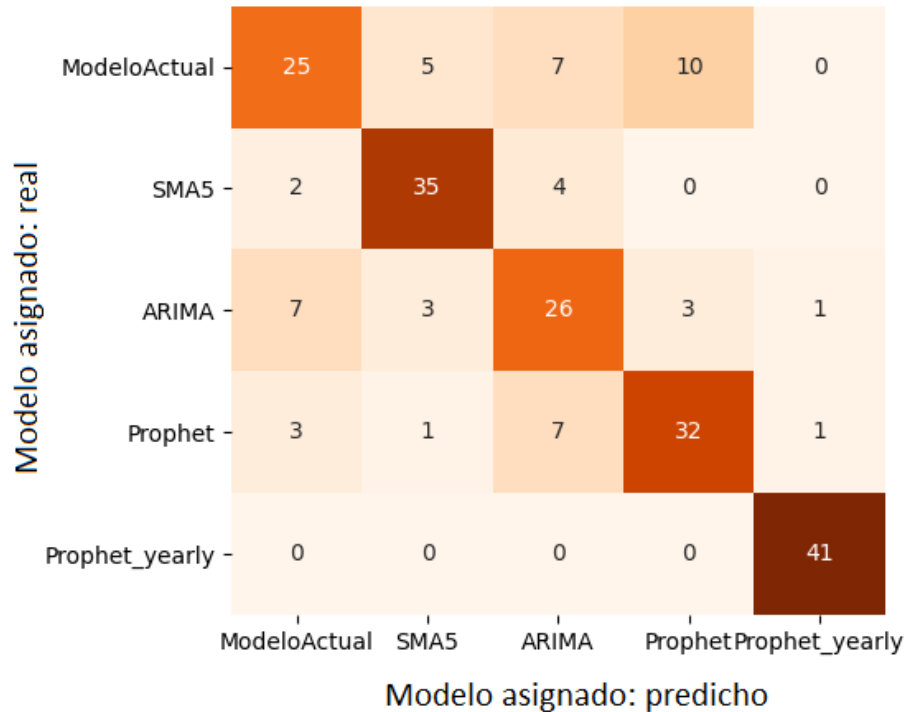


Figura 2.22: Matriz de confusión post entrenamiento con *oversampling*.

Aplicada la solución al set de datos utilizado, se obtuvieron nuevas métricas de rendimiento, las cuales en su mayoría mejoraron. La única excepción es la precisión y F1 de la técnica *Modelo Actual*. No obstante, esto se explica precisamente por el sobreajuste de esta clase generado por el primer clasificador en base al desbalance.

| | Precision | Recall | F1 |
|-----------------|-----------|--------|------|
| Modelo Actual | 63 % | 54 % | 58 % |
| SMA5 | 91 % | 76 % | 83 % |
| ARIMA | 62 % | 76 % | 68 % |
| Prophet | 85 % | 87 % | 86 % |
| Prophet forzado | 95 % | 100 % | 98 % |
| Accuracy | 79 % | | |

Tabla 2.12: Resumen métricas de rendimiento del clasificador post entrenamiento con *oversampling*.

De las 213 clasificaciones usadas para testear, 114 corresponden a datos reales los cuales, al aislarlos, siguen mostrando un buen comportamiento. Esto se observa en la figura 2.23, en primera instancia, y con más detalle en la tabla.

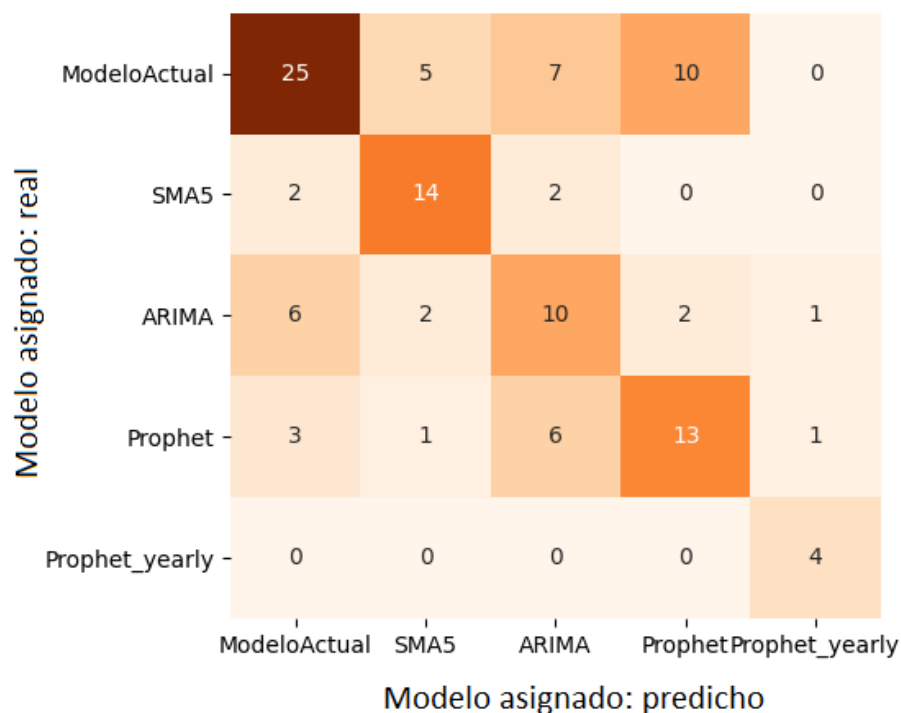


Figura 2.23: Matriz de confusión post entrenamiento con *oversampling*, solo datos reales.

| | Precision | Recall | F1 |
|-----------------|-----------|--------|------|
| Modelo Actual | 69 % | 53 % | 60 % |
| SMA5 | 64 % | 78 % | 70 % |
| ARIMA | 40 % | 48 % | 43 % |
| Prophet | 52 % | 54 % | 53 % |
| Prophet forzado | 67 % | 100 % | 80 % |
| Accuracy | 58 % | | |

Tabla 2.13: Resumen distribución de clasificaciones post entrenamiento con *oversampling*, solo datos reales.

Naturalmente, al quitar los datos artificiales generados por el sobremuestreo en las clases de menor tamaño, las métricas tomaron valores menores a las anteriores; aun así, siguen siendo mejores que sin la ayuda de dicha técnica.

En específico, series de tiempo como las asociadas a la técnica *Prophet con estacionalidad anual forzada*, se benefician de una buena clasificación debido a que sin esta técnica, los valores MAE y MAPE asociados a ella serían aún mayores.

Por ejemplo, al medicamento “*OMEPRAZOL - 20 MG - 60 - CÁPSULAS*” se había asignado anteriormente la técnica *Modelo Actual* con un MAE de 18,6 y MAPE de $\sim 96\%$, cuando con *Prophet con estacionalidad anual forzada* obtiene MAE de 14,8 y MAPE de $\sim 48\%$. Las diferencia son apreciables en la figura 2.24

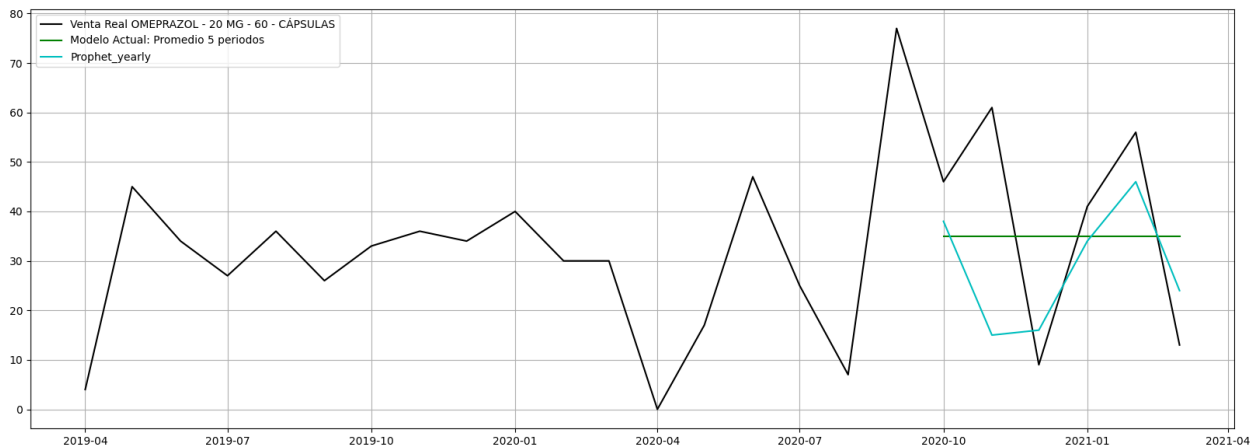


Figura 2.24: Ejemplo de medicamento asignado a *Prophet* con estacionalidad anual forzada.

A través de la tabla 2.14 se puede evidenciar la variación de rendimiento en MAPE en promedio para la predicción de series de tiempo con el modelo de clasificación utilizado, versus directamente asignar la mejor técnica posible para dicha serie nueva. El aplicar dicha serie (la mejor) implicaría la necesidad de entrenar y medir los resultados de esta con todas las técnicas disponibles y clasificarla a partir del rendimiento en alguna métrica en particular, así como se realizó en la sección 2.3.1 con el MAPE.

| | N° Fallas | Aumento MAPE promedio | MAPE promedio final |
|-----------------|--------------|--------------------------|------------------------|
| Modelo Actual | 22 | 8,2 % | 44,3 % |
| SMA5 | 4 | 8,3 % | 54,8 % |
| ARIMA | 11 | 14,8 % | 45,2 % |
| Prophet | 11 | 17,7 % | 38,4 % |
| Prophet forzado | 0 | 0 % | 39,5 % |
| General | 48 | 11,9 % | 44,7 % |

Tabla 2.14: Resumen de variación de rendimientos MAPE en base al uso del clasificador.

En esta implementación, se observa que 48 (42 %) series de tiempo no fueron correctamente clasificadas, costando en rendimiento la ganancia de 11,9 % MAPE en promedio para cada una de ellas.

De la misma manera, la métrica MAE asciende según grupo, es decir, en base al nivel de venta de cada uno. Específicamente se tienen los siguientes valores.

| | N° Fallas | Aumento MAE | MAE promedio final |
|---------|----------------------|------------------------|-------------------------------|
| Grupo 1 | 24 | 0,3 | 1,8 |
| Grupo 2 | 20 | 1,6 | 7,0 |
| Grupo 3 | 4 | 4,5 | 11,9 |
| General | 48 | 1,2 | 4,6 |

Tabla 2.15: Resumen de variación de rendimientos MAE en base al uso del clasificador.

Capítulo 3

Conclusiones

A partir del análisis de resultados de cada técnica implementada, fue posible conocer el rendimiento de cada una de ellas por separado, así como sus tiempos de ejecución. En particular se encontró que la técnica *Modelo actual* es la que mejor rendimiento tiene en base al MAPE, es decir, al error porcentual promedio que las series tienen respecto de la predicción. Pero por otro lado, se encuentra la técnica *SMA5*, la cual presenta un mejor rendimiento en base al indicador MAE. Esto representa un indicio de que esta técnica se comporta mejor con productos de venta promedio mayor en base a la técnica *Modelo actual*.

Este indicio fue comprobado a través de la clasificación de series de tiempo en grupos según su magnitud promedio de ventas lo cual, de hecho, llevó a la conclusión de que *SMA5* se comporta mejor con los grupos 2 y 3 (de mayor venta promedio) y *ARIMA* con las series del grupo 1 (las de menor venta promedio). Finalmente, la técnica *Modelo actual* queda desplazada en rendimiento al segundo lugar para cada grupo.

En base a lo anterior, es posible indicar que, en primera instancia, si fuera necesario elegir sólo una técnica para aplicar a la totalidad de las series de tiempo, entonces esta sería *Medias Móviles Simples de 5 periodos (SMA5)* pues, a pesar de no tener el mejor MAPE con todas las serie, sí implica un menor error promedio para los medicamentos de mayor venta y por lo tanto, menor impacto negativo en el negocio.

Ahora, al clasificar cada medicamento según el modelo al que mejor responde (en base a los resultados anteriores), se encuentra una evidente mejora promedio tanto en la métrica MAE como en MAPE y, por lo tanto, es conveniente que cada serie de tiempo sea analizada basados en dicha asignación a través de una etiqueta (generación de dicho campo en la base de datos). Aun cuando tal método resulta tomar más tiempo que el seleccionado antes (*SMA5*), se sugiere la inversión de 3 minutos extras para obtener resultados mejores a través de esta ejecución que incluye la utilización de todas las técnicas.

Con respecto al clasificador de series de tiempo, se esperaba conseguir un modelo que fuera capaz de asignar la técnica predictiva que mejor pudiera ajustar a la serie de tiempo sin la necesidad de probar cada una de ellas. Ante esto se obtuvo valores relativamente bajos en cuanto al rendimiento del modelo. En particular, se obtuvo a nivel general una exactitud de clasificación de 58%, es decir que clasificó correctamente ese porcentaje de series de tiempo entre las distintas alternativas disponibles aumentando en promedio un

11,9% del MAPE y 1,2 el MAE. Adicionalmente, si bien la ejecución de las predicciones tomó menos de 1 minuto, el entrenamiento del clasificador tomó en promedio 80 minutos por ocasión. Además, debe considerarse la necesidad de uso de la técnica de sobremuestreo para compensar el desbalance de tamaño entre clases (técnicas), lo cual podría generar confusiones al clasificador, sobreestimando el tamaño de las clases más pequeñas, induciendo a error y reflejando esto en los resultados.

Frente a lo anterior, es necesario cuestionar si se está conforme con el clasificador considerando la cantidad de tiempo invertido en comparación a ejecutar todos los modelos para una nueva serie y a partir de los resultados elegir la mejor técnica (así como se realizó en un principio para cada medicamento al inicio del modelamiento), teniendo mayor certeza del rendimiento de esta alternativa. Vale mencionar que la ejecución de todas las técnicas para cada nueva serie de tiempo, toma tan solo 1 minuto en promedio.

3.1. Recomendaciones de implementación

En base a los resultados obtenidos, se recomienda la utilización de los mejores modelos para cada serie de tiempo debido a que implica un mejor rendimiento general de las predicciones y un tiempo razonable de ejecución considerando las opciones probadas en este trabajo.

Considerando las necesidades de la administración técnica de la farmacia, los códigos asociados a la predicción deberán ser ejecutados una vez al mes, obteniendo de esta manera los valores de predicción para cada uno de los medicamentos de los siguientes 5 meses, además de la predicción del que se encuentre en curso y realizando las operaciones en base a esta información.

Se recomienda ejecutar la asignación de mejores modelos para la totalidad de los medicamentos al menos cada seis meses, de manera tal de mantener actualizadas las etiquetas de estos en la base de datos, asegurando así que la técnica asignada a cada serie es realmente la de mejor rendimiento en cada momento.

¿Por qué cada seis meses? Pues porque muchas de las series de tiempo que se tienen cuentan con una historia limitada (menos de dos años de datos), por lo cual, a medida que se tenga más información de ellos, será posible captar mejor el comportamiento de estas (aparición de estacionalidades, tendencias positivas o negativas, etc) y asignar una técnica que responda a las características de la serie de tiempo. Un periodo de seis meses en este caso, es una cantidad de tiempo prudente para incluir potenciales variaciones relevantes relativas a las series de tiempo que pudieran hacer cambiar el escenario y, por lo tanto, la técnica que mejor le ajusta.

Conforme a lo anterior, se sugiere la integración de los datos arrojados por las técnicas en la plataforma FarmaBot tan pronto se ejecuten los códigos, de manera que el usuario pueda tomar las decisiones de abastecimiento mes a mes. Por supuesto, dichas decisiones deben ir acompañadas de una revisión de inventario (el sistema FarmaBot ya cuenta con ella) que permita ajustar los niveles de compra en base al stock del momento en que se toman, es decir, si ya se cuenta con un stock alto (que sobrepase la demanda proyectada del mes actual y los venideros), el pedido no deberá sobredimensionarse más allá de lo recomendado por la

técnica usada.

Para dicha implementación, se sugiere mostrar los pedidos ya programados (caso CENA-BAST) de cada medicamento junto con los valores máximos de modificación ($\pm 30\%$) y con la predicción entregada por las técnicas para facilitar la toma de decisión de la encargada.

3.2. Recomendaciones de trabajo futuros

En base a lo observado en el estudio de esta memoria se recomiendan los siguientes trabajos futuros asociados al mejoramiento del actual sistema de administración FarmaBot:

- **Identificación de quiebre de stock:** implementación de campo que identifique en la base de datos la situación del stock de los medicamentos en base a su demanda proyectada. Por ejemplo, si se proyecta una demanda de 300 unidades de un producto para los siguientes 3 meses, sin embargo sólo se cuenta con 200 unidades, el sistema debiera lanzar una alerta de quiebre de stock para dos meses más. De esta manera, se facilita la tarea de quien realiza los pedidos de la farmacia pues la alerta indicaría el grado de urgencia que tiene cada uno de los productos para el reabastecimiento.
- **Registro de ventas perdidas o no realizada producto de quiebre de stock:** implementación de campo que identifique y caracterice en la base de datos, la razón de no venta de un medicamento cuando este sea demandado por un cliente. Esta acción permitiría trabajar con datos de demanda real (o cercana a la real) de los productos y no solo los datos de venta realizada.
- **Revisión y corrección de medicamentos:** si bien en el desarrollo de este trabajo se hicieron correcciones asociadas a los campos **dosis**, **cantidad** y **presentación**, es necesario generar una instancia de profundización en esta materia debido a que requiere la evaluación profesional farmacológica para no inducir a errores²³.
- **Identificación de medicamentos equivalentes:** se plantea la posibilidad de que medicamentos con mismo principio activo, dosificación y presentación sean equivalentes entre sí en alguna proporción y que por lo tanto sean reemplazables a la hora de ser comercializados²⁴. Esto apunta a mejorar la identificación real de la demanda por los medicamentos considerando que lo importante es el principio activo, la dosificación y la presentación y no la cantidad comercial en que se ofrece dicho medicamento.

Considerando lo anterior, se propone la continuidad del trabajo de esta memoria a través de las siguientes acciones:

- **Análisis de las series de tiempo en granularidad quincenal:** si bien, a medida que se trabaja con mayor detalle en la granularidad de las series de tiempo, los errores

²³ Por ejemplo, se identificó de manera separada el medicamento “Metformina” y “Metformina clorhidrato”, cuando farmacológicamente estas son equivalentes, por lo tanto en vez de ser consideradas como dos series de tiempo por separado, deberían expresarse en una. Este caso podría darse en distintos principios activos declarados, por lo cual se sugiere una revisión en este aspecto.

²⁴ Por ejemplo, "METFORMINA - 850 MG - 60 - COMPRIMIDO" podría ser reemplazado por la compra de dos cajas del mismo producto en formato de 30 comprimidos.

propagados son mayores al agregar los resultados (cálculo de demanda quincenal se agrega a demanda mensual), esto también implica una ventaja respecto del análisis con las distintas técnicas utilizadas. Esto debido a que se tiene mayor información utilizable por los algoritmos de las distintas técnicas (en particular, mayor cantidad de puntos para analizar) y por lo tanto se descartaría menor cantidad de series de tiempo por razones técnicas. Otro beneficio de esto es la mejor identificación de quiebres de stock en las series y por lo tanto, la menor subestimación de demanda.

- **Integración de hitos relevantes a la técnica de predicción *Prophet* y *Prophet* con anualidad forzada:** los hitos referidos son quiebres de stock, existencia de feriados y eventos que interrumpan el funcionamiento normal de la farmacias, entre otros que pudiesen considerarte relevantes para el caso.
- **Utilización de algoritmos de predicción con *machine learning*:** se sugiere el estudio de la demanda a través de esta opción debido a la capacidad de este método de integrar en un mismo modelo la interacción del comportamiento de los distintos productos, los cuales pueden ser claves a la hora de predecir el comportamiento de un medicamento en particular²⁵.

²⁵ Por ejemplo, se presume la existencia de medicamentos complementarios, es decir, se consume uno y probablemente o otro también (por ejemplo, con el uso de medicamentos antiinflamatorios acompañados de Famotidina u Omeprazol); o medicamentos equivalentes en acción, como lo pueden ser dos medicamentos que solo difieren en la presentación (cápsula vs. comprimido)

Bibliografía

- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>
- Cathalifaud, D. (2020, 10). *Reuniones virtuales con clientes*.
- CENABAST. (2017). *Quiénes somos*. <https://www.cenabast.cl/institucion/quienes-somos/>
- ChabotChile. (2020). *Crear agrupaciones a partir de una medida continua*. <https://www.farmabot.cl/about>
- Chávez Quisbert, N. (1997). Modelos arima. *Rev Cien Cult*(1), 23–30. http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S2077-33231997000100005&lng=es&tlng=es
- Logitec. (2014). Mermas del retail en chile. *Revista Logitec*. <https://www.revistalogitec.com/index.php/logistica/pymes/item/806-mermas-del-retail-en-chile>
- SanMiguel. (2018). *Informe diagnóstico comunal plan de desarrollo comunal de san miguel 2019-2024*. <https://web.sanmiguel.cl/farmacia-comunal/>
- SanMiguel. (2019). *Farmacia comunal*. <https://web.sanmiguel.cl/farmacia-comunal/>
- Tableau. (2020). *Crear agrupaciones a partir de una medida continua*. https://help.tableau.com/current/pro/desktop/es-es/calculations_bins.htm
- Taylor, S. J., y Letham, B. (2017). Forecasting at scale. *PeerJ Preprints* 5:e3190v2. <https://peerj.com/preprints/3190v2/>
- Zitko, P., y Aceituno, D. (2019). Proyección del estudio de carga de enfermedad. <https://salud-sociales.udla.cl/>. <https://salud-sociales.udla.cl/wp-content/uploads/sites/70/2020/08/Informe-Proyección-Carga-de-Enfermedad-a-2030-Chile-UDLA.pdf>

Anexo A

Resultados de la implementación de técnicas

A continuación se expone una muestra de los resultados de la ejecución de las cinco técnicas implementadas para la predicción de ventas de las series de tiempo de tres medicamentos, los cuales son comercializados por la Farmacia Comunal de San Miguel.

Estos son los datos utilizados para el cálculo de las métricas de rendimiento MAE y MAPE de cada técnica para cada producto.

| Producto | Fecha | Real | T1 | T2 | T3 | T4 | T5 |
|--|---------|------|-----|-----|-----|-----|------|
| LOSARTAN - 50MG - 30 - COMPRIMIDO | 10/2020 | 172 | 170 | 170 | 163 | 191 | -650 |
| LOSARTAN - 50MG - 30 - COMPRIMIDO | 11/2020 | 163 | 170 | 169 | 177 | 197 | 455 |
| LOSARTAN - 50MG - 30 - COMPRIMIDO | 12/2020 | 189 | 170 | 170 | 167 | 202 | -15 |
| LOSARTAN - 50MG - 30 - COMPRIMIDO | 01/2021 | 205 | 170 | 169 | 174 | 208 | 56 |
| LOSARTAN - 50MG - 30 - COMPRIMIDO | 02/2021 | 184 | 170 | 172 | 169 | 214 | 591 |
| LOSARTAN - 50MG - 30 - COMPRIMIDO | 03/2021 | 206 | 170 | 170 | 173 | 219 | 338 |
| METFORMINA - 850MG - 60 - COMPRI- MIDO | 10/2020 | 15 | 21 | 21 | 19 | 26 | 29 |
| METFORMINA - 850MG - 60 - COMPRI- MIDO | 11/2020 | 21 | 21 | 21 | 19 | 27 | 34 |
| METFORMINA - 850MG - 60 - COMPRI- MIDO | 12/2020 | 13 | 21 | 23 | 19 | 28 | 6 |
| METFORMINA - 850MG - 60 - COMPRI- MIDO | 01/2021 | 13 | 21 | 22 | 19 | 29 | 44 |
| METFORMINA - 850MG - 60 - COMPRI- MIDO | 02/2021 | 11 | 21 | 23 | 19 | 29 | 9 |
| METFORMINA - 850MG - 60 - COMPRI- MIDO | 03/2021 | 16 | 21 | 22 | 19 | 30 | 40 |
| QUETIAPINA - 50MG - 30 - COMPRIMIDO | 10/2020 | 4 | 2 | 2 | 2 | 2 | -26 |
| QUETIAPINA - 50MG - 30 - COMPRIMIDO | 11/2020 | 1 | 2 | 2 | 2 | 2 | -2 |
| QUETIAPINA - 50MG - 30 - COMPRIMIDO | 12/2020 | 3 | 2 | 2 | 2 | 2 | -4 |
| QUETIAPINA - 50MG - 30 - COMPRIMIDO | 01/2021 | 3 | 2 | 2 | 2 | 2 | -3 |
| QUETIAPINA - 50MG - 30 - COMPRIMIDO | 02/2021 | 3 | 2 | 2 | 2 | 2 | -4 |
| QUETIAPINA - 50MG - 30 - COMPRIMIDO | 03/2021 | 2 | 2 | 2 | 2 | 2 | -3 |

Tabla A.1: Ejemplos de resultados de la implementación de las técnicas de predicción. T1: Técnica Actual, T2: SMA5, T3: Arima, T4: Prophet, T5: Prophet_year.

Anexo B

Rendimientos de la implementación de técnicas

A continuación se expone una muestra de los resultados de rendimiento MAE y MAPE de la ejecución de las cinco técnicas implementadas para la predicción de ventas de las series de tiempo de tres medicamentos, los cuales son comercializados por la Farmacia Comunal de San Miguel.

Las técnicas para las cuales fue calculado el MAE y MAPE son: “Técnica actual”, “SMA5”, “ARIMA”, “*Prophet*” y “*Prophet_year*” (*Prophet* con anualidad forzada). Estos son los datos utilizados para la evaluación de rendimientos de los distintos análisis que se muestran en esta memoria. Además, se tomó este set de datos como parte de los *inputs* para la implementación y evaluación de la técnica de clasificación *Random Forest*.

| Producto | Técnica | MAE | MAPE |
|---------------------------------------|----------------|--------|------|
| LOSARTAN - 50 MG - 30 - COMPRIMIDO | Técnica Actual | 18,83 | 0,10 |
| LOSARTAN - 50 MG - 30 - COMPRIMIDO | SMA5 | 18,50 | 0,09 |
| LOSARTAN - 50 MG - 30 - COMPRIMIDO | ARIMA | 20,67 | 0,11 |
| LOSARTAN - 50 MG - 30 - COMPRIMIDO | Prophet | 18,67 | 0,10 |
| LOSARTAN - 50 MG - 30 - COMPRIMIDO | Prophet_year | 334,33 | 1,87 |
| METFORMINA - 850 MG - 60 - COMPRIMIDO | Técnica Actual | 6,17 | 0,48 |
| METFORMINA - 850 MG - 60 - COMPRIMIDO | SMA5 | 7,17 | 0,55 |
| METFORMINA - 850 MG - 60 - COMPRIMIDO | ARIMA | 4,83 | 0,37 |
| METFORMINA - 850 MG - 60 - COMPRIMIDO | Prophet | 13,33 | 0,99 |
| METFORMINA - 850 MG - 60 - COMPRIMIDO | Prophet_year | 15,17 | 1,03 |
| QUETIAPINA - 50 MG - 30 - COMPRIMIDO | Técnica Actual | 1,00 | 0,42 |
| QUETIAPINA - 50 MG - 30 - COMPRIMIDO | SMA5 | 1,00 | 0,42 |
| QUETIAPINA - 50 MG - 30 - COMPRIMIDO | ARIMA | 1,00 | 0,42 |
| QUETIAPINA - 50 MG - 30 - COMPRIMIDO | Prophet | 1,00 | 0,42 |
| QUETIAPINA - 50 MG - 30 - COMPRIMIDO | Prophet_year | 9,67 | 3,28 |

Tabla B.1: Ejemplos de métricas de rendimiento MAE y MAPE a partir de la implementación de las técnicas de predicción.