



UNIVERSIDAD
TORCUATO DI TELLA

MASTER IN MANAGEMENT + ANALYTICS

FORECASTING DE LENTES
OFTÁLMICAS BAJO UN ENFOQUE DE
MACHINE LEARNING

TESIS

Juan Francisco Fernandez Reguera

Mayo 2022

Tutor: Ignacio Mera

Resumen

La evolución de la capacidad de almacenamiento y de procesamiento informático con la que la sociedad cuenta hace algunas décadas ha permitido combinar el uso de modelos estadísticos y de diversos algoritmos para generar información que posee un valor incalculable para cualquier tipo de organización. En el presente trabajo se utilizarán técnicas de *Machine Learning* para predecir la demanda de un producto recientemente lanzado al mercado de lentes oftálmicas. La cantidad por predecir cuenta con una particularidad que es la alta cantidad de subproductos, es decir, lentes (SKU's) con la misma composición, pero con distinta graduación. Esto significa que, si bien la demanda en términos consolidados del producto puede considerarse relativamente estable, no ocurre lo mismo al observar de forma individual a cada uno de estos subproductos. La correcta predicción de la demanda podría traducirse en un ahorro significativo de recursos de una compañía, como lo pueden ser la reducción de su activo inmovilizado, menores costos asociados a cubrir quiebres de stock y ahorros considerables en el proceso de cadena de abastecimiento.

Abstract

The evolution of the storage and computer processing ability that society has had for the last few decades has made it possible to combine the use of statistical models and various algorithms to generate information that has incalculable value for any type of organization. In the present work, Machine Learning techniques will be used to predict the demand of a recently launched product on the ophthalmic lens market. The amount to be predicted has the particularity of a high number of sub-products, that is, lenses (SKU's) with the same composition, but with different graduation. This means that, although the demand in merged terms for the product can be considered stable, it does not occur the same when looking at each of these sub-products individually. The correct prediction of demand could be translated into significant savings in a company's resources, such as the reduction of its immobilized assets, lower costs associated with covering stock outages, and considerable savings in the supply chain process.

Índice

1. Introducción.....	5
1.1. Contexto y marco teórico.....	5
1.2. Abordaje de la problemática	7
1.3. Objetivo del presente trabajo.....	8
2. Obtención, evaluación y comprensión de los datos.....	9
2.1 Orígenes de los Datos.....	9
2.2 Análisis exploratorio de los datos	9
2.3 Tablas y gráficos representativos.....	11
3. Modelos y métricas de evaluación a utilizar	15
3.1 Confección de Dataset sintéticos.....	15
3.2 Ingeniería de atributos y elección de <i>features</i>	15
3.2.1 Cálculo de Leadtime.....	15
3.2.2. Política de stock mínimo.....	16
3.2.3 Descripción de variables del Dataset	18
4. Estrategias de validación de modelos.....	20
4.1. <i>Training</i> y <i>Testing</i> Sets	20
4.2 Método de validación seleccionado.....	22
5. Modelos investigados, descartados y utilizados.....	25
5.1 Modelos estudiados y descartados	25
5.2 Modelos seleccionados	26
5.2.a <i>Exponential smoothing state space model</i> (Suavizado Exponencial).....	26
5.2.b <i>Random Forest</i>	27
5.2.c <i>Generalized Boosted Regression Modeling</i>	29
5.2.d <i>XGBoost (Extreme Gradient Boosting)</i>	30
6. Métricas de Evaluación de Modelos	33
6.1 Métricas propias de los modelos utilizados.....	33
6.1.1 MSE	33
6.1.2 RMSE	34
6.1.3 MAE	34
6.1.4 sMAPE	34
6.2 - Métricas de evaluación del Negocio.....	34
6.2.1 Porcentajes del desvío sobre la cantidad pronosticada.....	34
6.2.2 Desvíos sobre la demanda en defecto.....	35
6.2.3 Quiebres de stock del pronóstico.....	35

6.2.4 Lenta rotación de SKU's. Rentabilidad del pronóstico.....	35
7. Ejecución y optimización de modelos elegidos	37
7.1 <i>Exponential smoothing state space model (ETS)</i>	38
Desarrollo del modelo.....	38
Optimización del modelo.....	38
Desempeño de la predicción general y unitario	38
Distribución de errores.....	39
7.2. <i>Random Forest</i>	40
Etapas iniciales del modelo	40
Optimización de Hiper-parámetros y principales variables.....	40
Desempeño de la predicción general y puntual del modelo	41
Distribución de errores	42
7.3. <i>Generalized Boosted Regression Modeling</i>	43
Etapas iniciales del modelo	43
Optimización de Hiper-parámetros	44
Desempeño de la predicción general y unitario	46
Distribución de errores.....	47
7.4 <i>Extreme Gradient Boosting (XGBoost)</i>	48
Desarrollo del modelo.....	48
Optimización de Hiper-parámetros	48
Desempeño de la predicción del modelo.....	49
Distribución de errores.....	50
8. Elección del modelo en base a los resultados obtenidos.	51
8.1 Performance de los modelos bajo métricas de <i>Machine Learning</i> :	51
8.2 Performance de los modelos bajo métricas del Negocio de lentes oftálmicas:	52
8.3 Distribución de los errores de los modelos presentados:.....	53
8.4 Elección de modelo bajo las diversas métricas.....	54
9. Conclusiones y recomendaciones de Negocio	55
9.1 Propuestas de mejoras al trabajo realizado	56
Referencias	58

1. Introducción

1.1. Contexto y marco teórico

Durante el mes de junio de 2021 una compañía de la industria del cuidado de la visión lanzó un nuevo producto al mercado bajo una reconocida marca. La propuesta de esta marca estaba orientada a un tipo de consumidor de bajo poder adquisitivo. Desde el punto de vista de la propuesta de valor, la lente tenía atributos muy interesantes en función de la relación entre tecnología del producto y el precio de lanzamiento, que era incluso menor que el de marcas genéricas que circulaban en ese entonces en el mercado.

A continuación, se realizará una breve introducción al negocio de las lentes oftálmicas para contextualizar la problemática:

Cuando el optometrista receta a un paciente lentes, el proceso comienza por la elección del armazón. Aquella elección es la que el consumidor suele priorizar y que contendrá las dos lentes que deberán tener determinadas características y una cantidad de atributos y combinaciones muy elevadas. Esas lentes recetadas pueden ser simétricas, ambos ojos poseen las mismas características, o pueden ser diferentes entre ellas. Las principales unidades utilizadas para recetar lentes medicinales tienen 3 coordenadas para determinar el tipo de lente que necesita un paciente: esférico (las graduaciones suelen ir de -12 a 12 en una escala de 0,25 en 0,25), cilíndrico (las graduaciones suelen ir de -10 a 0 en una escala de 0,25 en 0,25), y eje (0° a 180° de escala 1° en 1°). Como se puede observar, una receta del optometrista puede tener una gran cantidad de combinaciones y cada individuo también podrá optar por el material del que quiere en su anteojo, por ejemplo, si es una lente de sol, una blanca o una fotocromática. Existen otras elecciones del consumidor como son el tipo de protección del antirreflejo de forma tal que filtre determinados rayos de luz que dañan la visión como pueden ser la protección UV o la luz azul. Estas características que puede tener la lente que va a llevar un individuo es solo una pequeña porción de la cantidad de opciones a las que el consumidor puede acudir cuando va a realizarse un anteojo y eso hace que la demanda sea muy difícil de pronosticar y el negocio tenga un componente de servicio muy personalizado.

Para abastecer esta demanda la industria ha segmentado el negocio de lentes oftálmicas en dos grandes subdivisiones en función del grado de consumidores potenciales que tienen las lentes. El primer grupo está compuesto por las lentes de stock, que son aquellas que tienen mayor demanda y están también relacionadas generalmente a graduaciones menores, esféricos muy bajos de -2.00 a 2.00 y cilíndricos de hasta -2.00. En ellas se centra una gran proporción de las personas con problemas leves y moderados de la visión como puede ser el caso de la miopía o la hipermetropía. Las lentes de stock suelen ser producidas en grandes plantas de forma masiva y tienen un nivel de rotación muy elevado.

Con el objetivo de diferenciar las distintas graduaciones que puede tener el mismo producto, las lentes de stock se agrupan bajo un código denominado SKU, por sus siglas en inglés, *stock keeping unit* y son la forma en la que la compañía organiza y controla su stock. En líneas generales este código se asigna de acuerdo con el criterio interno de la empresa que por lo general se basa en la asignación de números a productos de forma secuencial o siguiendo algún tipo de estructura significativa. Para el caso de las lentes de stock, la identificación del SKU se basa en el código de barras que posee cada lente impresa en el sobre que la protege de rayaduras y polvo antes de ser colocada en un armazón. Si lo vemos desde una perspectiva logística, la cantidad de SKU's es muy

grande pero su alta rotación avala el hecho de producirlas a gran escala y almacenarlas en cantidades razonables.

Sin embargo, existe otra proporción de consumidores de lentes que tienen graduaciones combinadas, aumentos poco usuales o enfermedades menos asiduas que necesitan una lente de mucha mayor “personalización”, para las cuales no es conveniente tener en stock, debido a su baja rotación. En estos casos las grandes plantas producen una lente semielaborada denominada *block* al que laboratorios de pequeña escala tallan exclusivamente a demanda, otorgándole la graduación solicitada de forma exclusiva para ese consumidor. A estas lentes se las conoce en la jerga como lentes de laboratorio. Las lentes de laboratorio, por cuestiones de escalabilidad de costos, son entre 2 y 2,5 veces más costosas que una lente de stock. Si una lente de stock está en falta, se puede producir en un laboratorio, aunque no resulta tan rentable como comprarla terminada. Otra regla para tener en cuenta es que las lentes de stock y laboratorio no pueden ser combinadas en el mismo antejo, por lo que, si una receta posee una lente de stock y otra que debe ser tallada, la totalidad de la lente deberá ser producida en laboratorio.

Retornando al lanzamiento del producto, el mismo se trataba de una lente de stock, siendo uno de sus atributos más valorados por los minoristas su rango extendido de graduaciones; es decir, la compañía importaría de las plantas no solamente las graduaciones más demandadas sino un rango un poco mayor, y las comercializaría al mismo precio que la lente de stock de rango usual. Esto implicaba que una óptica minorista podría ofrecer un producto más barato a una mayor cantidad de personas, al no existir la necesidad de pasar por el laboratorio. Como contrapartida a esta mayor cantidad de SKU's en oferta, el riesgo del negocio radicaba en el potencial aumento de la tasa de lenta rotación para los SKU's pertenecientes al rango extendido. Por último, un *feature* o atributo adicional que poseía el producto lanzado era la tecnología de protección ante la luz led que irradian los dispositivos tecnológicos tan utilizados en la actualidad. Esta propuesta de tecnología, amplitud de rango y bajo costo, lanzada a través de una marca reconocida por la población, fueron los pilares de la estrategia de la gerencia de la empresa.

El lanzamiento de las lentes contaba con un espectro de aproximadamente 1024 SKU's, con una curva de distribución desigual de las cantidades en función de la experiencia previa de la demanda de lentes de stock. La estimación de las cantidades a comprar por SKU del stock inicial fue realizada por la gerencia de Operaciones. Desafortunadamente, el equipo de esta área no tenía experiencia previa en demanda de rangos extendidos y para junio de 2021 se realizó una compra de unas 60.000 lentes al proveedor que las producía en el exterior. Inicialmente la estrategia de surtido de stock se basaba en que el bajo costo de las lentes compensaría la baja rotación de los SKU con menor movimiento y de esa forma mantener el capital de trabajo en niveles aceptables.

En principio, no se conocen las razones, ya sea únicas o combinadas, de la excelente recepción por parte de la comunidad del cuidado de la visión. En este contexto, los stocks de muchos SKU's de los rangos extendidos se vieron desbordados por la demanda y obligaron a la compañía importadora a tallar localmente aquellos SKU's que, por ser un lanzamiento reciente, no podían resignarse a no abastecer. En menos de dos meses muchos SKU's sufrieron la ruptura del stock por lo que, comenzando el tercer mes de venta, se colocó una segunda orden de compra de reposición de stock por 62.000 lentes adicionales y así cumplir con una política de stock de 90 días de ventas. Es importante destacar que el costo FOB, es decir, el valor del producto en su país de origen antes de gastos de flete, seguro y aranceles aplicables se reducía en un 5% a partir de pedidos mayores a 30.000 unidades y eso motivó una nueva compra tan voluminosa. Cuando

arribó el refuerzo de lentes solicitado, otros SKU's volvieron a quebrar stock. Esto significa que para el quinto mes luego del lanzamiento, las 122 mil unidades adquiridas no contuvieron la demanda de varios SKU's del producto, a pesar de haberse vendido en los primeros 5 meses únicamente 48.000 lentes (40% de las compras realizadas al momento). Este desbalanceo de aprovisionamiento de stock encendió las alarmas sobre el error de pronóstico de demanda de SKU's ya que la producción local de lentes en falta comenzó a desgastar el margen bruto del producto.

Se ha estimado que la producción de lentes ocasionó costos adicionales a la compañía equivalentes al 8% de la rentabilidad lograda y que, al primer trimestre de ventas, el 5% de las lentes que se lanzaron al mercado habían sido talladas localmente. En este marco, la estimación de pronósticos, también llamado *forecasting de demanda*, cobra un papel protagónico para cada SKU si se desea mejorar la *performance* del producto en el mercado, sin disminuir sus niveles de rentabilidad.

Para abordar la introducción al marco teórico del forecasting se debe partir desde la definición de dicho concepto. Rob J Hyndman define que el forecasting “se trata de predecir el futuro con la mayor precisión posible, dada toda la información disponible, incluidos los datos históricos y el conocimiento de cualquier evento futuro que pueda afectar los pronósticos”ⁱ.

Para una correcta implementación del pronóstico de una compañía existen dos condiciones preliminares que enuncia Hyndman: objetivos y planificación.

Si se aplica esta definición al problema anteriormente mencionado se puede converger a un primer objetivo para esta problemática: Buscar predecir la demanda futura de cada SKU del producto lanzado utilizando información del mercado, características de los clientes y consumidores, las circunstancias del entorno económico y social, sumado a la mejor estimación posible que se pueda hacer de las condiciones futuras de aquella información que se esté utilizando.

Si bien las compañías deben realizar pronósticos de corto, mediano y largo plazo, el presente trabajo estará abocado a la planificación de los stocks de corto plazo. De esta forma se buscará pronosticar órdenes de compra para abastecer la demanda inmediata en un lapso determinado.

1.2. Abordaje de la problemática

Cuando nos referimos al *forecast* de demanda en un contexto de análisis de datos, la problemática suele asociarse a una estrategia de regresión, que puede o no ser lineal, y dentro de ella al estudio de series de tiempo. En la actualidad, se han desarrollado diferentes modelos de series de tiempo, por ejemplo, por Holt-Winters, ARIMA, SARIMAX, etc. Adicionalmente, para todos estos se pueden encontrar también diferentes enfoques para el manejo de series de tiempo en la literatura, tanto en modelos de regresión como en modelos aditivos. Sin embargo, en el trabajo de Pavlyshenkoⁱⁱ, se enumera la existencia de algunas limitaciones en los enfoques de series de tiempo para abordar los pronósticos de ventas a saber:

- La necesidad de tener datos históricos durante un largo período de tiempo para capturar patrones como la estacionalidad. Esta característica a menudo es difícil de obtener, por ejemplo, en el caso del lanzamiento de un nuevo producto. Como

respuesta a ello se utiliza para estos casos las series de tiempo en ventas de un producto similar donde se espera que nuestro nuevo producto tenga un patrón de desempeño similar.

- Los datos capturados pueden tener muchos datos atípicos y faltantes. La limpieza del dataset en estos casos se vuelve fundamental antes de utilizar un enfoque de series de tiempo.
- Será necesario tener en cuenta muchos factores exógenos que tienen un impacto en las ventas. En este sentido, el *forecasting* de la demanda de un producto pasa a asemejarse más a un problema de regresión que a un problema de series de tiempo. La práctica demuestra que el uso de enfoques de regresión a menudo puede brindar mejores resultados en comparación con los métodos de series de tiempo.

En el abordaje de nuestra problemática están presentes varias de las limitaciones anteriormente mencionadas. En primer lugar, el producto en cuestión fue recientemente lanzado al mercado y existe poca experiencia en la demanda del rango extendido de graduaciones. Respecto al segundo factor y, como se profundizará más adelante, los datos atípicos también han sido relevados y tratados en el presente trabajo. Por último, los factores exógenos pueden influir enormemente en la demanda de este tipo de productos, ya que la vida útil puede ser extendida casi indefinidamente si el contexto económico no acompaña al usuario. Esto último será el disparador de la incorporación de varios *features* de nuestro trabajo.

1.3. Objetivo del presente trabajo

Partiendo de las ideas anunciadas durante el transcurso de esta primera sección y, a modo de conclusión, se puede afirmar que las siguientes páginas estarán abocadas a encontrar un método de predicción de la demanda tal que el stock de la compañía pueda responder de forma acorde. De esta forma, se evitará el tallado en laboratorio a un mayor coste. Por otro lado, el análisis sopesará en forma negativa contar con volúmenes de stock ociosos muy altos. Para ello se evaluarán las condiciones de mercado sobre las que actúa la demanda y cómo ellos podrán influir en la predicción de la demanda.

Un tema no menor para considerar es el control de tiempos sobre el cual, si la compañía coloca una orden de compra en función a su pronóstico debe estimar también que tal orden se encuentre disponible al momento de inicio del período estimado. En este sentido será necesario incorporar al análisis el concepto de *leadtime* de los productos comercializados por la compañía. El *leadtime* se define como aquel tiempo que transcurre entre la colocación de la solicitud de compra al proveedor hasta que la misma se encuentra disponible para su comercialización. De esto se desprende que para realizar un pronóstico que sea eficiente se deberá obtener también una estimación lo más precisa posible de estos tiempo ya que en caso contrario se deberá recurrir al tallado de la demanda en exceso que es aquello que se desea evitar.

Para cumplir con la meta de este trabajo se utilizarán tanto modelos de series temporales convencionales como modelos de ensamble de tipo aditivo.

A su vez, un objetivo secundario será el de obtener una recomendación para el negocio que se base en un modelo de optimización de stocks mínimos de cada lente. Esto permitiría determinar la frecuencia y la cantidad de las reposiciones al momento de colocar una nueva orden de compra en la planta, con un costo de transporte eficiente.

2. Obtención, evaluación y comprensión de los datos

2.1 Orígenes de los Datos

Las bases de datos con las que se ha trabajado fueron extraídas en su mayoría de los registros de las compañías del grupo económico como también indicadores de actividad y económicos publicados por organismos públicos y universidades. Entre los más utilizados se encuentran los datos Solicitudes de Pedidos y de Stock de la compañía importadora y de sus subsidiarias (revendedoras minoristas a ópticas). También se utilizó la base de datos de la producción de las mercaderías faltantes.

Las bases de datos de pedidos contienen como principales inputs: el SKU, la descripción de la lente, cantidades demandadas/compradas, precios de lista, precios netos, graduaciones (esférico y cilíndrico), clientes, tipos de clientes y proveedores. Adicionalmente, a partir de estas bases de datos se generaron nuevas variables que buscan ayudar a los algoritmos utilizados a confeccionar un modelo más preciso de predicción de demanda. Por otro lado, se incluyeron variables exógenas relacionadas al marco macroeconómico y de la industria, como lo son el tipo de cambio, inflación, confianza de los consumidores y días de facturación para cada mes del período. A continuación, se enumeran de forma completa las bases de datos utilizadas:

- Movimientos de stock Compañía importadora.
- Listas de precios Compañía importadora y minoristas.
- Movimientos de stock Subsidiarias minoristas.
- Producción Laboratorio.
- Costos de Importación.
- Equivalencias de codificación de SKU entre compañías.
- Índice de Inflación INDEC.ⁱⁱⁱ
- Cotización USD divisa Banco Nación Argentina.^{iv}
- Índice de Confianza al Consumidor UTDT.^v
- Relevamiento de expectativas de Mercado (REM) del BCRA para perspectivas del USD e Inflación a futuro.^{vi}
- Días de facturación para los años en análisis.

2.2 Análisis exploratorio de los datos

Se ha destinado una cantidad considerable de trabajo al análisis de los datos de la demanda, antes de comenzar con la modelización de la problemática. En este sentido, se ha realizado la búsqueda de patrones y efectos temporales con el objetivo de confeccionar un dataset que incluya las variables principales de la demanda alineadas con la experiencia del mercado que posee la compañía. En función de ello se ha trabajado con los *features* de mayor significancia, con el objetivo de encontrar relaciones propias de los datos como pueden ser:

Tendencia: Se conoce como tendencia a aquella señal estadística que se desprende del análisis de los datos en un mismo sentido y que, generalmente suele mantenerse durante un periodo de tiempo. Un ejemplo de tendencia en el dataset analizado sería la tendencia a la mayor demanda de aquellas lentes con un esférico entre de poca graduación. Ver gráfico del punto 2.3 *Histograma de distribución de la demanda: Esféricos*.

Autocorrelación: los datos cercanos en el tiempo se parecen más entre sí con respecto a elementos más lejanos, solamente por el hecho de estar cerca.

Ruido: refiere a la variación inexplicable o al azar que se encuentra dentro de una muestra de datos determinada. De forma resumida, el ruido puede deberse principalmente a dos factores: errores y residuos. La eliminación del ruido dentro de un modelo estadístico permitirá mejorar la calidad predictiva de la estimación.

A su vez, también se buscará durante este análisis la búsqueda de relaciones causadas por el impacto de factores externos a la demanda y a la compañía. Algunos ejemplos de las relaciones que se investigarán serán: la relación de la demanda con el nivel general de precios (inflación), el impacto de los días laborales dentro de un período, la evolución de la demanda frente a la evolución de la cotización del dólar versus el peso argentino.

En los datos de ventas, también podemos observar valores extremos: valores atípicos que no representan la realidad de las variables y por ende se deberá establecer un plan de acción para estos datos. En caso de que se detecte alguno de estos eventos específicos y sean repetidos de forma periódica, se procederá a agregar esta característica de forma binaria que indicará estos eventos especiales y permitirá modelizar de mejor forma los valores extremos de la variable objetivo.

Uno de los casos más significativos que ha sido necesario excluir del análisis de proyección de la demanda es el de los stocks iniciales de revendedores de las lentes bajo análisis. Esto significa que un laboratorio de lentes oftálmicas decide gestionar su propio inventario para tener mayor agilidad en la prestación de sus servicios a bocas de expendio al consumidor. Si bien esta es una buena noticia para la compañía importadora por el alto nivel de facturación que implica, la confección de stocks iniciales es un factor poco predecible y que impacta de una forma tangencial al stock que se encuentre en cantidades de demanda elevadas. Durante el proceso de exploración de los datos se pudo detectar que la mayoría de los establecimientos de stocks iniciales han adquirido el mismo junto con el lanzamiento del producto, es decir durante los meses de junio y julio de 2021. Si bien han aparecido jugadores menores en los meses posteriores, se ha procedido a omitir la venta bajo estas condiciones para el análisis predictivo del trabajo.

Otro dato relevante detectado durante esta etapa fue el de un incremento de precios de lista adicional para un subgrupo de SKU's en el inicio del año 2022. Los precios de lista para los rangos extendidos han aumentado un 49% (debido a un ajuste para alinear el costo de reposición de dichos SKU al de las fábricas) contra el 10% de los rangos convencionales, lo que podría significar un punto de inflexión para la evaluación de la performance del modelo. Este evento fue incorporado al modelo mediante la creación de una nueva variable. En el gráfico que se expone a continuación, se puede observar si han existido variaciones significativas de la demanda a partir del cambio de este precio.

Año	Rango Regular	Rango Extendido
2021	73%	27%
2022	70%	30%

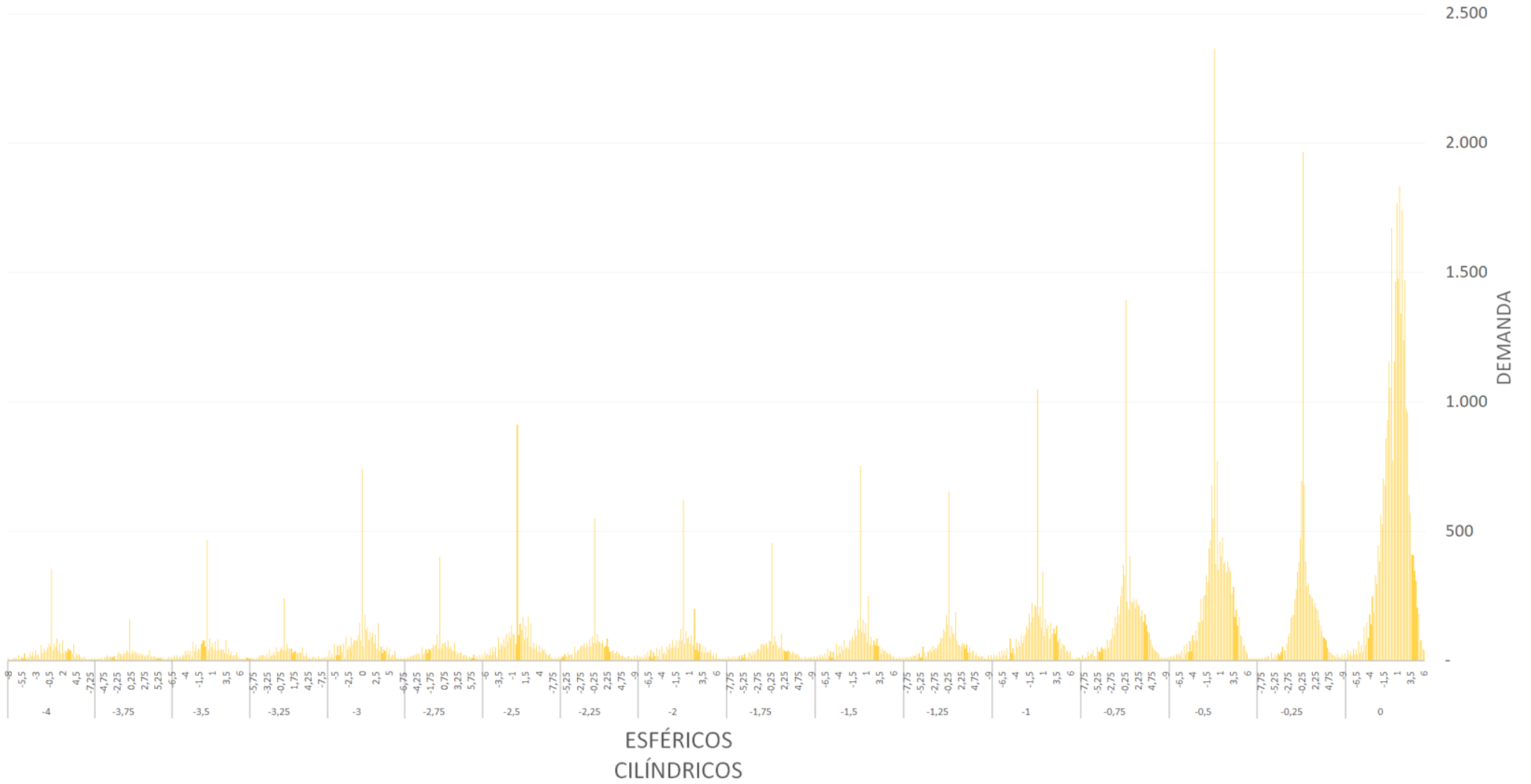
2.2. a Demanda de lentes según si son rangos extendidos o rangos regulares para cada año. Se puede observar que la variación de demanda de rangos extendidos no supone una diferencia significativa a pesar del aumento de precio.

2.3 Tablas y gráficos representativos

A continuación, se incluyen gráficos y tablas que han ayudado a determinar si las variables analizadas serán significativas o no para el modelo a desarrollar.

Otra de las características que consideramos que priorizará nuestro modelo a la hora de realizar una regresión de la cantidad demandada será el relacionado al esférico o al cilíndrico de la lente. Tal como se ha mencionado anteriormente, una lente de stock del mismo producto posee dos características principales que lo diferencian y que serán consecuencia de la receta de un oftalmólogo u optometrista, ellas son el esférico y el cilíndrico. En el gráfico 2.3.a, puede observarse como la demanda de esféricos disminuye a medida nos alejamos del 0 o neutro. Por otro lado, el cilíndrico neutro muestra la centralidad de la demanda, aunque este se va reduciendo a medida decrece.

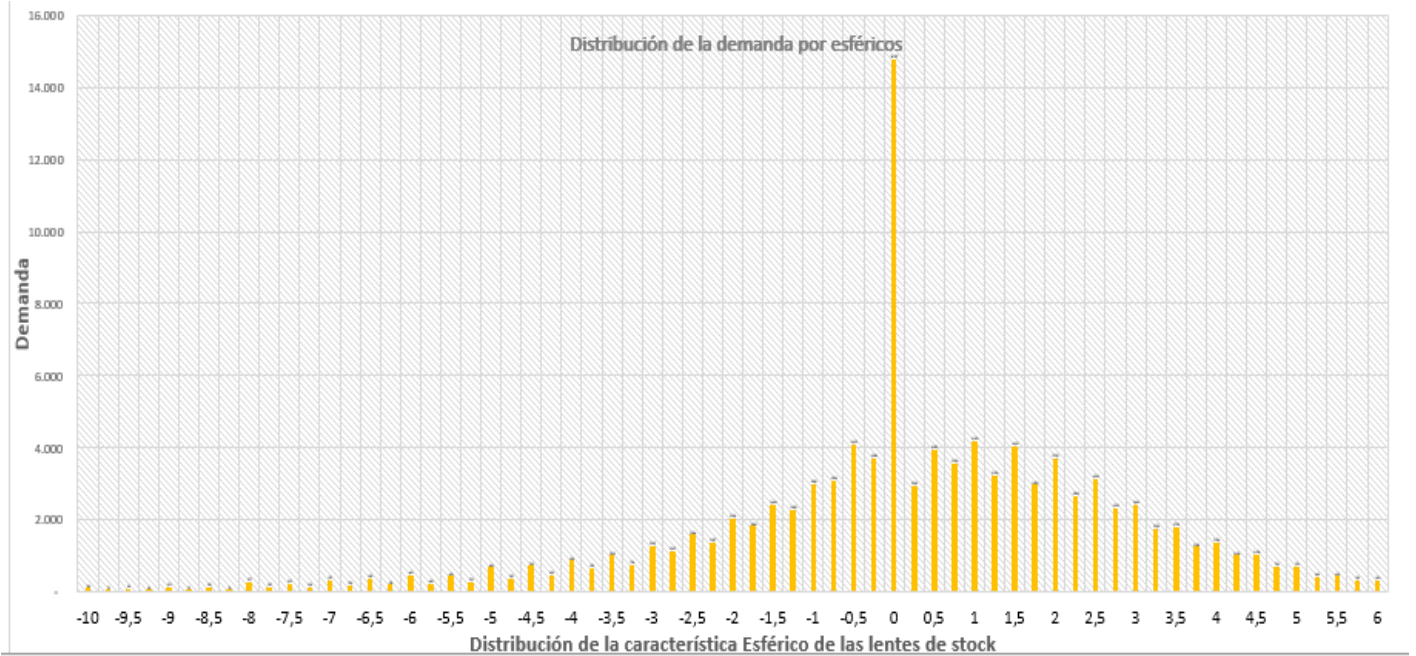
Histogramas de distribución de Demanda de Esféricos Agrupados por Cilíndricos



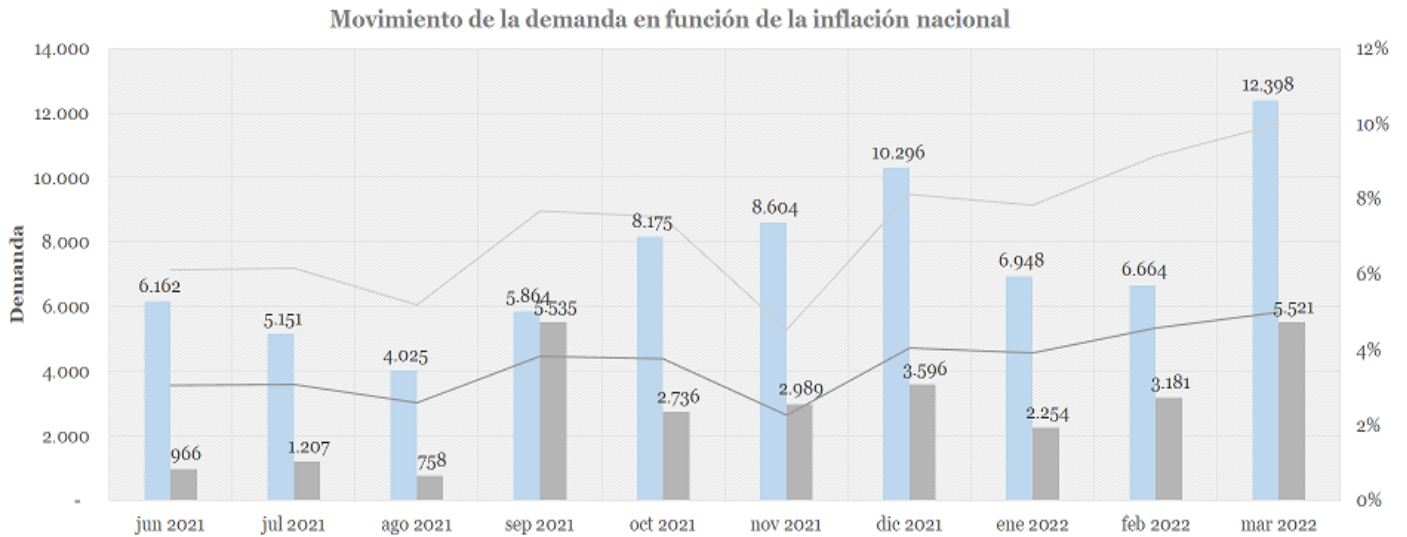
2.3.a Histogramas de distribución de la demanda de SKU agrupados por cilíndricos y subdivididos por esféricos. Se puede observar como al alejarse de la intersección de los ceros disminuye la cantidad demandada.

Histograma de distribución de cantidades demandadas de cilíndricos con esféricos anidados.

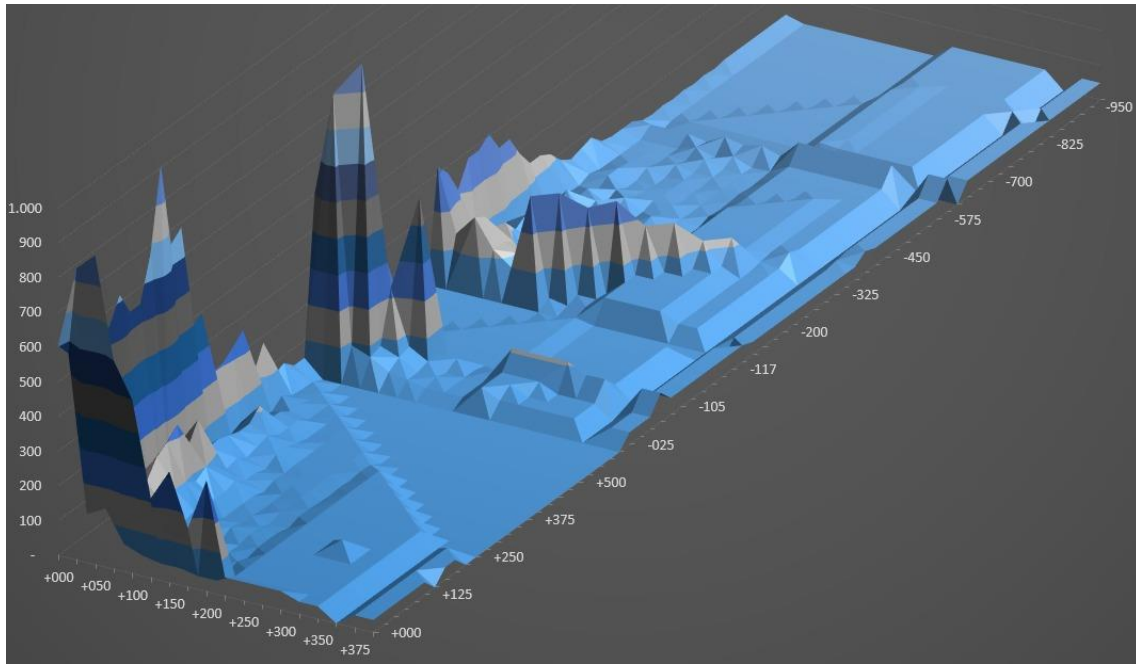
En el gráfico a continuación se puede observar cómo los esféricos incrementan su porción de la demanda a medida se acercan al 0, o neutro. Esta característica se presume será muy importante para que los modelos que sean llevados a cabo logren interpretar esta relación como un patrón a considerar para realizar un pronóstico correcto.



2.3.b Histograma de distribución de la demanda: Esféricos.

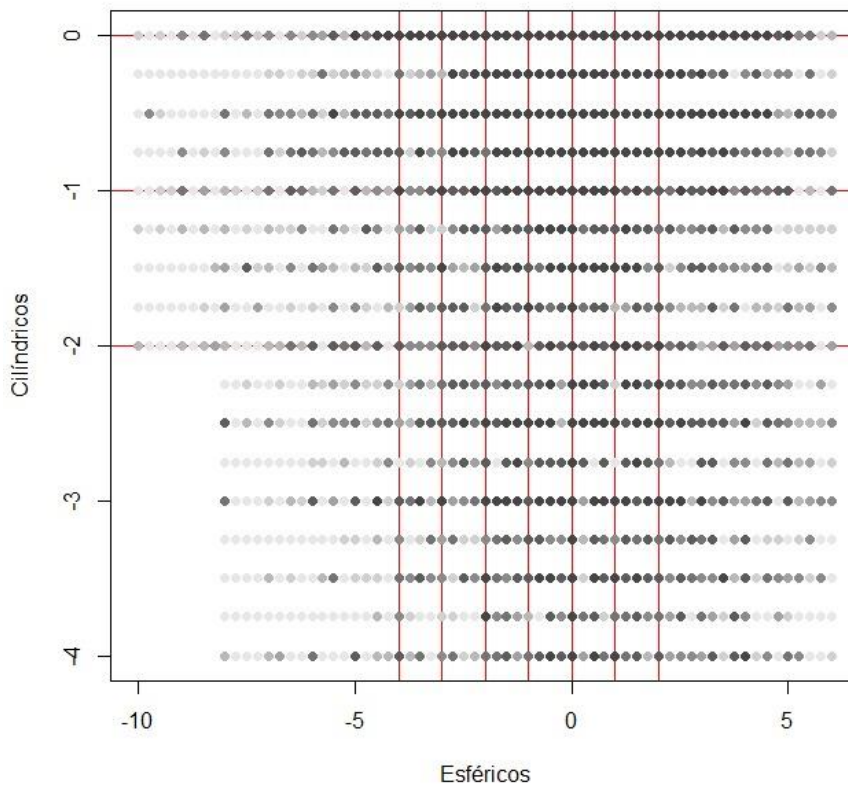


2.3.c Gráfico de evolución de la demanda (Rangos convencionales y Extendidos) comparados con el aumento del índice IPC/REM.



2.3.d Gráfico de distribución de stocks iniciales. Se puede observar cómo se ha concentrado la demanda inicial en aquellos stocks de bajos cilíndricos y esféricos, relegando de forma muy pronunciada al resto de los SKU. Este efecto conjugado con el armado de stocks iniciales sugiere que ha sido la causa de un quiebre sistemático de la producción de forma local de faltantes.

Distribución de la Demanda de Lentes por Graduación



2.3.e Distribución de la demanda de lentes en escala de grises, la intersección de líneas rojas define a los rangos regulares de lentes de stock de las de rango extendido. Aquí se puede observar la concentración de la demanda en las líneas verticales, aunque con mucha mayor oscuridad en los cilíndricos que componen el rango regular del stock.

3. Modelos y métricas de evaluación a utilizar

3.1 Confección de Dataset sintéticos

La confección del dataset ha sido un factor elemental durante el diseño del modelo. El primer factor que ha sido crucial en esta tarea fue la unificación de las demandas mayoristas y minoristas de la compañía que poseen dos sistemas de *web order* diferentes. A su vez, la codificación de los SKU para cada *web order* era distinta, por lo que el primer desafío para la construcción del dataset ha sido la creación de una base de datos de equivalencias de productos y, de esta forma, unificar las cantidades demandadas mensuales. Adicionalmente, mediante técnicas de extracción de datos se procedió a la generación de *features* clave como esférico y cilíndrico que estaban incorporados a la descripción de cada SKU. Una vez elaborada esta diferenciación se confeccionó una variable binaria de pertenencia al rango extendido del stock (72% de los SKU pertenecen al Rango Extendido). Estas características son las que en un principio se establecieron como iniciales y comunes a todos los meses de demanda.

Otro factor para tener en cuenta y para realizar una correcta medición de la demanda es la de la omisión de los stocks iniciales de los clientes revendedores del producto. Básicamente, y como se ha enunciado en la fase de exploración de datos, la demanda de stock iniciales es considerados un dato atípico ya que son eventos únicos donde un cliente establece un stock propio de lentes (oscila entre 600 y 3.000 lentes y la cantidad de SKU's no es la misma).

Por último, se procedió a generar, para el mes y año de la cantidad demandada, regresores o features adicionales que se consideraron de interés. La inflación, el aumento porcentual de la lista de precios, la variación del dólar mayorista, indicadores como la confianza del consumidor (índice ICC UTDT), etc.

3.2 Ingeniería de atributos y elección de *features*

Para realizar la ingeniería de atributos se realizó un análisis adicional al de la exploración de los datos. El objetivo principal de esto era poder contar con información real al momento de la colocación de la orden de compra al proveedor en el extranjero y que sea replicable para cada período futuro. También se tuvieron en cuenta los estándares que poseía la compañía en cuanto a parámetros claves como los días de stock o el *leadtime*. En función de estos dos parámetros se obtuvieron dos disparadores elementales para la confección del dataset final.

3.2.1 Cálculo de Leadtime

El primero de ellos ha sido la estimación de días de *leadtime* con el objetivo de poder establecer con qué información real se contaría al momento de la colocación de la orden de compra. Mediante el análisis de los despachos de importación del año 2021 se arribó a un *leadtime* promedio de 69 días y si se excluye los envíos rápidos por *courrier* se eleva a 76 días. Como se puede ver en la tabla 3.2.2.a que se expone a continuación, la fábrica recibe el pedido, se realiza en la misma el proceso de producción/*picking* y se despacha a destino mediante transporte aéreo. Un dato para destacar sobre la colocación de pedidos de lentes de stock es que la misma no posee una cantidad mínima de SKU para ordenar, tampoco máxima, aunque si el pedido total supera las 30.000 unidades existe una bonificación del costo FOB de 5%. En el gráfico 3.2.2.b se puede observar el efecto que produce la reducción del costo de importación y su relación con el aumento de la cantidad solicitada a las plantas. En función de la información recopilada se determinó que el tiempo prudencial para establecer un horizonte de proyección de ventas era de 90

días. Ejemplificando con base al modelo desarrollado, se han generado modelos con un dataset que contiene la información hasta noviembre de 2021. Una vez entrenados estos modelos se realiza el forecasting de la demanda del período de 90 días posteriores, es decir, marzo de 2022.

3.2.2. Política de stock mínimo

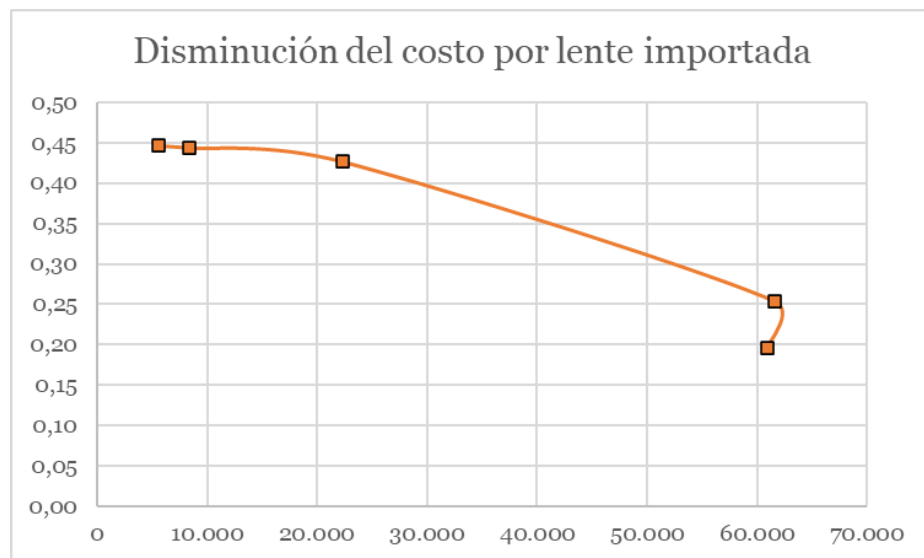
El segundo dato obtenido del relevamiento del proceso formal de la compañía es el requisito de los días de stock mínimos. Esta regla implica que el departamento de logística debe surtir con una cantidad de un SKU equivalente al menos a los 90 días de demanda que éste haya tenido. La conclusión arribada fue que utilizar la demanda de los últimos 90 días para realizar un *forecast* de los 90 días posteriores podía tener oculto algún patrón no tan lineal por lo que se utilizaron diversas variables acumulativas e individuales para cada SKU.

Tabla cálculo Leadtime

Orden de Compra	Fecha Orden de Compra	FOB (US\$)	Fecha Invoice	Fecha Recepción	Fecha Despacho	Lead time	Importe FC	Unidades	Flete	Seguro	Tipo Orden
2049	14-02-22	479	16-02-22	31-03-22	14-03-22	28,00	479	380	305	8	Courrier - Entrega rápida
1729	11-01-22	17.217			31-03-22	79,00	17.217	16.811			Pedido estandar *
1349	29-11-21	22.887	19-01-22	17-02-22	09-02-22	72,00	22.887	22.377	9.546	324	Pedido estandar
1090	28-10-21	8.620	07-12-21	28-12-21	22-12-21	55,00	8.620	8.326	3.696	123	Pedido estandar
845	23-09-21	6.102	10-11-21	14-12-21	30-11-21	68,00	6.102	5.640	2.518	86	Pedido estandar
556	24-06-21	60.427	21-06-21	10-09-21	31-08-21	68,00	60.427	61.586	15.653	761	Pedido estandar
423	26-01-21	53.226	27-04-21	28-05-21	20-05-21	114,00	53.226	60.899	11.893	651	Pedido estandar
						69,14	168.957	176.019	43.611	1.954	

Tabla 3.2.2.a Relevamiento y cálculo del leadtime a través de la obtención de la información entregada por la compañía.

* Corresponde a una importación arribada durante los primeros días de abril, al momento del relevamiento no se contaba con la información final del embarque.



3.2.2.b Gráfico de costos de importación en función de las unidades importadas.

En el siguiente gráfico se puede observar cómo existe una disminución significativa de los costos de importación a medida se coloca una orden superior a 20.000 unidades. Posteriormente, al obtener un 5% de descuento en el precio FOB a partir de las 30.000 unidades puede verse como este costo unitario de importaciones acelera su baja. El objetivo de la línea naranja busca reflejar de forma tentativa la tendencia a la baja que poseen los costos de importación, no obstante, no puede considerarse que la línea de tendencia sea concluyente dado que existe una fluctuación que la compañía no controla sujeta a los costos de flete, seguro y aranceles.

3.2.3 Descripción de variables del Dataset

A continuación, se describen las variables que conforman el dataset creado a partir de las distintas bases de datos obtenidas.

Target/Y/Forecast: Cantidad demandada para el SKU en el mes y año determinado. Variable objetivo de los modelos.

Esf: Característica esférico del SKU oscila entre -10.00 y +6.00, en una escala de 0.25 dioptrías entre cada SKU.

Cil: Característica cilíndrico del SKU oscila entre -4.00 y 0.00, en una escala de 0.25 dioptrías entre cada SKU.

Ano: identifica período (año) en análisis.

Mes: identifica período (mes) en análisis.

Las variables presentadas a continuación han sido creadas por el usuario con el objetivo de generar información adicional que permita a los modelos realizar mejores pronósticos:

RE_binaria: indica la pertenencia o no al rango extendido de la oferta de SKU que realiza la compañía. La misma ha sido incorporada con el objetivo de entender si la característica del SKU de pertenecer al rango extendido del producto incide en la demanda de este.

USD: Tipo de cambio mayorista que utilizan como base de cálculo las empresas que realizan importaciones. Tanto la compañía como sus clientes toman gran parte de las decisiones de abastecimiento en función de la cotización de la divisa o de la proyección que tendrá la misma.

Inflación_del_mes: coeficiente que surge de la división del indicador de mes sobre el del mes anterior. Se busca con este *feature* identificar si la demanda mayorista o minorista sufre alteraciones con el incremento o disminución del indicador general a nivel país.

Coef_lista_precios_100_06.2021: Armado de una serie base 100 con la lista de precios al 30.06.2021, fin de mes del lanzamiento inicial de producto.

ICC: Índice de Confianza al Consumidor publicado por la Universidad Torcuato Di Tella. El uso de este indicador busca establecer una relación, ya sea positiva o negativa, entre la confianza del consumidor orientada al consumo minorista y bocas de expendio de lentes y la demanda que puede tener la compañía para el mismo período.

Invoicing_days: Días hábiles o de facturación, es una medida muy utilizada en el consumo masivo y el *retail* donde los días de facturación implican un mayor flujo de ventas que meses cortos o con una cantidad mayor de días no laborables.

Venta_cuatro: Demanda del SKU para el último mes con el que se cuenta con información antes de colocar una orden que llegará a los 90 días aproximadamente. Esto significa la demanda del cuarto mes o bien la demanda entre los días 120 y 90 mensualizada.

Venta_cinco: corresponde a la demanda para el quinto período anterior al que se encuentra bajo análisis. El objetivo de esta variable es que el algoritmo detecte patrones

en la demanda de un ciclo de aprovisionamiento (120 días o 4 meses) y, de esta, forma evaluar si existe una relación entre los meses previos de demanda y el mes bajo análisis.

Venta_seis: corresponde a la demanda para el sexto período anterior al período bajo análisis. Es análogo a Venta_cuatro y Venta_cinco en cuanto a objetivos de confección de la variable.

Venta_siete: corresponde a la demanda para el séptimo período anterior al período bajo análisis. Es análogo a Venta_cuatro, Venta_cinco y Venta_seis en cuanto a objetivos de confección de la variable.

Venta_acum_cinco: indica la venta acumulada para dos períodos inmediatos anteriores (Venta_cuatro + Venta_cinco). El objetivo de este tipo de variables es que el algoritmo pueda detectar la saturación de la demanda estándar del SKU y pueda utilizarla para realizar mejores predicciones. Como ejemplo orientativo se podría decir que si la demanda acumulada de los últimos n días ha sido estadísticamente mucho menor (o mayor) a la demanda promedio, el modelo podría interpretar que es una señal de incrementar la previsión de demanda o bien el modelo podría interpretar la señal opuesta.

Venta_acum_seis: indica la venta acumulada para dos períodos inmediatos anteriores (Venta_cuatro + Venta_cinco + Venta_seis). El objetivo de esta variable es análogo al de la variable Venta_acum_cinco, aunque, con una mayor amplitud temporal.

Venta_acum_siete: indica la venta acumulada (al igual que en las dos variables anteriormente mencionadas) para dos períodos inmediatos anteriores (venta_cuatro + venta_cinco + venta_seis + venta_siete).

4. Estrategias de validación de modelos

Tanto los modelos más tradicionales de forecasting como aquellos compuestos por algoritmos muy complejos deben ser evaluados por el usuario que corroborará si el modelo elegido ha tenido el desempeño deseado. Aquellos modelos más complejos recientemente mencionados se encuadran bajo el marco teórico de las problemáticas de *machine learning*. La definición estas pueden comprenderse bajo la definición de Arthur Samuel en 1959: “el estudio y construcción de algoritmos que pueden aprender de datos y hacer predicciones sobre ellos”. Como consecuencia de la realización de estas predicciones mencionadas, las técnicas de evaluación asociadas a las mismas serán fundamentales para determinar la calidad de la predicción y el motivo de su estudio en el siguiente apartado.

4.1. *Training y Testing Sets*

Para el presente trabajo se ha utilizado un enfoque de conjunto de validación o *validation set*. Esta es una estrategia de validación simple y muy utilizada para problemas de *machine learning*. El proceso implica dividir aleatoriamente el conjunto disponible de observaciones de aproximación en dos partes, un conjunto de entrenamiento y un conjunto de validación. El conjunto de entrenamiento se utiliza para desarrollar el mejor modelo posible y dicho modelo ajustado se usa para predecir las respuestas de las observaciones en el conjunto de validación. La tasa de error del conjunto de validación resultante se somete a una o más métricas de error que, como se verá más adelante, proporciona una estimación de la tasa de error de la prueba y una eficacia del modelo.^{vii}

Dado que en el presente trabajo se están utilizando series de tiempo hay algunas consideraciones para tener en cuenta antes de aplicar un conjunto de validación. Una de ellas es que tomar series de tiempo de forma aleatoria para todo el dataset sería cometer *data leakage*, una especie de trampa con los datos, ya que una serie de tiempo futura podría incorporar de forma implícita el desempeño de la variable en el pasado. Por ejemplo, si se predice la demanda de un SKU en el período $n-2$, entrenando con datos de los meses $n-3$, $n-1$ y n , la predicción muy posiblemente tendrá algún tipo de información oculta, como pueden ser patrones de comportamiento estacional u otros efectos producto del aprendizaje que permitan al modelo predecir con un nivel bajo de error, pero con información que no se contará al momento de predecir con datos no ocurridos o inciertos.

Para todos los modelos evaluados se ha seguido el enfoque de Hyndman que establece que el entrenamiento debe realizarse con períodos consecutivos y el proceso de validación deberá realizarse dejando intervalos de tiempo entre medio. Esta metodología propuesta y estudiada por el autor puede subdividirse en tres propuestas de validación cruzada para series de tiempo:^{viii}

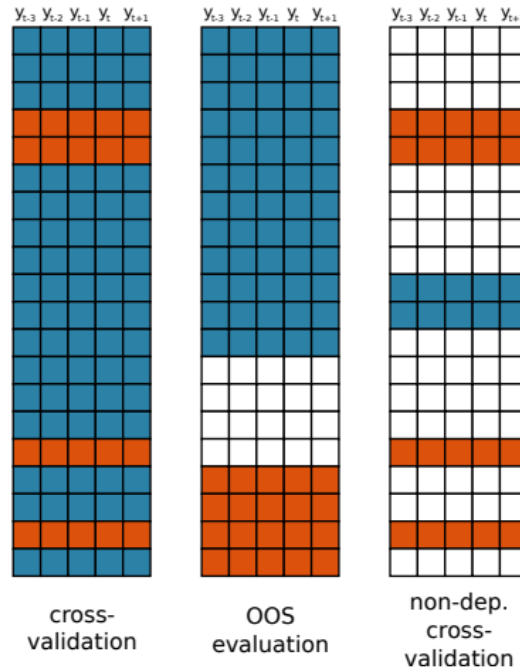
K-Fold Cross validation: este proceso implica la subdivisión del dataset en una cantidad n de *subsets*, los modelos se entrenan con $n-1$ de ellos y el último es utilizado para validación. Este proceso se repite para que cada *fold* sea el de validación una vez.

Out of Sample (OOS): se utiliza un bloque de datos del final de la serie para la evaluación.

Non-dependent CV: Se usan los mismos *folds* que para *K-Fold Cross Validation* anterior pero las filas se eliminan del conjunto de entrenamiento si tienen una distancia entre períodos menor a p de una fila en el conjunto de prueba, donde p es el orden máximo del

modelo. Esta práctica hace que se entrene un modelo con una cantidad de datos mucho menor al dataset original.

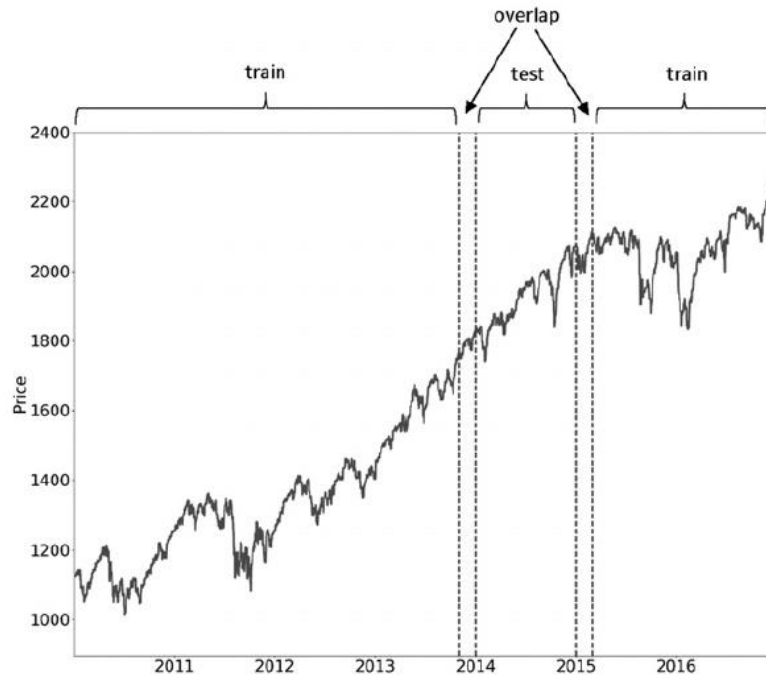
A continuación, se presenta un gráfico del autor donde se comparan las tres estrategias de validación. Las celdas azules indican sets utilizados para entrenar el modelo, las celdas naranjas aquellas utilizadas para validarlo y, las que no poseen color no han sido incluidas en ninguno de estos grupos.



4.1.a Esquemas de validación de series temporales según R. Hyndman

Otro autor que ha hecho un estudio intensivo de series de tiempo y recomienda el uso de sets de validación separados temporalmente del de entrenamiento es López de Prado. Si bien este autor se especializa en series financieras, el análisis de predicciones para series temporales tiene la misma justificación sobre la correlación serial que éstas arrastran. El autor toma un enfoque de Validación Cruzada con Purga y Embargo.

La Purga del conjunto de entrenamiento consiste en eliminar todas las observaciones cuyas etiquetas coincidan en el tiempo con las etiquetas o patrones similares incluidas en el conjunto de prueba. El proceso de Embargo surge como consecuencia de las características de series temporales que exhiben una correlación serial (como en los procesos autorregresivos de medias móviles). Para estos casos es menester eliminar del conjunto de entrenamiento las observaciones que siguen inmediatamente a una observación en el conjunto de prueba.^{ix}



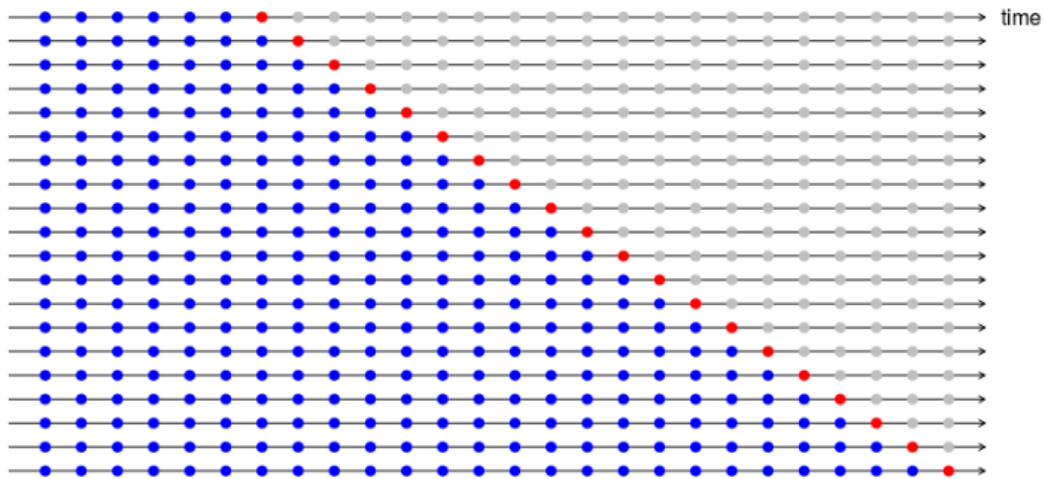
4.1.b Gráfico de Purga y Embargo planteado por López de Prado

4.2 Método de validación seleccionado.

Luego de la exhaustiva investigación de los métodos de validación de pronósticos sugeridos por los autores anteriormente nombrados y del análisis de la información disponible para el producto bajo estudio, se procedió a confeccionar la estrategia de validación de los resultados de los modelos que fueran a ser testeados. De la observación de los datos disponibles, la cantidad de períodos de demanda fue determinante a la hora de establecer la metodología de validación. Adicionalmente, si bien se evaluó ampliar la serie de tiempo hacia atrás utilizando la demanda de productos similares comercializados por la compañía durante los años 2019 y 2020, la alternativa fue descartada por encontrar a las demandas de los rangos extendidos disponibles solamente en el negocio minorista.

Otra posibilidad sopesada estuvo relacionada al cambio de unidades de tiempo, es decir, utilizar demandas semanales en vez de utilizar la demanda mensualizada para cada SKU. Esta última alternativa no resultaba armónica ya que, desde el punto de vista de la cadena de suministro, realizar un pronóstico de forma semanal no resultaba económicamente factible y se estaría dejando fuera del período de evaluación a la misma cantidad de información. En función delo detallado anteriormente se tomó la decisión de utilizar la metodología impulsada por R. Hyndman que, si bien puede considerarse que resulta muy similar a la estrategia de Embargo enunciada por López de Prado, la disponibilidad de datos para una serie de tiempo tan corta no parecía ser concordante con las propiedades que anunciaba el autor.

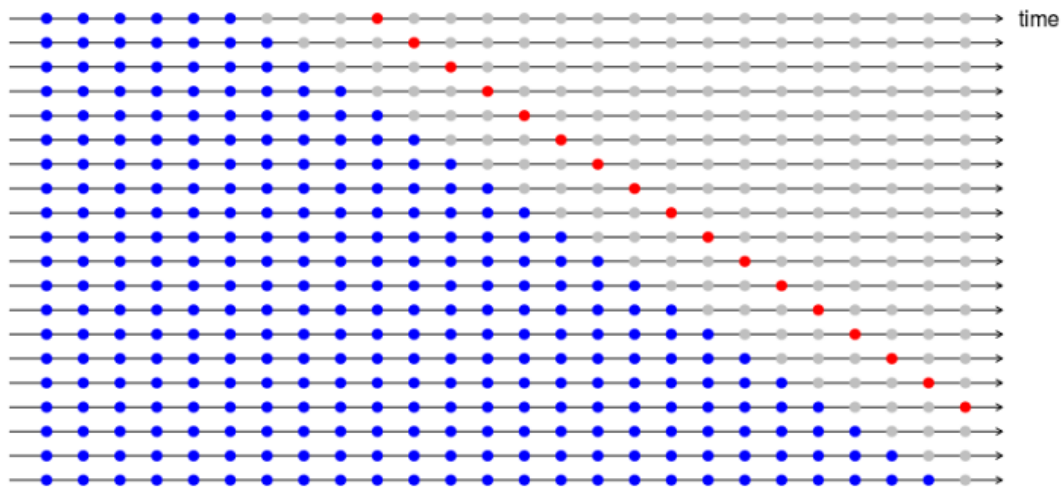
Al ahondar aún más en la aplicación de métodos de validación se han encontrado varias aristas de aplicabilidad de las técnicas. La siguiente imagen ilustra la serie de conjuntos de entrenamiento y prueba enunciados por R. Hyndman, donde las observaciones azules forman los conjuntos de entrenamiento y las observaciones rojas forman los conjuntos de prueba. Como se podrá notar, las observaciones rojas son el período inmediato al de las observaciones de entrenamiento.



4.2.a Diagrama de validación inmediata de R. Hyndman

Esto implicaría utilizar como dataset de entrenamiento los meses de junio a noviembre y realizar la predicción del mes inmediato, es decir, diciembre. Si bien esta estrategia es válida, resulta poco adaptable al contexto estudiado donde el abastecimiento posee un *leadtime* de aproximadamente 76 días. Por lo tanto, en el mes de noviembre no será necesario pronosticar la demanda de diciembre, sino más bien la demanda de febrero del año siguiente. Si cada vez que se realizara la colocación de un pedido de mercadería se obtuviera durante el mes posterior, es muy probable que se hubiera utilizado esta forma de validación para aprovechar una mayor cantidad de datos disponibles. Gracias a estas características del presente trabajo se ha preferido utilizar un modelo de validación de tipo *multi-step*, que adicionalmente posee algunas ventajas frente al diagrama anterior. Un tipo de validación de varios períodos tendrá una menor correlación serial con los datos modelizados y, por lo tanto, los modelos elegidos probablemente tengan una mejor calidad predictiva. Adicionalmente, una evaluación de pasos múltiples se amolda mucho más al contexto de comercialización de lentes oftálmicas que se ha investigado

Si bien una validación cruzada de series de tiempo con el período subsiguiente inmediato puede tener cierta relevancia, Robert Hyndman considera mucho más relevantes aquellos pronósticos de más de un período posteriores al de entrenamiento de los modelos. En este caso, el procedimiento de validación cruzada basado en un origen de pronóstico continuo puede modificarse para permitir el uso de errores de varios pasos. Dada toda esta enunciación de características del conjunto de validación, la estrategia *Out of Sample (OSS)* resulta encajar a la perfección con el tiempo de evaluación de aquellos *leadtime* enunciados en el apartado 3.2.1. De esta forma se logra arribar al segundo esquema planteado por el autor^x:



4.2.b Diagrama multi step según R. Hyndman

Es muy importante mencionar que para el autor la precisión del pronóstico de ambos casos se calcula promediando los errores de los conjuntos de prueba. Este procedimiento se conoce como "evaluación sobre un origen de pronóstico continuo" porque el "origen" en el que se basa el pronóstico se sucede a través del tiempo.

Como conclusión, es fundamental dejar en claro que bajo ningún aspecto se podrían utilizar observaciones futuras para construir un pronóstico. A esto se le suma que no es posible obtener un pronóstico confiable basado en un pequeño conjunto de entrenamiento y que, al mismo tiempo, dicho conjunto sea el inmediato posterior al de los modelos entrenados. Por estos dos motivos se concluyó que era necesario utilizar una validación de varios pasos hacia adelante. Finalmente se podría decir que el presente trabajo contendría de forma análoga a lo que Hyndman expone en la primera línea del diagrama 4.2.b, y que la verdadera esencia del presente proyecto de forecasting de lentes se basa en una continuación histórica de este proceso y que computará como error de modelo al promedio de los errores que vayan surgiendo en la progresión de periodos para optimizar la calidad predictiva conforme pasen los periodos bajo análisis.

5. Modelos investigados, descartados y utilizados

Antes de comenzar a investigar los modelos que se iban a utilizar se descartó la posibilidad de interpretar este problema de stock como un problema de clasificación. Si bien no resultó una decisión muy difícil de tomar se contempló la posibilidad de desarrollar un algoritmo de clasificación múltiple luego de la realización de un *clustering* entre los distintos SKU. Esto fue finalmente descartado ya que un problema de regresión se adaptaba mucho mejor a las principales características de la demanda estudiada en los apartados anteriores.

Una vez establecido el enfoque del problema se acudió a bibliografía relacionada a las diversas maneras de utilizar métodos estadísticos y de *machine learning* para la predicción de demandas en una serie de tiempo. En este sentido podríamos clasificar a los algoritmos estudiados dentro de dos categorías:

- Modelos de regresiones lineales: entre ellas podrían mencionarse las regresiones simples con o sin regresores, las *lasso* o las *ridge*.
- Modelos de regresiones no lineales: aquí existe un universo mucho más variado de herramientas como los *splines*, árboles, máquinas de vector soporte e incluso redes neuronales.

Es importante destacar que ambos tipos de modelos poseen ventajas como desventajas. Por este motivo, para el final del trabajo se presenta al menos uno de cada clasificación previamente enunciada, con la premisa de demostrar que en el caso de una correcta implementación de cada tipo de modelo se podría llegar a resultados satisfactorios.

Los modelos de regresión lineal asumen un comportamiento asociado al de las series de tiempo, esto quiere decir que poseen un sesgo natural orientado por la cercanía temporal de los datos. Por otro lado, el modelado de estas series debe ser estacionario y requiere interacción con sus usuarios. Sin embargo, estos modelos tienen muchas ventajas como por ejemplo el hecho de que el modelo es posible de interpretarse a través de aquellas variables que expliquen el comportamiento de la salida de los datos procesadas. También es importante destacar que este tipo de modelos no tienden a sobre-ajustar los datos, sino que suelen ser más rígidos. Por convención, este sobreajuste de los datos es denominado en la jerga como *overfitting* y, en caso opuesto, se denomina *underfitting*.

En segundo lugar, se encuentran los modelos de regresiones no lineales. Estos son mucho más flexibles y poseen mayor libertad para encontrar relaciones entre los datos, adicionalmente, no requieren la detección de una estacionalidad para ser analizados. Como contrapunto a estos modelos, es muy difícil poder interpretar los resultados de forma analítica cuando se corre un modelo no lineal y adicionalmente a ello el riesgo de sobre ajuste es muy elevado, por lo que requiere un mayor nivel de optimización del modelo para no caer en ilusión estadística del desempeño de estos.^{xi}

A continuación, se muestran algunos modelos que han sido investigados e incluso otros bajo los cuales se ha generado un forecast de predicción de SKU, aunque con una performance menor a la de los modelos finalistas. Por otro lado, se expone el marco teórico sobre los que se erigen aquellos modelos seleccionados para el presente trabajo.

5.1 Modelos estudiados y descartados

- *Prophet*: desarrollo de *Meta Platforms* para el *forecasting* de impresiones en redes sociales. Fue descartado por necesitar establecer la estacionalidad y por

arrojar resultados muy similares al de los modelos exponenciales suavizados.

- Modelos ARIMA: un modelo utilizado por la bibliografía durante varias décadas y con varios ajustes en función del diseño de los datos. Al igual que Prophet se consideró dentro de las predicciones finales por terminar siendo muy similar a los modos más tradicionales de forecasting de demanda.
- Redes Neuronales: La implementación de redes neuronales ha sido explorada principalmente debido a su gran poder predictivo en series de tiempo. Desafortunadamente el algoritmo desarrollado estuvo en varias oportunidades algunas horas intentando converger a un mínimo local pero no lo logró y se terminó descartando frente a otros modelos con buenas performances.
- *Bagging*: durante el desarrollo del algoritmo de *Random Forest* se ha comprobado que la respuesta a la problemática correcta no era la más eficiente bajo estas características. En ese sentido, se descartó el presente modelo de ensamble por encontrar un modelo de su mismo tipo con mejor performance.

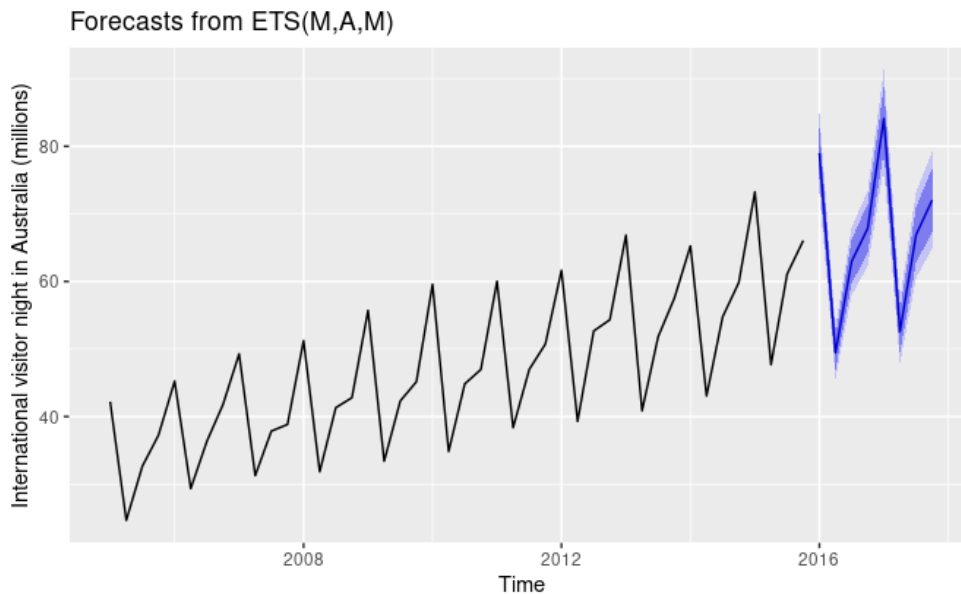
5.2 Modelos seleccionados

5.2.a *Exponential smoothing state space model* (Suavizado Exponencial) ^{xii}

El suavizado exponencial (o también *ETS* como abreviatura de aquí en adelante) es una vieja técnica de pronósticos propuesta a fines de la década de 1950 y ha motivado el desarrollo de métodos de pronóstico considerablemente exitosos. Los pronósticos producidos se basan en métodos de suavizado exponencial; esto significa que las predicciones son promedios ponderados de las observaciones del pasado y la forma de otorgarle mayor o menor peso predictivo a una observación decae exponencialmente a medida que éstas envejecen. En otras palabras, cuanto más reciente sea la observación, mayor será el peso asociado de la predicción del suavizado. Este marco genera pronósticos confiables de forma rápida y para una amplia gama de series de tiempo, lo cual es una gran ventaja para su aplicación en diversas industrias.

Existen varios métodos de suavizado exponencial y por lo general la selección del método se basa en reconocer los componentes clave de la serie temporal, como pueden ser la tendencia o la estacionalidad y la forma en que éstos ingresan al método de suavizado. La forma en la que se incorpora puede ser, por ejemplo, de manera aditiva, amortiguada o multiplicativa. Los modelos generan pronósticos puntuales idénticos a los métodos, pero también generan intervalos de predicción. Además, este marco estadístico permite una selección de modelo genuina entre modelos competidores.

Una característica de este tipo de modelos es que con ellos se pueden generar intervalos de predicción, algo que no se puede hacer con la totalidad de las metodologías aplicadas en el presente trabajo. Los intervalos de predicción podrán variar en función de la utilización de modelos con métodos aditivos o multiplicativos.



5.1.a Gráfico de pronóstico utilizado por R. Hyndman del suavizado exponencial multiplicativo bajo método Holt-Winters' y errores multiplicativos.

Si bien para el enfoque de predicción de la demanda de lentes oftálmicas no se han utilizado los intervalos para la comparación de resultados con los otros modelos, esta característica sería de mucha utilidad en la práctica del abastecimiento si se contara con un embarque óptimo y cómo completar éste en exceso o en defecto dentro de las cotas de tolerancia del modelo aplicado.

5.2.b *Random Forest*

Es uno de los métodos basados en árboles de decisión y una versión más integral de *Bagging*. *Random forest* o árboles aleatorios en castellano, combina técnicas de remuestreo o *bootstrapping* de *Bagging* y por ende mantiene su bajo sesgo. Sin embargo, a diferencia de éste, logra disminuir la varianza al promediar entre varias alternativas detectadas de forma aleatoria, en donde se hace foco durante la optimización del modelo.

Como bien enuncian Hastie, Tibshirani y Friedman en su libro “*The elements of Statistical Learning*”, si bien *Bagging* es una técnica para reducir la varianza de una función de predicción estimada, el método parece funcionar especialmente bien para procedimientos de alta varianza y bajo sesgo, como los árboles. Para la regresión, simplemente se ajusta el mismo árbol de regresión muchas veces para obtener versiones muestreadas de los datos de entrenamiento. Finalmente, obtenidos los árboles de regresiones promedia el resultado. Esta idea es la base que utiliza *Random Forest* que ha sido desarrollado por Leo Breiman en 2001^{xiii} y comprende una modificación sustancial al método de *Bagging* que crea una gran colección de árboles decorrelacionados y, para el caso específico de la regresión, los resultados son promediados mientras que para la clasificación ejecuta una elección por mayoría. En muchos problemas, el rendimiento de los *Random Forest* es muy similar al *Boosting*, que se analizará más adelante, pero a diferencia de esta técnica estos son más fáciles de entrenar y de optimizar. Como consecuencia, el método de *Random Forest* para las regresiones y las clasificaciones son muy populares y se implementan en una gran variedad de problemáticas.

Para explicar el funcionamiento de *Random Forest* se partirá desde el enfoque que se utiliza en *Bagging*. En este modelo ensamblado, al promediar muchos modelos ruidosos, pero aproximadamente imparciales, tiende a reducir su varianza. En este sentido, los

árboles son candidatos ideales para ello, ya que pueden capturar estructuras de interacción complejas en los datos y, si crecen lo suficientemente en profundidad, se obtendría un sesgo relativamente bajo.

Este aspecto de detección de patrones entre los datos fue uno de los factores determinantes para la elección del algoritmo para la predicción de la demanda ya que se consideró que podría tener un excelente poder predictivo al relacionar a los esféricos y los cilíndricos con la variable binaria del rango extendido de los SKU.

Dado que los árboles son notoriamente ruidosos, el hecho de promediarlos genera un beneficio significativo para su poder predictivo

Siguiendo la definición de los autores mencionados anteriormente, continuación se repasa el algoritmo de regresión basado en el método *Random Forest* que se utilizará para la predicción de la demanda de lentes oftálmicas:

1. Para $b = 1$ en B y sea B la cantidad de remuestreos a realizar:
 - (a) Se toman muestras del *dataset* Z^{*B} de tamaño N , siendo esta la cantidad de observaciones que contendrá cada uno de estos *subsets*.
 - (b) Cada uno de esos *subsets* se utilizarán a como datos de entrenamiento.
 - (c) Se genera el crecimiento de un árbol de bosque aleatorio T_b , luego se seguirán de forma recursiva los siguientes pasos para cada nodo terminal del árbol, hasta alcanzar el tamaño mínimo de nodo $nmin$ definido por el usuario (para la regresión se sugiere una cantidad mínima de 5 nodos):
 - I. Se procede a seleccionar m variables al azar de las p variables disponibles, hay que recordar que *Bagging* utiliza todas ellas. Por lo general, los valores de m recomendado por diversos autores son de $p/3$ para la regresión o incluso tan bajos como 1 . En este aspecto b se caracteriza por ser el b -ésimo árbol aleatorio en términos de variables o puntos de corte para cada nodo y valores de nodo terminal. Como puede intuirse, la reducción de m reducirá la correlación entre cualquier par de árboles en el conjunto y ayudará a reducir la varianza del promedio.
 - II. Se procede a elegir la mejor variable o el mejor punto de división entre los m .
 - III. Se procede a dividir el nodo en dos nodos secundarios.
2. Se obtiene la salida del conjunto de árboles: $\{T_b\}^{B_1}$

Obtenido esto, para hacer una predicción en un nuevo punto x se utilizará, como se dijo anteriormente, la suma de cada respuesta b sobre el cociente de los B árboles totales:

$$Regression: \hat{f}_{rf}^B(x) = \frac{1}{B} \sum_{b=1}^B T_b(x)$$

Para finalizar con la descripción de este modelo de predicción de demanda resta enunciar el método de validación a utilizar en este modelo. Tanto *Bagging* como *Random Forest*, utilizan un enfoque de fuera de la bolsa u *Out of the Bag* (OOB, por sus siglas en inglés) para la estimación del error del modelo. En estos casos, el remuestreo que hace el algoritmo reestima el error al dejar fuera del cómputo las distintas entradas del *dataset*,

es decir, que en este caso la medición del error continúa realizándose a lo largo del camino del remuestreo de árboles. Una vez que el error OOB se estabiliza, se puede terminar el entrenamiento. En el presente trabajo, adicionalmente al remuestreo anunciado anteriormente y, en concordancia con el punto 4.2 del presente trabajo, existe un conjunto OOB adicional que ha sido el mismo para el resto de los modelos de predicción que es el del mes de marzo de 2022, junto a la exclusión de aquellos datos atribuibles a los meses de *leadtime*.^{xiv}

5.2.c Generalized Boosted Regression Modeling

Estos modelos son conocidos como máquinas *Boosting*, y al igual que *Random Forest* se basan en árboles de decisión. Al igual que *Bagging*, *Boosting* es un enfoque general que se puede aplicar a problemas de aprendizaje supervisado tanto de regresión como de clasificación.

Cuando se realizó el análisis de *Bagging*, se mencionó que se basaba principalmente en varios remuestreos del mismo set de entrenamiento con el que se promediaba la decisión de cada uno para llegar a un veredicto sobre el valor de la variable objetivo. Las máquinas *Boosting* funcionan de manera análoga, con la excepción de que los árboles se generan de una forma secuencial. Esto significa que cada árbol utiliza la información de los árboles generados anteriormente. Esto significa que los modelos *Boosting* no remuestran B veces sobre el mismo set de datos, sino que el nuevo árbol se ajusta a una nueva versión del conjunto de información original. Siguiendo la lógica del procedimiento y a los conceptos revisados en árboles, en los métodos *Boosting* se incorporará un nuevo parámetro llamado λ (*lambda*) que controlará la tasa a la que el modelo aprende de los datos al generar una mayor o menor cantidad de particiones del mismo set de datos.

Tal como se hizo con *Random Forest* en el apartado anterior, se presenta a continuación el algoritmo de regresión de máquinas *Boosting* incorporado en “*Introduction to Statistical Learning*” donde se expone el razonamiento algebraico sobre el que se realizará la predicción de la demanda de cada SKU para un período determinado^{xv}:

1. Se determina $\hat{f}(x) = 0$ con $r_i = y_i$ donde r significa el error residual que inicia siendo igual al valor de la variable objetivo y a predecir para cada i en el set de entrenamiento.
2. Para $b = 1, 2, \dots, B$, se repite recursivamente:
 - a. Ajustar un árbol \hat{f}^b y dividirlo en d divisiones equivalentes a $d+1$ nodos para entrenar el set de entrenamiento (X, r) . En este punto, se ajusta la función objetivo del modelo junto al error residual del modelo (aquel que inició siendo igual a la variable objetivo).
 - b. Actualizar \hat{f} en la porción disminuída del set de entrenamiento con un nuevo árbol:

$$\hat{f}(x) = \hat{f}(x) + \lambda \hat{f}^b(x)$$

- c. Se procede a actualizar los errores residuales del modelo, el parámetro *lambda* en este sentido será aquel que regule la tasa de aprendizaje bajo la cual se aspira a reducir el error de los residuales.

$$r_i \leftarrow r_i - \lambda \hat{f}^b(x)$$

3. Se obtiene el modelo de salida *Boosting*,

$$\hat{f}(x) = \sum_{b=1}^B \lambda \hat{f}^b(x)$$

Como se desprende de la función existen tres maneras de optimizar el algoritmo. El primero es en función de B , que son las cantidades de árboles. Si B es demasiado grande, se generan muchos residuos en función de la cantidad de árboles y como consecuencia comienza un lento proceso de sobreajuste del modelo. En segundo lugar, se encuentra el parámetro de contracción λ que controla la velocidad a la que aprende *Boosting* y por ello suele ser un número pequeño pero que, dependiendo el problema, se puede ajustar. Si bien los valores típicos señalados por la teoría son de 0,01 o 0,001, el usuario del modelo será encargado de optimizar este patrón de aprendizaje. Por último, se debe considerar el número d de divisiones en cada árbol, es decir, la profundidad de interacción entre los árboles y las variables. Esta característica ayuda a controlar los factores prioritarios para la interacción del modelo potenciado con d cantidad de variables.

La manera por lo general de optimizar estos hiper-parámetros del modelo es mediante la validación cruzada. Adicionalmente, este modelo funciona de una forma aditiva bajo las cuales los árboles operan sobre los residuos dejados por los árboles anteriores. Como consecuencia esto suele generar un proceso de aprendizaje mucho más lento comparado con otras metodologías. Como última ventaja del modelo, el uso de árboles más pequeños sobre los residuos puede otorgar cierta interpretabilidad a estos modelos en aquellos casos de baja profundidad de árboles. En el caso del presente trabajo se podrá observar cómo las distintas variables aportan mayor o menor significancia a la reducción del error entre los pronósticos y los datos del set de entrenamiento.^{xvi}

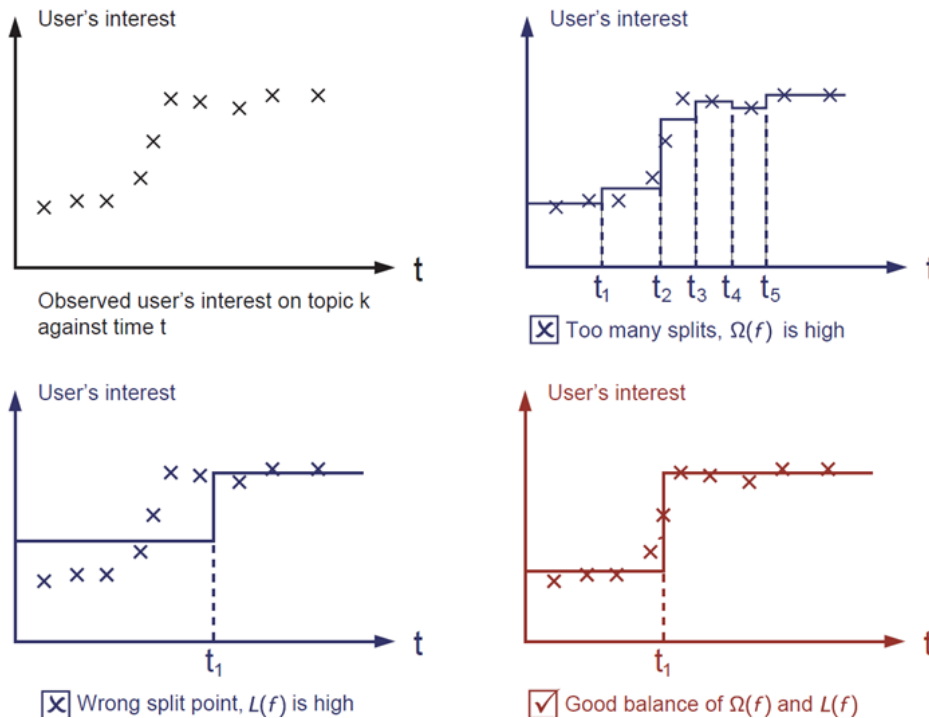
5.2.d *XGBoost (Extreme Gradient Boosting)*

Las máquinas *Boosting* han demostrado tener una gran versatilidad para resolver todo tipo de problemas a medida que la capacidad de procesamiento de datos fue aumentando a lo largo del tiempo. Esto resulta de esta manera ya que el proceso lento de aprendizaje de un algoritmo *Boosting* se comenzó a optimizar con el aumento de la capacidad de procesamiento de los equipos de computación y que, en consecuencia, se vuelve redituable el uso de estos algoritmos en problemáticas que antes se resolvían con cálculos menos sofisticados como el suavizado exponencial simple que se ha presentado anteriormente.

En este sentido, *XGBoost* es una mejor formalización y una implementación inteligente de GBM. En líneas generales utiliza una formalización de modelo más regularizada que ayuda a controlar el sobreajuste de los modelos y, en consecuencia, le da una mejor performance.^{xvii} A continuación se muestran algunas de las características agregadas a este modelo que sigue los principios enunciados en el apartado anterior.

Como se dijo anteriormente, la regularización de un modelo resulta ser una etapa clave en la que se puede establecer una mejora sustancial para todo modelo de *machine learning*. Esta premisa está presente tanto para modelos demasiado rígidos como el suavizado exponencial donde el algoritmo tiende a simplificar mucho la complejidad del problema, como para modelos donde la regularización también sea un factor donde el

usuario pueda accionar. A su vez, la regularización de modelos suele tener mayor relevancia para aquellos modelos donde la cantidad de formas de arribar a una predicción correcta no es tan clara ni lineal y la flexibilidad del algoritmo llega a distorsionar la performance general del modelo final cuando se lo enfrenta a la problemática real. Como ya fue mencionado anteriormente, esta distorsión es conocida como *overfitting* y suele ser la consecuencia de sobre adaptarse a los datos de entrenamiento.



5.2.d.1 Ejemplificación de regularizaciones correctas e incorrectas de modelos según la documentación de XGBoost

La cantidad de hiper-parámetros disponibles para regularizar un modelo de *forecasting* de series temporales con una cantidad de información relativamente escasa fue la principal razón por la cual se ha considerado la utilización de este modelo. A su vez, y como se verá más adelante, esta herramienta posee una versatilidad que suele consumir un mayor nivel de tiempo en comparación a las vistas anteriormente.

A continuación, se enumeran varios de los hiper-parámetros que XGBoost incorpora al algoritmo de *Boosting* con el objetivo de mejorar la performance de las predicciones:

- Enunciadas en el modelo General:

Nrounds: es la cantidad de árboles (B) es decir, el número máximo de iteraciones que se van a entrenar.

Eta (tasa de aprendizaje o λ): como se mencionó anteriormente, es el parámetro que administra la tasa de aprendizaje del modelo. Después de cada *Boosting* este parámetro trabajará en aprender los patrones más relevantes con el objetivo de prevenir el sobreajuste del modelo.

Max_depth (*d*) o Profundidad máxima de un árbol: aumentar este valor hará que el modelo sea más complejo y propenso a sobre ajustarse.

- Incorporaciones de la herramienta *XGBoost*

Gamma: es la mínima reducción del error necesaria para aceptar un nuevo corte. Cuanto mayor sea gamma, más conservador será el algoritmo.

Min_child_weight: es la cantidad mínima de observaciones que debería tener un nodo para poder realizar un corte. Cuanto mayor sea, más rígido será el algoritmo que defina el modelo.

Subsample: indica qué proporción de los datos serán considerados para cada árbol. Si por ejemplo se estableciera en 0.5, XGBoost muestrearía aleatoriamente la mitad de los datos de entrenamiento antes de comenzar a trazar árboles. El objetivo de esta variable será evitar el sobreajuste del modelo. El submuestreo ocurrirá una vez en cada iteración de cada *Boosting*.

max_delta_step: cuando toma un valor positivo, puede ayudar a equilibrar el balanceo de clases. Para el presente trabajo, este parámetro no será utilizado.

sampling_method: es el método que el usuario determina para muestrear al set de entrenamiento. Este puede ser uniforme (combinado con *subsample* \geq 0.5) o *gradient based*, método restringido a algunos tipos de árboles que implica que la probabilidad de selección en el muestreo es proporcional al valor absoluto regularizado de los gradientes.

Colsample_bytree: bajo este parámetro se le indica al modelo a diseñarse qué proporción del total de las variables se van a considerar dentro del entrenamiento de cada árbol.

colsample_bylevel: es la relación de *subsampling* de columnas para cada nivel. Cada vez que se genera un corte para crear un nuevo árbol, este vuelve a tomar un nivel de muestras sobre el total de columnas del dataset para generar este árbol.

6. Métricas de Evaluación de Modelos

Como se ha mencionado en capítulos anteriores, el problema del presente trabajo tendrá un enfoque de aprendizaje automático basado en técnicas de regresión. Para evaluar estas técnicas de aprendizaje y poder seleccionar las mejores interpretaciones que se hayan realizado en los modelos desarrollados, es necesario tener un conjunto de métricas para validar su efectividad. Adicionalmente, estas métricas muy posiblemente ayuden a encontrar ese tan necesario equilibrio entre el sesgo y la varianza del modelo. A continuación, se detallan las métricas elegidas en función de la selección del autor y la experiencia relevada en otros trabajos estudiados.

En una segunda instancia, se abarcarán aquellas métricas que podrían considerarse económicas o propias del negocio. Esto quiere decir que, si bien el modelo poseerá métricas propias de las buenas prácticas del aprendizaje automático, existirán otras que, como contrapartida, deberán tender a establecer pautas de eficiencia del negocio bajo análisis.

Es necesario destacar que para ambos tipos de métricas se utilizará la notación matemática tradicional que será acompañada de dos variables elegidas por el usuario para hacer referencia a los dos escenarios posibles. En función de ello, se ha determinado denominar como *Actual* a aquella demanda de un SKU para el período bajo análisis y *Predicted* como la variable predicha por el modelo bajo análisis. Adicionalmente, se utilizará un subíndice i en referencia al espacio temporal, es decir, el mes bajo análisis sobre el cual se pueda ubicar esta variable y referenciar tanto a períodos posteriores como anteriores.

Por último, más allá de cualquier métrica sobre la que un usuario base la elección del mejor modelo, es fundamental tener en cuenta que lo que verdaderamente importa a la hora de definir la calidad predictiva de un modelo es cómo es su performance por fuera de los datos de entrenamiento. Es decir, un modelo puede tener una métrica baja, cualquiera sea la utilizada, sobre los datos de *train*, pero ser mucho más elevado sobre los datos de validación. Esto la posicionaría peor ante otra técnica que funcione de manera inversa.

6.1 Métricas propias de los modelos utilizados.

6.1.1 MSE

El MSE, por sus siglas en inglés, se traduce como el Error Cuadrático Medio (ECM) y mide qué tan lejos se encuentran las predicciones realizadas por el modelo de los verdaderos valores observados. Las variables a las que se aplica la fórmula se desprenderán de $Predicted_i$, que denota las predicciones de demanda obtenidas en los modelos seleccionados y donde $Actual_i$ es la variable que indica los valores reales observados de demanda del SKU en cuestión. Cuanto menor sea esta medida, mejor será el modelo implementado. Es por esta razón que, a la hora de optar por un modelo en lugar de otro, lo recomendable es inclinarse por aquel que tenga el menor ECM sobre un conjunto de validación. Como características generales de esta métrica se puede mencionar que cada error influye en el total en proporción a su cuadrado. Como resultado, los errores grandes tienen una influencia relativamente mayor en el error cuadrático total que los errores más pequeños.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Actual_i - Predicted_i)^2$$

6.1.2 RMSE

La raíz cuadrada del error cuadrático medio (RMSE por sus siglas en inglés) mide la cantidad de error que hay entre dos conjuntos de datos. Se obtiene al aplicarle la raíz cuadrada al MSE. xviii

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(Actual_i - Predicted_i)^2}{n}}$$

6.1.3 MAE

El error absoluto medio es una medida de la diferencia absoluta entre las dos variables, lo que en el trabajo presente son los *Predicted_i* versus los *Actual_i* y sobre ella, la división sobre la cantidad de observaciones. Se diferencia en comparación a las métricas mencionadas anteriormente en que no escala el error a medida que se incrementa la varianza. En este sentido, no es inequívoca la preferencia de esta métrica sobre las anteriores en la bibliografía consultada.

$$MAE = \frac{1}{n} \sum_{i=1}^n |(Actual_i - Predicted_i)|$$

6.1.4 sMAPE

El error porcentual absoluto medio simétrico es una medida de precisión basada en errores porcentuales. Definido inicialmente por Armstrong (1985), y luego refinado en su versión actual por Flores (1986) y Makridakis (1993). El sMAPE fue propuesto para enmendar los inconvenientes de la métrica MAPE que, a diferencia del sMAPE, solo incluye en el numerador a la variable predicha, la desventaja que presenta es que los errores en variables de gran tamaño poseen un menor peso nominal en la sumatoria total.

Por definición, los valores de sMAPE oscilan entre 0 y 200 . Cuanto menor sea el valor de sMAPE, mejor será la precisión predictiva de un modelo determinado.^{xix}

$$sMAPE = \frac{1}{n} \sum_{i=1}^n 200 \times \left| \frac{(Actual_i - Predicted_i)}{(Actual_i + Predicted_i)} \right|$$

6.2 - Métricas de evaluación del Negocio

Son aquellas que se han definido para lograr la mejor performance posible de la utilización económica de los activos de la compañía. Para este caso se han utilizado principios de contabilidad de Costos y Contabilidad Financiera bajo estándares internacionales IFRS.^{xx}

6.2.1 Porcentajes del desvío sobre la cantidad pronosticada

La cantidad total pronosticada es evaluada en términos del producto sin importar el SKU. Esta es una métrica general elaborada para lograr entender si el modelo ha logrado capturar la tendencia de la demanda en general. Esto será calculado mediante una comparación entre el total de demandas reales durante el período de evaluación del modelo contra las cantidades pronosticadas para cada SKU para cada una de las técnicas de aprendizaje automático que fueran llevadas a cabo.

$$\text{Porcentaje de desvío } Q \text{ total} = \sum_{i=1}^{SKU} \frac{(\text{Actual}_i - \text{Predicted}_i)}{\text{Actual}_i}$$

6.2.2 Desvíos sobre la demanda en defecto.

Aquellos SKU sobre los que se haya pronosticado una demanda por debajo de la estimada serán considerados como quiebres de *forecasting*, es decir sin considerar el stock. Estos errores serán evaluados en cantidad total como en porcentajes sobre el total demandado. Aquí la cantidad de SKU sobre los que se produzca el quiebre pierde importancia ya que uno de los objetivos de esta métrica es ponderar a aquel modelo que menor traslado a producción hubiera tenido si no se contara con el respaldo del stock. A diferencia de la demanda en exceso, la demanda en defecto posee un traslado a costos inmediato y por eso es potencialmente más importante que la demanda en exceso, mientras que la primera será penalizada en el análisis de lenta rotación, pero con la posibilidad de ser compensada durante los períodos de reposición que puedan existir en la compañía.

$$\text{Quiebres del forecasting (no incluye stock)} \rightarrow \sum_{i=1}^{SKU} 0 > (\text{Actual}_i - \text{Predicted}_i)$$

6.2.3 Quiebres de stock del pronóstico

Llamamos quiebres de stock a aquella cantidad de ventas del mes en análisis que nuestro modelo de predicción de demanda no ha podido responder aun con el apoyo del stock al cierre del mes anterior. Esto implica que la lente hubiera tenido que producirse en un laboratorio y en consecuencia generar costos adicionales.

$$\text{Producción por quiebres de stock} \rightarrow \sum_{i=1}^{SKU} 0 > (\text{Actual}_i - \text{Predicted}_i) + \text{Stock}_{i-1}$$

6.2.4 Lenta rotación de SKU's. Rentabilidad del pronóstico.

La compañía tiene definida una política de desvalorización de sus stocks. La política consta en desvalorizar al 50% al exceso de stocks que se desprendan de la demanda de los últimos 6 meses para cada SKU. Esta política funciona como un contrapeso de la métrica de quiebres de stock ya que excederse demasiado en el cálculo de demanda genera pérdidas por desvalorización que tardarían varios semestres en reducirse. En función de lo dicho anteriormente, la eficiencia de los modelos también será evaluados por la relación de equilibrio que presenten entre la producción de lentes con exceso de demanda y el sobre aprovisionamiento de mercadería.

Se define a continuación el cálculo de la penalización de la lenta rotación:

Lenta rotación :

$$\frac{(\text{Actual}_i + \text{Actual}_{i-1} + \text{Actual}_{i-2} + \text{Actual}_{i-3} + \text{Actual}_{i-4} + \text{Actual}_{i-5}) - \text{Stock}_i}{2} \times \text{Costo}$$

Luego, y asumiendo que PU es el precio unitario de venta de cada unidad demandada y que el costo de la producción de una lente tallada es 2.5 veces el costo de una unidad

convencional más la utilización de la ecuación de Lenta rotación se produce la fórmula de rentabilidad de la predicción como se expone a continuación:

Eficiencia costos vs lenta rotación (considera stock)

$$\frac{[(Actual_i \times PU - (Lenta rotación) - (Producción por Quiebras \times Costo \times 2.5)]}{(Actual_i \times PU)} \times 100\%$$

Donde, como puede observarse, la cantidad más elevada a la que puede llegar este cálculo es del 100% si se logra minimizar en su totalidad el impacto de la lenta rotación y de los quiebras.

7. Ejecución y optimización de modelos elegidos

Previo al inicio de este apartado se considera necesario realizar una introducción a los datos contra los que se contrastaron cada uno de los modelos, es decir, mostrar un resumen de los datos del escenario contra los que se compararán los *validation sets* obtenidos mediante las diversas técnicas.

Durante el mes de marzo de 2022, el mes bajo el cual se han preparado los 4 modelos para verificar su poder de predicción y adaptabilidad a la realidad, han sido vendidas un total de 17.919 lentes entre los 1041 distintos SKU que componen el universo del producto lanzado. De esta cantidad, un total de 12.398 pertenecen al rango regular de graduaciones mientras que 5.521 son del rango extendido, es decir el 30,8%, un porcentaje estadísticamente comparable con el histórico de ventas de rangos extendidos que se puede ver en el análisis exploratorio de los datos. Por otro lado, el número de tallados afrontados por la compañía importadora ascendió a 305, lo que implica que han existido 305 episodios bajo los que el stock del producto en graduaciones puntuales se vio superado por la demanda durante del mes. Esta cifra, si se compara con los tallados realizados durante el promedio histórico, resulta más baja que la media, la cual oscila en las 356 lentes, es decir un 14% más bajo que la media histórica.

El stock al inicio del período marzo 2022 era de 68.516 lentes, lo que significa que durante marzo 2022 se vendió un 26% del stock. Este porcentaje implica un cálculo general de rotación de stock para el producto de 114 días de stock. Es posible que esta relación superior a 90 días explique la merma de los tallados para el mes bajo análisis.

A continuación, se presentan los resultados y aquellas características salientes de los modelos finales ejecutados. Si bien el objetivo es el de presentar de forma homogénea los resultados obtenidos para cada modelo y bajo qué aspectos se trabajó la mejora de la performance interna de cada uno, es necesario destacar que, por sus diversos aspectos evaluados no se buscará en este apartado la determinación del mejor modelo sino el análisis de los resultados arrojados por cada uno de forma puramente individual y aislada del resto. Durante el transcurso del apartado de evaluación de los resultados se buscará generar una interpretación integral de todos ellos y seleccionar el modelo que mejor haya captado las necesidades del presente trabajo.

7.1 Exponential smoothing state space model (ETS)

Desarrollo del modelo

Para ello fue necesario actualizar al dataset a una clase de datos de series de tiempo mensuales para poder procesarlo de forma correcta en las librerías "Forecast" publicadas por R. Hyndman. A su vez, se utilizó el argumento que habilita el *multi step series validation (nmse)* como ha sido enunciado en el cuarto apartado. Si bien la respuesta del algoritmo puede contemplar intervalos de confianza para el *output*, los mismos no fueron activados ya que el objetivo era poder generar un escenario de comparabilidad con el resto de los modelos que arrojan estimaciones puntuales.

Optimización del modelo

El modelo permite la utilización de los siguientes criterios de optimización: *mse*, (error cuadrático medio), *amse* (promedio del error cuadrático medio), *sigma* (desvío estándar de los residuos), "*mae*" (media de los residuales absolutos) o *lik* (log -probabilidad, el valor por defecto del modelo). Si bien se experimentó con los cinco criterios, el promedio del error cuadrático resultó ser el que mejores predicciones realizó para el mes de marzo de 2022 y, por ende, resultó seleccionado para este apartado.

Otra característica del algoritmo sobre la que no se pudo trabajar es la de relacionar al modelo con una potencial estacionalidad y utilizarlo como regresor del modelo. Esto no fue posible debido a la inexistencia de datos con más de un año de antigüedad.

Finalmente, cuando se realizó el análisis de utilización de la tendencia como regresor aditivo o multiplicativo, se estableció que la tendencia aditiva fue la de mejor desempeño para este modelo puntual.

Desempeño de la predicción general y unitario

A continuación, se presentan los resultados generales para el modelo ejecutado:

ETS

	Real	Pronóstico	Var.	Var. (%)
Cantidades demandadas	17.919	11.844	-6.075	-34%
Tallados por quiebres	-305	-83	222	-73%
Tallados/Total	-1,7%	-0,46%	1,2%	-73%

	Cantidad SKU	Unidades
SKU en exceso	535	1.192
SKU en defecto	506	-7.267
Total	1.041	-6.075

El modelo ha tenido un desempeño llamativo, principalmente porque a priori, al ver un desvío de cantidades reales versus cantidades pronosticadas de aproximadamente el 51%, no se esperaba una disminución del porcentaje de SKU's cuya demanda fue superada por el pronóstico de este modelo. Adicionalmente, al adicionar el stock al 28/02/2022, el modelo alcanzó excelentes performances de rentabilidad donde ésta hubiera sido del 96%; la baja en la venta proyectada ayudó a producir una menor penalización por lenta rotación junto a los tallados. La cantidad de tallados pronosticados paradójicamente fue más de 3 veces menor a la real.

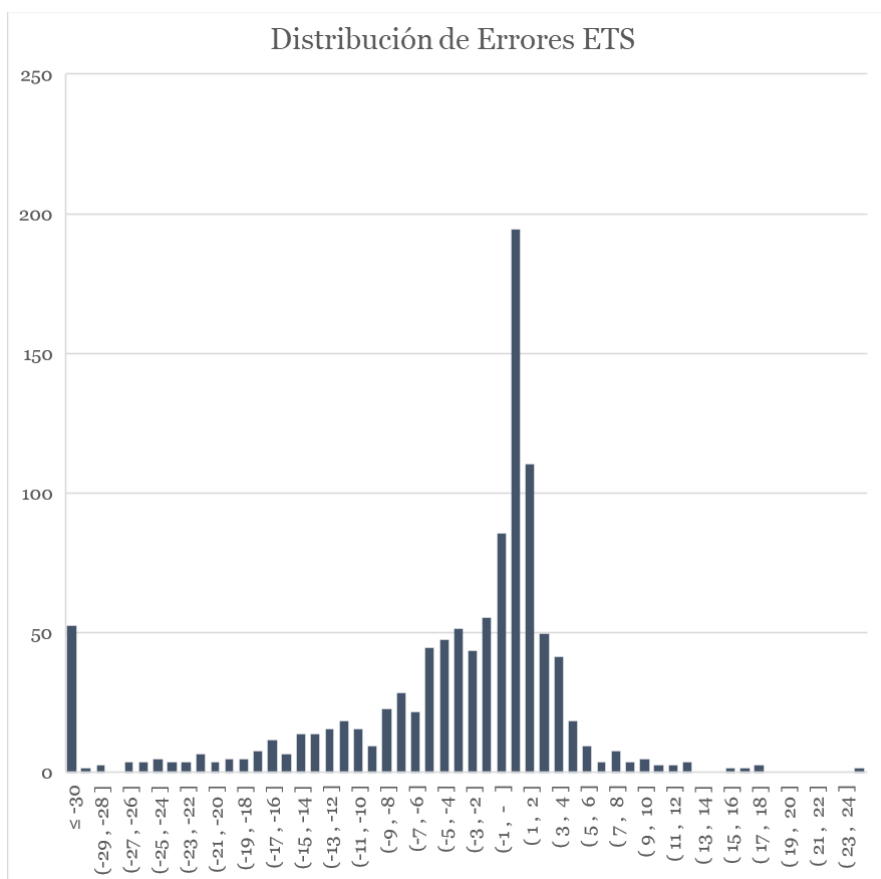
Distribución de errores

A continuación, se presentan en forma conjunta las 8 métricas bajo las cuales se ha evaluado al modelo ETS junto a una gráfica que concentra las distribuciones de errores nominales del modelo.

ETS

Métricas		Errores	
Porcentaje de desvío Q total	51,29%	MSE	474,14
Quiebres del forecasting (no incluye stock)	-40,55%	RMSE	21,77
Producción por quiebres / Venta	-0,46%	MAE	8,13
Eficiencia costos vs lenta rotación (considera stock)	95,98%	sMAPE	87,36

La distribución de errores refleja una gran concentración de errores de SKU's en el intervalo de $[-5, 2]$ con alrededor de 650 observaciones dentro de este rango. Por otro lado, fuera de estos valores enunciados anteriormente, se puede observar una inclinación al error negativo de los errores menores a -30 (predicción en defecto). Este rango de errores menores a -30 representa más del 50% de los pronósticos en defecto del modelo en contraposición a los errores positivos no se elevan significativamente en el extremo positivo del histograma 7.1.a. Esto puede interpretarse con cierta razonabilidad al observar que el porcentaje de desvío de la cantidad total pronosticada versus la cantidad total demandada estuvo sustancialmente inclinado en esta dirección.

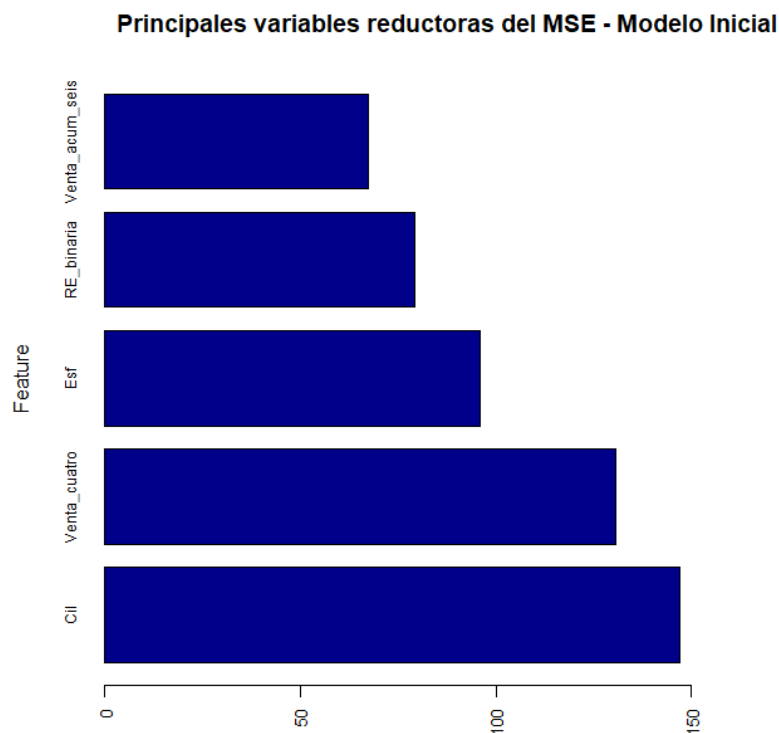


7.1.a Distribución de errores para el modelo ETS con escala de uno en uno y acotado a ± 30 observaciones.

7.2. *Random Forest*

Etapas iniciales del modelo

Luego de llevar a cabo el proceso de formación del dataset de las demandas de lentes se procedió a entrenar un modelo de *Random Forest*. Se dio inicio al modelo utilizando la métrica de profundidad sugerida por los libros de texto, es decir la cantidad de *features* del set de datos sobre 3. Por otro lado, se estableció una cantidad inicial de 500 árboles y una cantidad máxima de nodos de 10^{xxi} . Como puede observarse se utilizaron características de un modelo estándar para luego realizar una mejora del modelo a través de técnicas algorítmicas. A continuación, se pueden observar las variables que el modelo inicial presentó como principales aportantes a la reducción del error:



7.2.a Principales variables que explican la reducción de errores cuadráticos medios del modelo *Random Forest* al inicio.

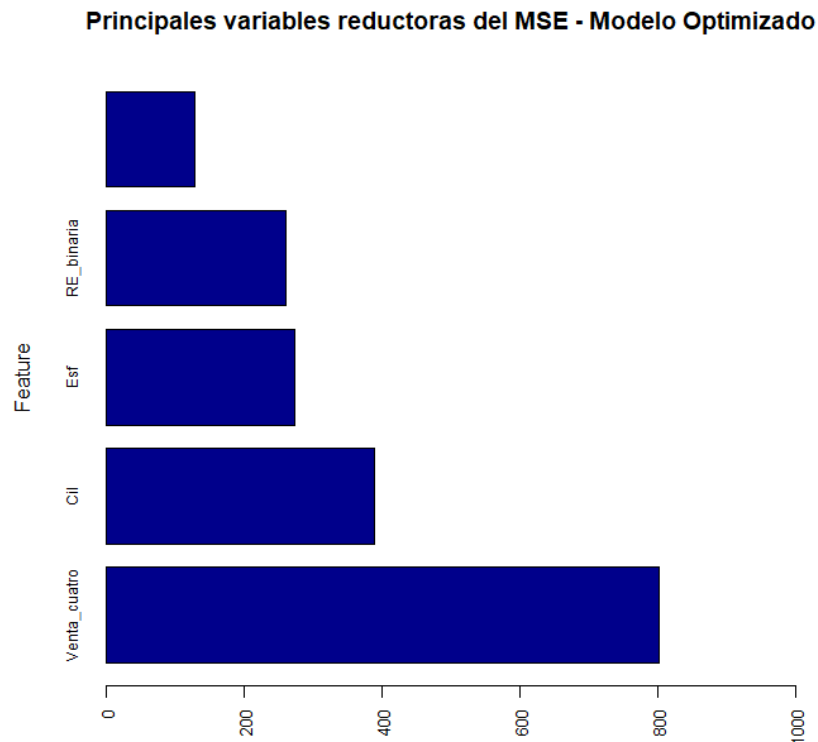
Optimización de Hiper-parámetros y principales variables

Con el objetivo de mejorar el poder predictivo del modelo, se definió un algoritmo para que pudiera iterar entre las características posibles que podrían tener los hiper-parámetros a definir dentro del modelo. Aquellas características sobre las que el usuario del modelo puede trabajar son el *maxnode*, *mtry* y *ntree*. El desarrollo de esta estrategia significa que se iteró con una cantidad finita de valores para estos tres hiper-parámetros dentro de un vector de posibilidades y se combinaron entre ellos para minimizar el *mse* (error cuadrático medio) del modelo. A continuación, se presentan los intervalos sobre los que se iteraron los distintos hiper-parámetros del algoritmo:

- Para establecer los intervalos para *maxnode* se definió un vector *c* con los siguientes valores a iterar: $c(5, 10, 25, 50, 75)$

- Para establecer los intervalos la cantidad de columnas del dataset (*mtry*) a utilizar se definió un vector *c* con los siguientes valores a iterar: *c*(3, 4, 5, 6)
- Para establecer los intervalos para *n tree* se definió un vector *c* con los siguientes valores a iterar: *c*(500, 1.000, 2.000, 2.500, 3.000)

Una vez ejecutado el algoritmo se obtuvieron las características bajo las cuales nuestro modelo debería tener una performance mejorada frente al set de validación. Finalmente, se obtuvo la importancia de las variables explicativas para el modelo optimizado ejecutado:



7.2.b Principales variables que explican la reducción de errores cuadráticos medios del modelo Random Forest luego de la optimización de hiper-parámetros

Como puede observarse en el gráfico, las principales variables predictoras no han cambiado entre las diversas variables existentes, sino que se ha distribuido su peso en la predicción de forma consistentemente distinta. Un ejemplo de ello es el de aquellas características de los SKU como el cilíndrico y el esférico que habían justificado una mayor proporción a la variabilidad del modelo. Al optimizar el algoritmo, estas variables han perdido el protagonismo frente al *feature* de la venta del mes anterior. Esta cantidad demandada del último mes utilizado ha sido la más determinante mientras que han quedado en niveles similares las características de las lentes como lo son el cilíndrico, esférico y su pertenencia a los rangos extendidos del stock.

Desempeño de la predicción general y puntual del modelo

A continuación, se presentan los resultados del modelo optimizado que ha sido ejecutado y su comparación contra los datos reales.

Random Forest

	Real	Pronóstico	Var.	Var. (%)
Cantidades demandadas	17.919	21.867	3.948	22%
Tallados por quiebres	-305	-13	292	-96%
Tallados/Total	-1,7%	-0,07%	1,6%	-96%

	Cantidad SKU	Unidades
SKU en exceso	796	7.112
SKU en defecto	245	-3.164
Total	1.041	3.948

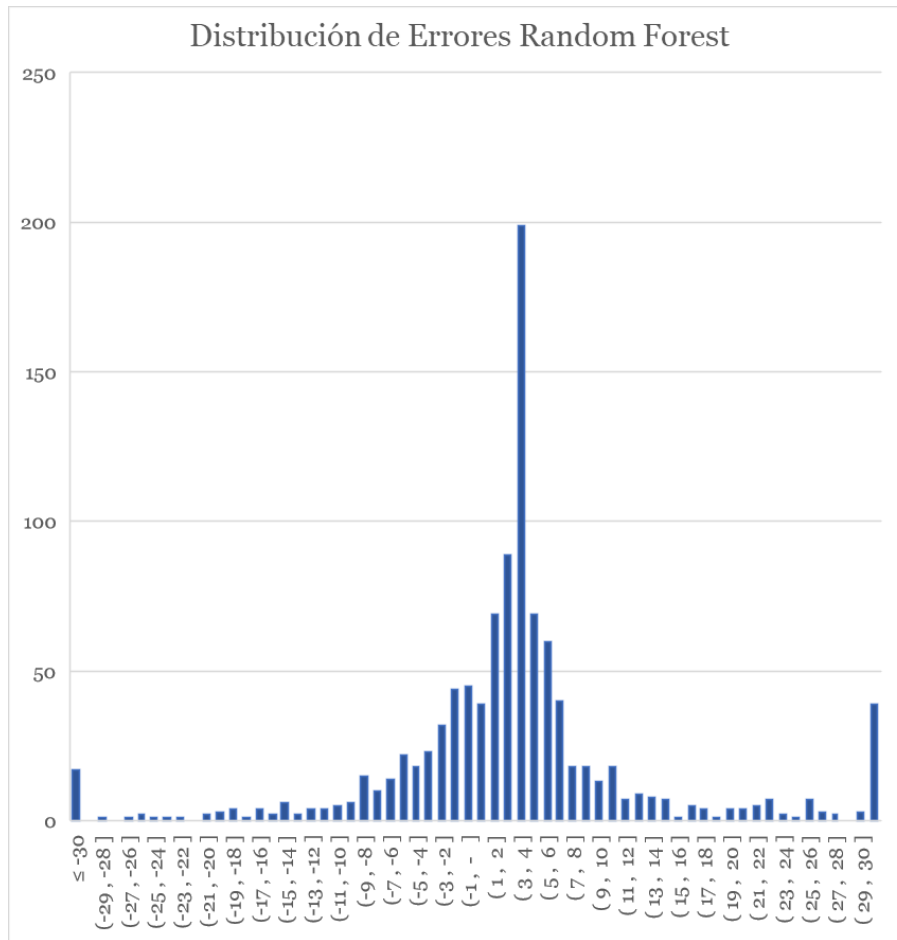
Al observar los desvíos del modelo frente al set de datos reales se puede apreciar que la performance de este modelo no ha sido buena en diversos aspectos. A priori, un exceso de 18% en la estimación de la demanda podría ser un parámetro razonable si no fuera a causa de que el índice de quiebres de stock ha sido superior al 17% del total de los SKU a estimar, esto significa que muchos artículos se han pronosticado con un alto nivel de exceso y, por otro lado, una cantidad significativa de ellos ha tenido errores superiores al defecto de -30 unidades explicando más del 50% de los pronósticos en defecto. El dato más alentador ha sido el ratio de tallados sobre la producción que ha mostrado una excelente performance como consecuencia de una predicción en exceso del 76% del total de SKU's. Otra consecuencia de este desvío al alza ha sido la eficiencia del modelo, que ha rondado el 80%, un número que no suena competitivo incluso bajo modelos básicos de estimación de demanda. Aun así, el modelo ha obtenido puntajes razonablemente comparables frente a las métricas de *machine learning*, sobre todo en el relacionado al error porcentual medio absoluto (sMAPE).

Random Forest

Métricas		Errores	
Porcentaje de desvío Q total	-18,05%	MSE	524,75
Quiebres del forecasting (no incluye stock)	-17,66%	RMSE	22,91
Producción por quiebres / Venta	-0,07%	MAE	9,89
Eficiencia costos vs lenta rotación (considera stock)	80,07%	sMAPE	89,25

Distribución de errores

Como era esperable luego del análisis de la tabla anterior, el hecho de tener un desvío de alrededor del 18% en exceso sobre la cantidad de ventas proyectadas, la acumulación de errores se centró en la parte positiva del histograma, no obstante, se puede observar un cierto desequilibrio en los extremos del histograma de errores del modelo claramente sesgados al exceso de demanda. Esto significa que si bien una parte central de los errores se encuentran dentro del intervalo de [-5 , 5] las altas cantidades ubicadas en los extremos anuncian una fuerte penalización en las métricas de rentabilidad económica.

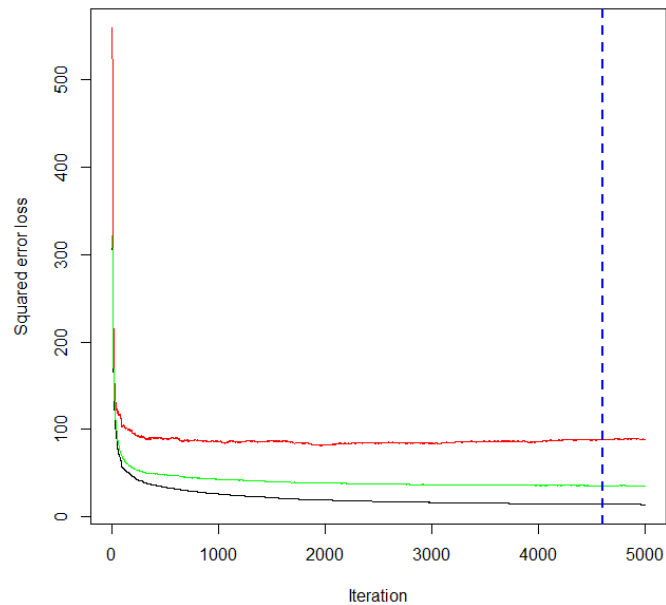


7.2.c Distribución de errores: Random Forest con escala de uno en uno y acotado a ± 30 observaciones.

7.3. Generalized Boosted Regression Modeling

Etapas iniciales del modelo

Para el procesamiento de este modelo se ha utilizado el mismo data set que el utilizado en el apartado de *Random Forest*. En un principio se utilizaron hiper-parámetros sugeridos por la bibliografía como lo son un tamaño muestral de 5.000 *n.trees*, un parámetro de aprendizaje $\lambda = 0,1$ y una profundidad del modelo de hasta 3 niveles. El modelo tuvo una buena performance con métricas de *machine learning* bajas y de negocios consideradas intermedias. Como respuesta a estos buenos resultados no se utilizó un algoritmo de búsqueda exhaustiva de ajuste de hiper-parámetros, sino que se exploró de forma iterativa la potencial mejora del modelo mediante una pequeña grilla exploratoria. Esto ha sido posible debido a la evaluación de los buenos resultados y del alto costo computacional y de tiempo que podría implicar en el ajuste del modelo sin obtener una mejora significativa del mismo. A continuación, se muestra la función de pérdida del error cuadrático medio del algoritmo inicial donde se puede observar una optimización del error cuadrático alrededor de las 4.600 iteraciones.



7.3.a Función de pérdida del MSE antes de optimización

Optimización de Hiper-parámetros

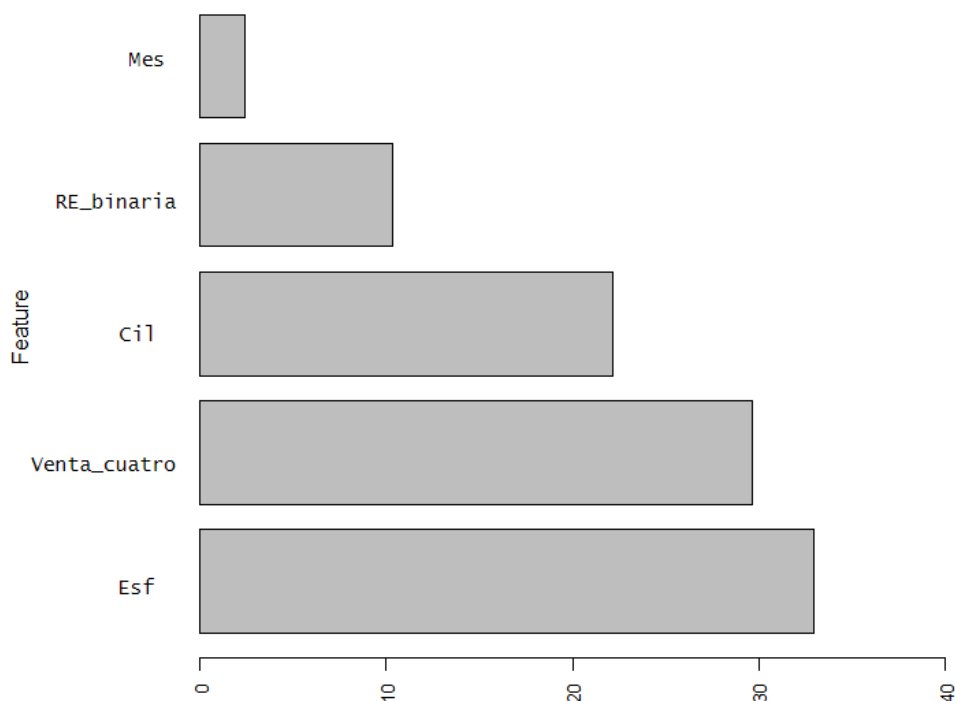
La optimización llevada a cabo para este algoritmo se basó en una iteración muy acotada, como se mencionó anteriormente, debido a la búsqueda de un equilibrio entre performance y costo computacional del modelo. En este sentido el modelo solo sufrió variaciones limitadas a dos hiper-parámetros en el orden del $\pm 50\%$ cada uno. Los mismos han sido la cantidad de muestras del modelo y la profundidad iterativa de este.

En la siguiente tabla se puede observar las combinaciones llevadas a cabo y resaltado aquella que resultó mejor de entre todas.

GBM	Valor Convencional			Mejor combinación
Muestras (<i>n.trees</i>)	2.500	5.000	7.500	
Profundidad (<i>Interaction depth</i>)	2	3	6	

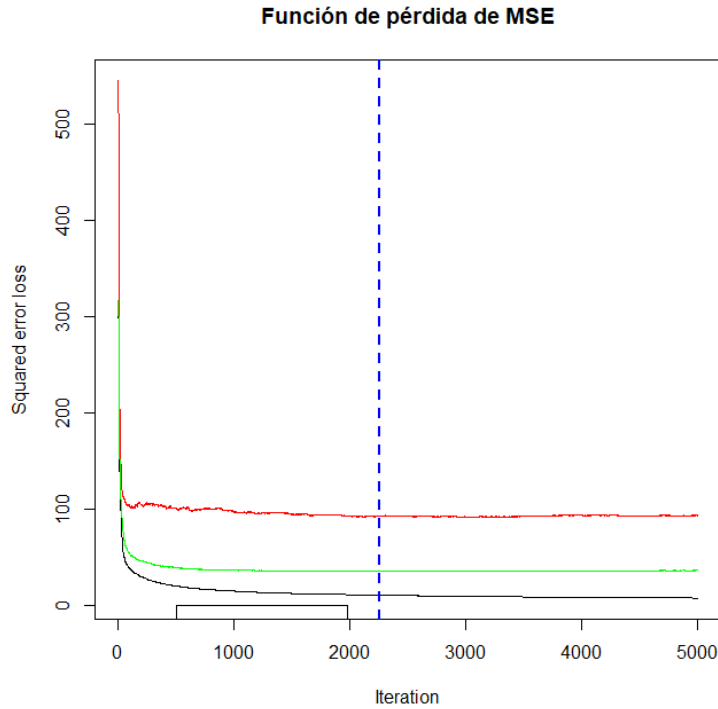
Una vez encontrado el mejor modelo ajustado, se procedió a generar una interpretabilidad para el modelo a través de aquellas variables que mayor injerencia habían tenido para reducir los errores del pronóstico.

Principales variables predictivas - Modelo Optimizado



7.3.b Principales variables que explican las predicciones del modelo basado en GBM

Como se puede observar, la tendencia de variables explicativas en modelos anteriores se mantiene dentro de las variables de esféricos y cilíndricos propios del SKU y la venta del último mes con datos reales. Por otro lado, como podrá observarse en la nueva función de pérdida del error cuadrático medio, la cantidad de muestras se reduce drásticamente al aumentar sensiblemente la cantidad de interacciones en profundidad del algoritmo. El dato que despierta cierto asombro es la aparición de la variable binaria que clasifica a los SKU del rango extendido de graduaciones, esto quiere decir que el modelo de *GBM* ha ponderado de manera significativa que una lente pueda, o no, pertenecer a esta categoría y, en función de ello, reducir el error de la predicción.



7.3.c Función de pérdida del MSE luego de la optimización

Desempeño de la predicción general y unitario

Luego de haber pasado por un limitado proceso de optimización, la mejora de resultados ha sido sensible y el modelo ha mostrado una buena adaptabilidad a la problemática. Es necesario destacar que el desempeño del modelo ha sido razonable ya sea desde las métricas estándar de *machine learning* como en variables de rentabilidad del negocio. Las cantidades demandadas, como pueden verse en las tablas contiguas, fueron pronosticadas en defecto en un -16% y la producción teórica de tallados, , ha estado por debajo del -0.3% del total, siendo unas 5 veces menos que la cifra real para el mismo período. Otro dato llamativo de las predicciones realizadas por el modelo GBM es que la cantidad de SKU en exceso fue superior a la de aquellos en defecto a pesar de que estos fueron en mayor cantidad. Este punto significa que es muy probable que, el modelo pueda ser mejorado con el correr de los meses y la estabilización de la demanda del producto en una tendencia creciente.

GBM

	Real	Pronóstico	Var.	Var. (%)
Cantidades demandadas	17.919	15.414	-2.505	-14%
Tallados por quiebres	-305	-46	259	-85%
Tallados/Total	-1,7%	-0,26%	1,4%	-85%

	Cantidad SKU	Unidades
SKU en exceso	649	2.973
SKU en defecto	392	-5.478
Total	1.041	-2.505

Distribución de errores

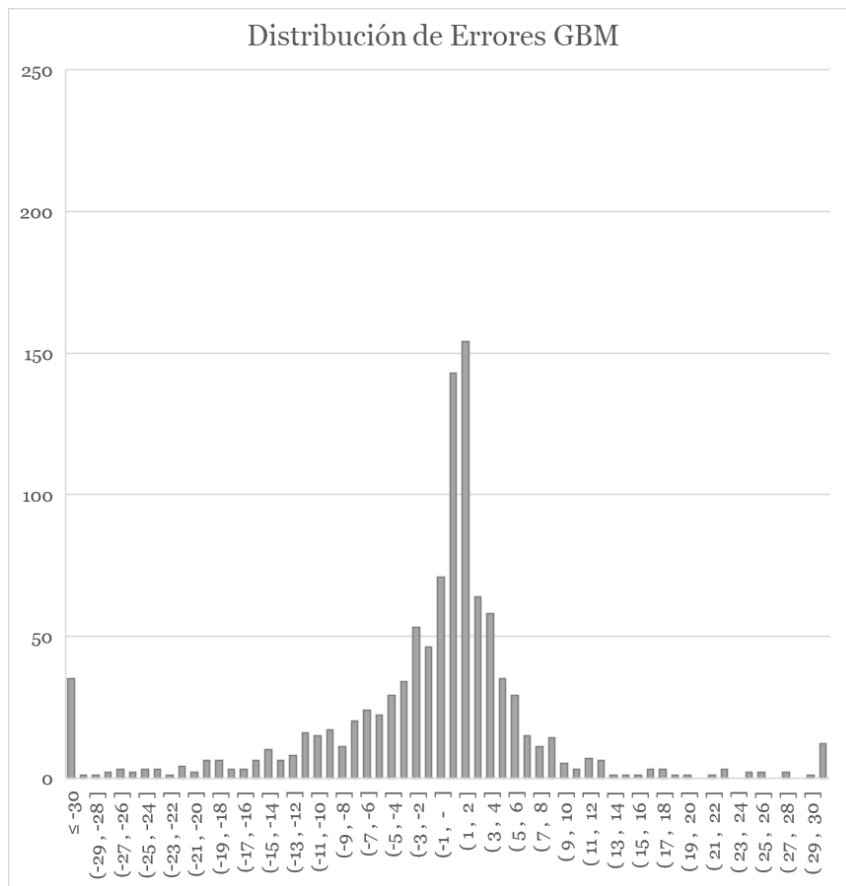
En la siguiente tabla se podrá observar las métricas obtenidas luego de la ejecución del modelo final que serán comparadas con los otros modelos en función de su performance de métricas de la metodología en sí como del negocio.

GBM

Métricas		Errores	
Porcentaje de desvío Q total	16,25%	MSE	442,83
Quiebres del forecasting (no incluye stock)	-30,57%	RMSE	21,04
Producción por quiebres / Venta	-0,26%	MAE	8,10
Eficiencia costos vs lenta rotación (considera stock)	91,36%	sMAPE	85,51

La performance del modelo demuestra métricas muy interesantes y razonables, con desvíos en métricas económicas aceptables y muy atractivos en cuanto a la eficiencia y la cantidad de tallados realizados.

Finalmente, puede observarse en el siguiente gráfico, la distribución de los errores de este modelo que se presenta es relativamente equilibrada, aunque puede observarse un desvío hacia los errores negativos, esto es explicado por una predicción en defecto. El error contra la cantidad total demandada sugiere cierta razonabilidad rondando un -1% en unidades totales. Por otro lado, es importante destacar que al encontrarse más de dos terceras partes de los desvíos dentro del rango de -5 a 5 lentes para cada SKU. También se puede interpretar que estamos ante un modelo equilibrado y que posiblemente ajuste mejor en función del paso del tiempo y cuyos días de stock parecieran estar en sintonía con una evolución sustentable del stock.



7.3.d Distribución de errores - Modelo GBM

7.4 Extreme Gradient Boosting (XGBoost)

Desarrollo del modelo

Para el desarrollo de este modelo se partió desde el inicio con una función iterativa llamada *random grid*. Esta forma de optimización de hiper-parámetros se basa en establecer para cada uno de ellos una cota máxima y otra mínima de los valores que estos podrían asumir. Una vez determinados estos puntos, el algoritmo ejecutará n modelos combinando de forma aleatoria entre cada uno de los intervalos seleccionados, hasta encontrar la mejor reducción del error entre esas n pruebas. A continuación, se presentan aquellos intervalos elaborados para el muestreo aleatorio del modelo cuyo objetivo será encontrar la mejor combinación que ajuste de mejor manera a la problemática planteada.

```
rgrid <- random_grid(size = 30,  
                    min_nrounds = 200, max_nrounds = 1500,  
                    min_max_depth = 5, max_max_depth = 8,  
                    min_eta = 0.0025, max_eta = 0.2,  
                    min_gamma = 0, max_gamma = 1,  
                    min_colsample_bytree = 0.6, max_colsample_bytree = 1,  
                    min_min_child_weight = 2, max_min_child_weight = 10,  
                    min_subsample = 0.75, max_subsample = 1)
```

7.4.a Extracto de código utilizado para la iteración de hiper-parámetros de la librería XGBoost.

Adicionalmente a esta configuración se utilizó como función objetivo una regresión clásica de minimización del error cuadrático medio. Se descartaron otras posibilidades de funciones objetivo como son una regresión logística o una regresión con una función logarítmica de pérdida.

Optimización de Hiper-parámetros

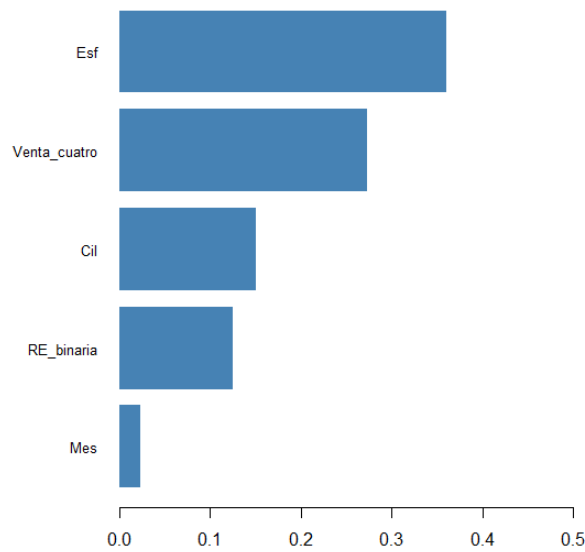
De los 30 modelos ejecutados, el de mejor performance resultó ser el compuesto por los siguientes valores para cada una de estas variables de regularización del modelo:

i	nrounds	max_depth	eta	gamma	colsample_bytree	min_child_weight	subsample
30	630	5	0.1909146	0.0641332	0.6356640	9.707238	0.7661778

Es necesario señalar que la cantidad de modelos fue una decisión arbitraria del usuario para lograr un equilibrio entre consumo de tiempo computacional y una exploración razonable de alternativas.

Posterior a la obtención de los hiper-parámetros óptimos, se procedió a evaluar la importancia de las diversas variables que han aportado información predictiva al modelo *XGBoost*. El aporte de las variables ha sido el mismo y en el mismo orden que el del modelo *GBM*, no obstante, como se verá a continuación, los resultados de la predicción son sensiblemente distintos entre ambos modelos.

Principales variables para el modelo XGBoost



7.3.b Principales variables que aportan a la reducción del error para el modelo final de XGBoost.

Desempeño de la predicción del modelo

La variación entre cantidades pronosticadas y reales fue en defecto de 5.814 unidades. Si bien el número parece ser la sentencia de un modelo con bajo nivel de efectividad, la cantidad de faltantes a la producción ha sido sensiblemente baja, significando una mejora del 55% contra el escenario real. Por otro lado, como podrá observarse en las tablas de resultados del modelo que se detallan a continuación:

XGBoost

	Real	Pronóstico	Var.	Var. (%)
Cantidades demandadas	17.919	12.100	-5.819	-32%
Tallados por quiebres	-305	-197	108	-35%
Tallados/Total	-1,7%	-1,10%	0,6%	-35%

	Cantidad SKU	Unidades
SKU en exceso	436	2.194
SKU en defecto	605	-8.008
Total	1.041	-5.814

A simple vista se puede observar que los errores de los SKU's han estado fuertemente orientados hacia el defecto de demanda de unidades. No obstante, la cantidad de artículos resultó ser mucho menos desbalanceada con 436 en exceso y 605 en defecto. Por último, la rentabilidad de la predicción se eleva a un 93% lo que significa que los stocks se encontrarían reforzados en aquellos artículos pronosticados en defecto.

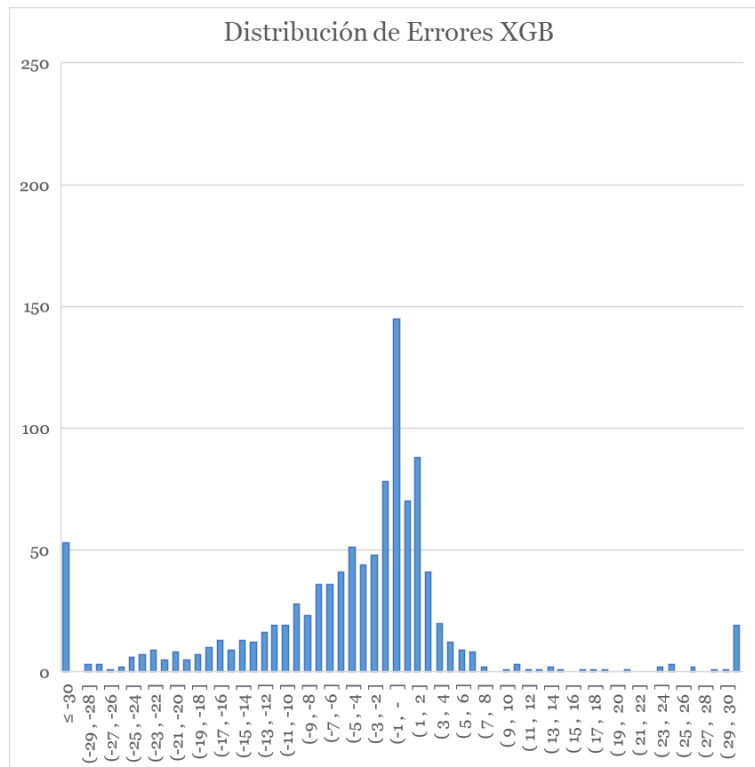
Luego de hacer un repaso por las métricas del negocio, se puede observar que las métricas de estadísticas o de *machine learning* no han logrado resultados sobresalientes, y en términos económicos se los puede catalogar como erráticos, como se puede ver en la tabla contigua:

Métricas		Errores	
Porcentaje de desvío Q total	48,09%	MSE	524,69
Quiebres del forecasting (no incluye stock)	-44,69%	RMSE	22,91
Producción por quiebres / Venta	-1,10%	MAE	9,80
Eficiencia costos vs lenta rotación (considera stock)	93,33%	sMAPE	104,85

Ejemplo de estos resultados erráticos son la eficiencia económica del modelo, que posee una tasa muy baja de tallados y con un muy buen nivel de eficiencia económica, que debería funcionar como un contrapeso al sobre aprovisionamiento de stock. Esto puede interpretarse como un consumo de los días de stock del inventario de la compañía, es decir de su capital de trabajo. Los desvíos en cantidades por encima del 40% estarían indicando que el modelo no ha incorporado correctamente los patrones principales de la demanda del producto.

Distribución de errores

En el siguiente apartado se realizará un análisis de la distribución de los errores del modelo en base a la gráfica 7.4.c presentada de forma contigua. Como se puede observar, existe una alta concentración de errores en el intervalo comprendido entre $[-6, 4]$ unidades de desvío. Esta cantidad representa más del 60% de los SKU bajo análisis y podrían ser considerados como razonables. Sin embargo. Existe una alta cantidad de desvíos para valores inferiores a -6 unidades sobre las que el modelo parece mostrar debilidad. Sin el objetivo de buscar una conclusión definitiva sobre este modelo, se puede avizorar cierta detección de patrones relevantes y posiblemente posean un desvío de forma estructural que hace que este modelo resulte ambiguo en cuanto a resultados de metodología como de performance del negocio.



7.4.c Distribución de los errores del modelo XGBoost

8. Elección del modelo en base a los resultados obtenidos.

Durante el siguiente apartado se abordarán de manera integral las experiencias recabadas durante el proceso de ejecución de todos los modelos y serán comparados en función de las métricas del negocio, errores y las ventajas y desventajas de su procesamiento.

En segundo lugar, se establecerá un método de evaluación de los modelos con el objetivo de establecer cierta imparcialidad en la elección de un modelo sobre otro.

8.1 Performance de los modelos bajo métricas de *Machine Learning*:

En la tabla presentada a continuación se exhiben los resultados provenientes de la evaluación de los pronósticos en función de aquellas métricas estadísticas definidas durante el apartado 6 del presente trabajo:

Evaluación de errores	Random Forest	GBM	XGBoost	ETS
MSE	524,75	442,83	524,69	474,14
RMSE	22,907	21,04	22,906	21,77
MAE	9,89	8,10	9,80	8,13
sMAPE	89,25	85,51	104,85	87,36

Como se puede observar, el método de *GBM* ha obtenido los mejores rendimientos, aunque no siempre con una amplia ventaja. Si bien han existido métricas donde las diferencias no podrían considerarse estadísticamente significativas (como el MAE y el sMAPE) no parece haber cuestionamientos desde este punto de vista que el modelo *GBM* ha sido el que se ha destacado por sobre los demás.

Por otro lado, el método de *ETS* ha demostrado tener un nivel de errores muy competitivos a pesar de tratarse de un modelo mucho menos complejo computacionalmente. Si se evaluara únicamente en función de los errores, el mismo se ubicaría en segundo lugar en todas las métricas excepto en sMAPE detrás de *Random Forest*. La performance obtenida por este método a nivel de errores indica que a pesar de no ser el mejor en cuanto a desempeño, sería totalmente válido en casos donde no se puede desarrollar soluciones que precisen mucha demanda de cómputo, de tiempo o cuyo beneficio económico justifique el desarrollo a medida de un algoritmo complejo y trabajoso. Esto explica que esta metodología sea tan utilizada por su excelente relación entre el tiempo de ejecución y su poder predictivo.

Los modelos basados en las metodologías *Boosting* han demostrado ser herramientas muy útiles y capaces de encontrar patrones complejos dentro de grandes cantidades de información que no se podrían descubrir fácilmente si un usuario se propusiera hacerlo. Sin embargo, dentro de estas dos herramientas la que ha demostrado ser mucho más eficaz al momento es *GBM* y que *XGBoost* no ha podido mostrar su potencial en esta ocasión a pesar de haber sido el algoritmo al que más tiempo computacional consumió. No obstante, si bien este tipo de modelos muestran tener un buen poder predictivo es importante tener en cuenta que la incorporación de cambios a los modelos desarrollados para aplicar mejoras o ampliaciones serían mucho más difíciles de implementar y eso podría desalentar su uso si los resultados no fueran sensiblemente mejores.

8.2 Performance de los modelos bajo métricas del Negocio de lentes oftálmicas:

En la siguiente tabla se presentan los resultados de las cuatro metodologías exploradas en función de las métricas del negocio:

Métricas del Negocio	Random Forest	GBM	XGBoost	ETS
Porcentaje de desvío Q total	-18,05%	16,25%	48,09%	51,29%
Quiebres del forecasting (no incluye stock)	-17,66%	-30,57%	-44,69%	-40,55%
Producción por quiebres / Venta	-0,07%	-0,26%	-1,10%	-0,46%
Eficiencia costos vs lenta rotación (considera stock)	80,07%	91,36%	93,33%	95,98%

Si se analiza la primera métrica que aborda la predicción de cantidades totales demandadas, el modelo GBM ha tenido el menor de los desvíos en defecto. No obstante, el método ha quedado relegado en el resto de las variables. *Random Forest* ha sido el único modelo que ha predicho una cantidad superior a la demanda real y en función de ello existió una penalidad más alta en la performance de rentabilidad.

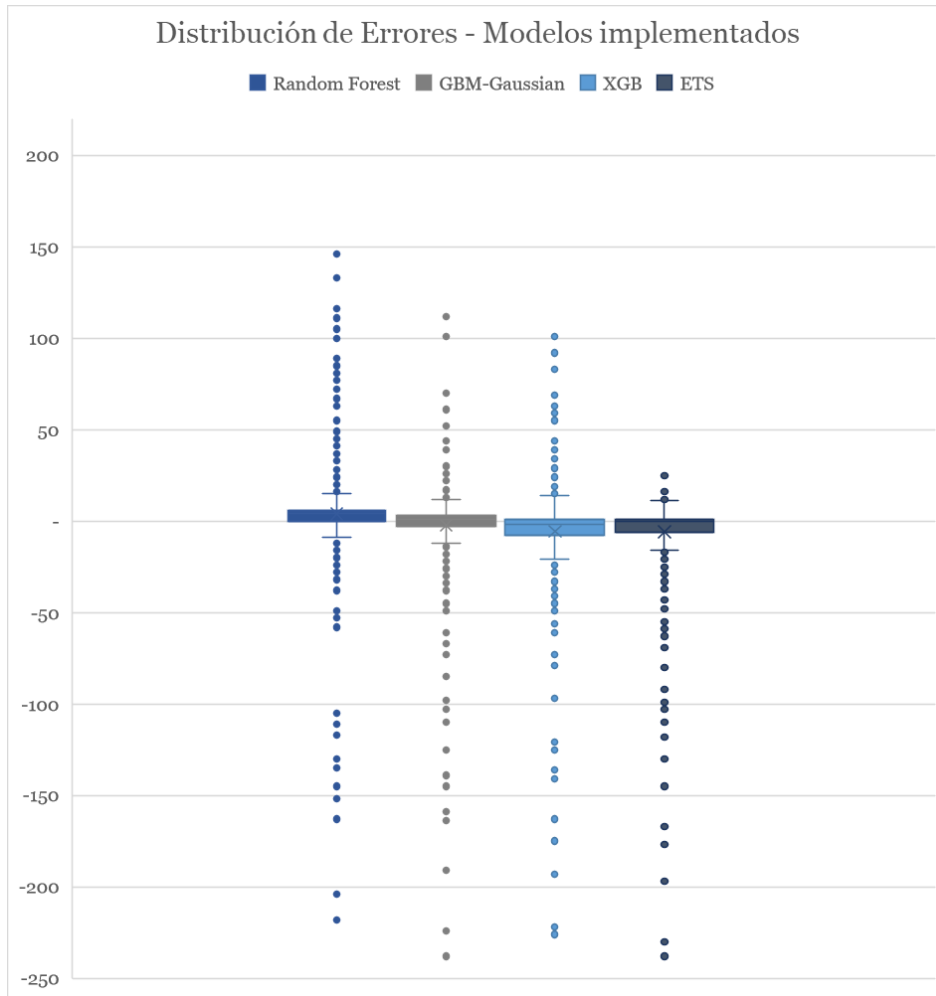
Cuando se analizan los desvíos en defecto a nivel SKU los modelos ETS y *XGBoost* han tenido similares rendimientos, aunque el modelo ETS ha resultado mejor. Esta métrica es de vital importancia para observar si el modelo ha captado el patrón de minimización de las cantidades para lograr no consumir el capital de trabajo del stock acumulado. Dicho en otras palabras, si un modelo capta con mayor facilidad las demandas en defecto por sobre las que son en exceso, la realización del cálculo de reposición de stock tenderá a compensar al modelo sin incurrir en mayor cantidad de tallados o mantener la lente en stock durante un tiempo excesivo.

Al poner foco en la estimación de las producciones por quiebres de stock se observa que el modelo *XGBoost* ha vuelto a tener la mejor performance y reafirma que es el modelo que mejor ha aprovechado la distribución de cantidades entre los SKU, aunque puede verse que ha estado bastante sesgado a predecir en defecto a aquellos artículos de mayor movimiento.

Por último, cuando se realiza el análisis de la rentabilidad, se puede observar que en contraposición a los modelos que ofrecieron predicciones en defecto de cantidades se ven beneficiados en este cálculo principalmente gracias al ahorro que ha implicado el cálculo de la lenta rotación. Esto también significa que el modelo ha consumido capital de trabajo establecido y que en los períodos futuros las reposiciones deberán ser mucho mayores en comparación a los otros dos modelos, implicando un cierto riesgo a futuro para la calidad de los modelos que predijeron en defecto. Por el contrario, en el caso de los sobre pronósticos, la penalización de lenta rotación afecta significativamente la rentabilidad, pero se reducen de forma drástica el tallado de productos faltantes. Finalmente, es importante destacar que el rendimiento del modelo *GBM* ha resultado mucho más equilibrado al relacionar la eficiencia económica con el desvío de la demanda total del mes y esto es lo que lo ha llevado a ser el modelo que mejor ha ponderado los contrapesos de las métricas sin consumir en exceso el capital de trabajo de la compañía.

8.3 Distribución de los errores de los modelos presentados:

El siguiente gráfico contiene la distribución de errores de manera consolidada para los cuatro modelos analizados durante el presente trabajo:



8.1.c.1 Gráfico de distribución de errores consolidado para los modelos ejecutados.

Como puede observarse en el caso de *Random Forest*, la dispersión de los errores está mucho más equilibrada en función de diferencias en exceso y en defecto, aunque al mismo tiempo poseen una amplitud mucho mayor a la del resto de los métodos.

También puede observarse el caso del algoritmo ETS tiene muy pocas observaciones con errores positivos. Este sería otro argumento para explicar el desvío en defecto de la demanda total en unidades.

Los errores de las metodologías *Boosting* poseen una estructura bastante similar, aunque en el caso del GBM la tendencia de concentración del bigote es menor y levemente orientada hacia una distribución de errores positivos

8.4 Elección de modelo bajo las diversas métricas

Para la elección del método más apropiado que haya reflejado de mejor manera el mes de validación utilizado, se ha llevado a cabo una ponderación del rendimiento de cada modelo en función de cada una de las métricas abarcadas durante el presente trabajo.

Cada performance de los modelos ha sido evaluada bajo un puntaje de 1 a 4 según su rendimiento en comparación al de los demás modelos. Esto implica que el peor modelo puntuaría con un solo punto contra los cuatro que obtendría aquel con mejor puntaje.

Es necesario aclarar que esta forma de evaluación ha sido definida de manera arbitraria por el usuario con el objetivo de obtener una evaluación cuantitativa y metódica de los resultados de los modelos para que exista una forma de comprobación de la elección en función de todas las métricas preestablecidas. Esto no quiere decir que éste sea el único método válido de elección del modelo y que la compañía podría priorizar entregarle más peso a alguna métrica por sobre otras. Sin desmedro de esto, es importante aclarar que la mayoría de los modelos han arrojado resultados razonables y que con la ampliación de los universos de datos y con la optimización de sus características, todos serían razonablemente eficientes para abordar la problemática planteada.

Métrica	Random Forest	GBM	XGBoost	ETS
MSE	1	4	2	3
RMSE	2	4	1	3
MAE	1	4	2	3
sMAPE	3	4	1	2
Error/Método	1	2	3	4
Procentaje de desvío Q total	3	4	2	1
Quiebres del forecasting (no incluye stock)	4	3	1	2
Producción por quiebres / Venta	4	3	1	2
Eficiencia costos vs lenta rotación (considera stock)	1	2	3	4
Totales	20	30	16	24

Tabla de ponderaciones de los modelos bajo las diversas métricas propuestas

De las ponderaciones presentadas en la tabla anterior se puede observar que el modelo con mejor puntaje ha sido el *GBM*; no obstante, ha sido seguido por el modelo *ETS*. El modelo *GBM*, como se mencionó anteriormente, parece ser más equilibrado en métricas de negocio y ha tenido una excelente performance a nivel de la evaluación de sus errores. Sorprendentemente, el modelo *XGBoost* ha tenido un desempeño inverso a pesar de tratarse de algoritmos bastante similares al de *GBM*.

Finalmente, a través de este método definido se refuerza la selección del modelo *GBM* como el método con mejor performance frente al resto de sus competidores incluyendo, excelentes métricas de errores y un comportamiento equilibrado en variables del negocio.

9. Conclusiones y recomendaciones de Negocio

Una de las principales cuestiones arribadas durante el inicio de este trabajo fue el manejo de aquellas situaciones inesperadas bajo las que los stocks o sus costos se movieran de una manera inesperada y, aun así, pudiesen ser previstas. Dentro de las experiencias relevadas durante el presente trabajo, la conformación de stocks iniciales por parte de clientes mayoristas (a los que se consideran poco usuales luego de un lanzamiento), afectarían de una forma significativa al equilibrio entre los planes de la compañía y el surtido de su stock. Bajo este contexto la primera propuesta para lograr una mejor eficiencia del negocio es la consideración de una venta por encargo a las fábricas de estos stocks iniciales con una bonificación por parte de la compañía, en función de la estimación de ahorro que tendrá en quiebres de stock y otras pérdidas asociadas, ya que este riesgo se estaría exteriorizando. La penalización de pedidos mayoristas no puede ser considerada como una buena práctica comercial por lo que no debe ser considerada como recomendación de negocio.

Otra recomendación operativa luego del relevamiento realizado es la sugerencia de unificación, en caso de ser posible, de los *web order* mayoristas y minoristas de la compañía y con ello la homogenización de la codificación de cada producto. Si bien este cambio en los procesos de la compañía ya no resultaría necesario dada la matriz de equivalencias permitiría que la compañía procese los pronósticos de manera más rápida y eficiente.

Como se ha observado durante el relevamiento de la cadena de suministro de los rangos regulares y los extendidos, los costos FOB de los productos han estado diferenciados desde un principio. Mas allá de que se considere como un error sensible el de establecer durante el lanzamiento el mismo precio de lista para ambos tipos de rangos, la traslación proporcional de estos costos debería ser evaluada en función de la inmovilización de capital que le insume a la compañía importadora en función de su propia evaluación de la demanda. Esto podría cubrir a la compañía de pérdidas adicionales o, en caso de una evaluación más favorable de la rotación de algunos SKU, favorecer la agilidad del movimiento del producto.

Otro dato relevante que ha destacado sobre la evaluación de los pronósticos realizados fue el hecho de que cuando el nivel de días de stock se encontraba por encima de los 90 días, el porcentaje de producciones locales de la venta descendía de forma significativa frente a otros meses, como lo fue el caso de marzo 2022. Esto implica que posiblemente los días de stock se encuentren subestimados para una mejor provisión. Incluso podría proponerse un mayor nivel de días de stock para rangos extendidos frente a los rangos regulares, sin un impacto significativo en capital inmovilizado. Este último punto tiene un segundo nivel de impacto que es el de los embarques óptimos. Como se ha visto en el relevamiento de la cadena de suministro, los mejores costos unitarios se comienzan a acelerar cuando se superan las 30.000 unidades como consecuencia del descuento del precio FOB y de la optimización de costos de importación asociados. En este sentido, si la compañía pudiese evaluar el costo del capital, ya sea financiero como de utilización del espacio, con el objetivo de aumentar en 30 días adicionales sus stocks por SKU y explorar la posibilidad de realizar compras bimestrales mayores a 30.000 unidades. Esta resultaría una buena alternativa para que la empresa se encontrase mejor protegida ante los armado de stocks iniciales por parte de la clientela mayorista.

9.1 Propuestas de mejoras al trabajo realizado

El primer punto de mejora para el modelo es que el mismo podría ampliarse no solo a nivel producto sino también se podría utilizar de manera conjunta para completar embarques del mismo proveedor. El estudio de esta modalidad de oferta de rangos extendidos podría ser el puntapié inicial para que la compañía los comercialice en otro tipo de productos como podrían ser las gamas premium que se comercializan en el país.

Una mejora sobre la que se podría trabajar es la aplicación de esta metodología sobre el cálculo de los productos semielaborados. En caso de la producción de lentes por quiebres de stock disminuyan drásticamente podría conllevar a la realización de ajustes a la política de abastecimiento de los *blocks*.

En futuras iteraciones también podrá ser viable considerar agregar nuevas variables al dataset. Una alternativa de ello podría ser la regionalización de la demanda, por ejemplo, las regiones centro, noroeste, noreste, sur y la zona metropolitana del país con el objetivo de reducir el tiempo de entrega a clientes alejados del depósito general estableciendo stocks en depósitos intermedios del grupo económico.

Otra alternativa podría buscar definir la segmentación de la clientela entre mayorista o minorista. Un objetivo un poco más ambicioso con el correr del tiempo podría ser la predicción de los 4 meses de *leadtime* acumulados, este reto, a priori, parece que requeriría de varias iteraciones y acumulación de datos para poder ser llevado a cabo.

Adicionalmente, las técnicas de análisis de datos que se aplicaron podrían aplicarse en mayor medida y crear nuevas variables sobre los mismos datos originales. En el futuro la creación de una mayor cantidad de variables binarias como lo fue la del rango extendido sería muy interesante para experimentar. Ejemplo de esta sugerencia podría ser: si sus cilindros son positivos o no; negativos o no; neutros o no y de igual forma para los esféricos.

También en el aspecto de aquellas variables exógenas a utilizar como el caso de la inflación del mes o de los índices de confianza del consumidor la compañía podría hacer un relevamiento más sofisticado para incorporar al set de datos, como por ejemplo un estudio de mercado de largo plazo para el nicho de mercado al que pertenece.

Es importante que cuando el producto supere los dos años de su lanzamiento se considere la existencia de la estacionalidad en la demanda de las lentes. Si bien esto no fue posible durante el transcurso del presente trabajo dado que la idea de generar algoritmos que aprendan rápidamente los patrones de la demanda ni bien se lanzó el producto está claro que cualquiera de estas técnicas que se decida aplicar deberá contemplar la aparición de este factor que aquí no ha podido verificar su relevancia.

Finalmente, es necesario destacar que, si bien la evaluación de los modelos se ha hecho en base a un solo período y que, a pesar de que la tendencia se estima será similar en el futuro, debería contemplarse seguir aplicando esta técnica tal como se enunció en el abordaje teórico de la metodología. Esto tiene sustento en que el desarrollo y el hundimiento de la inversión en cuanto a la mayor cantidad del tiempo ya ha sido realizada, los algoritmos ya se han optimizado y en función de ello, resulta lógico que se mantengan bajo aplicación de estos métodos por algunos períodos más para descartar la posibilidad de hechos no contemplados en el modelo.

Posiblemente sea una buena idea realizar una evaluación conjunta para racionalizar el desempeño de los modelos, como lo pueden ser promedios entre métodos, o la utilización de las cotas del modelo ETS para complementar pedidos de embarques. La aplicabilidad

de una continuidad de los modelos requeriría un esfuerzo relativamente pequeño y la actualización de los escenarios se acercaría mucho más a la idea planteada de utilizar los *multi-step forecasts* propuestos por R. Hyndman, que sugieren ser un camino muy interesante para el desarrollo de la técnica del forecasting más allá del modelo que se utilice.

Referencias

-
- ⁱ Hyndman, Rob J and Athanasopoulos, George. Forecasting: Principles and Practice (2nd ed). April 2018. <https://otexts.com/fpp2/>
- ⁱⁱ Pavlyshenko, B.M. (2019) “Machine – Learning Models for Sales Time Series Forecasting”, Data, 18 de Enero 2019, Vol. 4, N° 15.
- ⁱⁱⁱ Instituto Nacional de Estadísticas y Censos. Índice de precios al consumidor. Marzo 2022. <https://www.indec.gov.ar/indec/web/Nivel4-Tema-3-5-31>
- ^{iv} Banco de la Nación Argentina. Cotizaciones de las principales divisas. <https://www.bna.com.ar/Personas>
- ^v Índice de confianza al consumidor. Marzo 2022. Universidad Torcuato Di Tella. <https://www.utdt.edu/download.php?fname=164873132113817000.pdf>
- ^{vi} Banco Central de la República Argentina. Relevamiento de Expectativas de Mercado (REM) http://www.bcra.gov.ar/PublicacionesEstadisticas/Relevamiento_Expectativas_de_Mercado.asp
- ^{vii} Gareth, J., Witten, D., Hastie, T., Tibshirani, R. (2013). An Introduction to Statistical Learning with Applications in R. New York: Springer. pp. [176-178] .
- ^{viii} A Note on the Validity of Cross-Validation for Evaluating Autoregressive Time Series Prediction Christoph Bergmeira , Rob J Hyndmanb, Bonsoo Koo.
- ^{ix} Lopez de Prado, Marcos. Advances in Financial Machine Learning. Wiley, 2018.
- ^x Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. International Journal of Forecasting, 22(4), 679–688.
- ^{xi} Improved supply chain management based on hybrid demand forecasts. Luis Aburto, Richard Weber. https://www.researchgate.net/publication/222568533_Improved_supply_chain_management_based_on_hybrid_demand_forecasts
- ^{xii} Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: The state space approach*. Berlin: Springer-Verlag.
- ^{xiii} Breiman, L. (2001). Random forests, Machine Learning 45: 5–32.
- ^{xiv} Gareth, J., Witten, D., Hastie, T., Tibshirani, R. (2013). An Introduction to Statistical Learning with Applications in R. New York: Springer. Pp.[312-316]
- ^{xv} Gareth, J., Witten, D., Hastie, T., Tibshirani, R. (2013). An Introduction to Statistical Learning with Applications in R. New York: Springer. Pp.[321-323]
- ^{xvi} Gareth, J., Witten, D., Hastie, T., Tibshirani, R. (2013). An Introduction to Statistical Learning with Applications in R. New York: Springer. pp. 323 .
- ^{xvii} <https://xgboost.readthedocs.io/en/latest/tutorials/model.html>
- ^{xviii} Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance Cort J. Willmott, Kenji Matsuura Center for Climatic Research, Department of Geography, University of Delaware. Newark, Delaware 19716, USA.
- ^{xix} Chicco D, Warrens MJ, Jurman G. 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science. <https://doi.org/10.7717/peerj-cs.623>

^{xx} IFRS Foundation, IAS 2 Inventories, 2003.
<https://www.ifrs.org/content/dam/ifrs/publications/pdf-standards/english/2022/issued/part-a/ias-2-inventories.pdf>

^{xxi} Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning. 12th printing, 2017. Springer p 593.