



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

# Desarrollo de un modelo predictivo para una cadena de suplementos alimenticios

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

PABLO JAVIER ARAYA REBOLLEDO

PROFESORA GUÍA:  
ALEJANDRA PUENTE CHANDÍA

MIEMBROS DE LA COMISIÓN:  
CAROLINA SEGOVIA RIQUELME

BLAS DUARTE ALLEUY

SANTIAGO DE CHILE

2022

## Estimación de demanda para una cadena de suplementos alimenticios: Redes neuronales recurrentes y modelos clásicos de series de tiempo

El proceso de estimación de demanda es una tarea transversal en cualquier empresa que ofrece distintos bienes a sus usuarios. Esta labor se torna aún más relevante cuando los productos que se comercializan son alimentos procesados con fecha de caducidad, por lo que, una estimación más precisa no solo se traducirá en la disminución de ventas perdidas, sino que también la baja en perdidas debido a mermas de productos.

La información que se utiliza en el presente estudio corresponde a las boletas de 5 puntos de ventas de una cadena que comercializa suplementos alimenticios y productos para el bienestar. Los datos abarcan el periodo comprendido entre enero de 2015 y diciembre de 2020, sin embargo, se utilizará la información hasta marzo del último año debido a que las restricciones al comercio posterior a dicha fecha distorsionan la demanda real a la que está sometida un determinado punto de venta.

El objetivo de la presente investigación es desarrollar distintos modelos que sean capaces de predecir la demanda a nivel mensual, para 6 y 20 categorías con niveles de agregación distintos, para cada uno de los 5 locales. Además de lo anterior, cada modelo debe ser capaz de proyectar la demanda a un futuro de 12 meses. Se utilizan los modelos de suavización exponencial y ARIMA, además de 4 configuraciones distintas de redes neuronales recurrentes.

La etapa de elección de los mejores modelos se basa en la versión normalizada de la métrica RMSE (NRMSE) y MAPE, donde, además de comparar para cada serie cual es el modelo con mejor resultado, también se realizan los test-t Student's y Wilcoxon para testear la significancia en la diferencia de la distribución de los errores de cada uno de los 8 modelos de manera global.

En base a los resultados, y las métricas NRMSE y MAPE, no existe un modelo claramente mejor en términos globales, sin embargo, para aquellos locales o categorías que presenten una mayor varianza, es preferible utilizar redes neuronales recurrentes, por el contrario, los modelos de suavización exponencial simple y ARIMA se comportan mejor en series estables.

# Tabla de contenido

Capítulo I – Introducción.....	1
I.1 Antecedentes generales.....	1
I.1.1 Marco Regulatorio.....	1
I.1.2 Puntos de venta.....	3
I.2 Justificación.....	6
I.3 Objetivos.....	11
Capítulo II – Alcances y resultados esperados.....	12
II.1 Alcances.....	12
II.2 Resultados esperados.....	13
Capítulo III – Marco teórico.....	14
III.1 Conceptos de series de tiempo.....	14
III.1.1 Definiciones (tendencia, estacionalidad, ciclos y errores).....	14
III.1.2 Proceso estocástico y estacionariedad.....	15
III.1.3 Autocovarianza, autocorrelación y autocorrelación parcial.....	16
III.2 Modelos clásicos de series de tiempo.....	17
III.2.1 Suavización exponencial simple.....	17
III.2.2 Modelo de Holt’s.....	19
III.2.3 Modelo de Holt/Winters.....	20
III.2.4 ARIMA (Autoregressive integrated moving average).....	21
III.2.4.1 Autoregresion (AR).....	21
III.2.4.2 Medias móviles (MA).....	22
III.2.4.3 Integración (I).....	22
III.2.4.4 Autoregressive integrated moving average (ARIMA).....	23
III.3 RNN (Recurrent neural networks).....	24
III.3.1 Algoritmos supervisados, no supervisados y de reforzamiento.....	24
III.3.2 RNN: Arquitectura y optimización.....	25
III.3.2.1 Forma de una RNN.....	25

III.3.2.2 Ecuaciones y unidad neuronal .....	27
III.3.3 Aprendizaje y optimización .....	28
III.3.3.1 Función de perdida.....	28
III.3.3.2 Gradiente descendente.....	29
III.3.3.3 Gradiente descendente estocástica .....	31
III.3.3.4 Backpropagation.....	32
III.3.4 Hiperparametros .....	35
III.4 Preprocesamiento de una variable .....	37
III.4.1 Normalización .....	37
III.4.1 Estandarización .....	37
III.5 Métricas de error.....	38
III.5.1 Error cuadrático medio (Mean squared error, MSE).....	38
III.5.2 Error absoluto medio (Mean absolute error, MAE) .....	38
III.5.3 Raíz del error cuadrático medio (Root mean squared error, RMSE) .....	39
III.5.4 Raíz del error cuadrático medio normalizada (NRMSE).....	40
Capítulo IV – Datos.....	41
IV.1 Descripción de datos disponibles .....	41
IV.2 Limpieza y preprocesamiento de datos.....	49
Capítulo V – Metodología .....	52
V.1 Modelos clásicos de series de tiempo.....	52
V.1.1 Suavización exponencial simple .....	54
V.1.2 Suavización exponencial doble (Holt’s).....	56
V.1.3 Suavización exponencial triple (Holt’s/Winters) .....	58
V.1.4 ARIMA .....	59
V.2 Redes neuronales recurrentes .....	62
V.2.1 RNN-12 Time steps, one layer (RNN-12-1).....	66
V.2.2 RNN-24 Time steps, one layer (RNN-24-1).....	66
V.2.3 RNN-12 Time steps, two layers (RNN-12-2).....	67

V.2.4 RNN-24 Time steps, two layers (RNN-24-2).....	68
Capítulo VI – Resultados y análisis .....	69
VI.1 Distribución de los errores .....	69
VI.2 Test-t Student’s pareado .....	78
VI.3 Test de Wilcoxon signed rank .....	83
VI.4 Análisis mensual .....	85
VI.5 Conclusiones y futuros trabajos.....	89
Bibliografía.....	92
Anexos.....	94
Anexo A-Descripción de Variables .....	94
Anexo B-Resultados y análisis.....	99

# Capítulo I – Introducción

## I.1 Antecedentes generales

### I.1.1 Marco Regulatorio

En Chile la industria de los suplementos alimenticios y nutricionales esta representada por la Asociación de Productores e Importadores de Alimentos y Suplementos Saludables A.G (ALIMSA A. G). El gremio está compuesto por 7 actores relevantes del mercado, los cuales son: Nutrapharm, Nutrmarket, Nutraline, New Science, Maver, Medcell y Omnilife.

Según datos de ALIMSA<sup>1</sup>, la industria genera alrededor de US\$ 200 millones al año, representando cerca del 20% de la venta de alimentos y bebidas procesadas en Chile. Junto con lo anterior, la agrupación estima que el rubro genera más de 8.000 empleos en territorio nacional. Por otro lado, según Euromonitor International, en Chile, el consumo de productos relacionados con el control de peso se duplicó desde 2011 a 2016, alcanzando los \$53.544 millones de pesos en ventas este último año. A su vez, la misma fuente señala que los productos de nutrición deportiva tuvieron el mismo crecimiento en dicho periodo, alcanzando los \$13.662 millones en 2016. Las proyecciones de Euromonitor para estas dos categorías en 2021, es de \$66.000 millones para productos relacionados con la obesidad, y \$21.000 millones para nutrición deportiva.

La demanda por productos saludables en Chile se encuentra en crecimiento, debido en parte al aumento de la conciencia de la salud, una exigencia de productos más sofisticados y un incremento en la disponibilidad de mejores ingresos en los estratos medios y altos. Estos

---

<sup>1</sup> ALIMSA A. G, s.f. Los suplementos alimentarios. Recuperado de sitio web [www.camara.cl](http://www.camara.cl)

últimos, se encuentran en búsqueda constante de nutrición deportiva, comidas de control de peso, vitaminas y suplementos de dietas<sup>2</sup>.

El marco regulatorio que rige la importación y elaboración de suplementos alimenticios en Chile está supeditado al Código Sanitario. Específicamente, en el artículo 534, se define por lo que se entiende como un suplemento alimenticio en Chile:

*“Aquellos productos elaborados o preparados especialmente para suplementar la dieta con fines saludables y contribuir a mantener o proteger estados fisiológicos característicos tales como adolescencia, adultez o vejez.*

*Su composición podrá responder a un nutriente, mezcla de nutrientes y otros componentes presentes naturalmente en los alimentos, incluyendo compuestos tales como vitaminas, minerales, aminoácidos, lípidos, fibra dietética o sus fracciones.*

*Se podrán extender en diferentes formas de liberación convencional, tales como polvos, líquidos, granulados, grageas, tabletas, cápsulas u otras propias de los medicamentos”*

Junto con la definición anterior, el artículo 538 del mismo código, define cual es el procedimiento por el cual se determinan las concentraciones permitidas en los mismos suplementos, haciendo referencia a que el Ministerio de Salud es quien establece los niveles máximos y mínimos según la naturaleza del compuesto:

*“Los niveles, máximo y mínimo, de vitaminas, minerales y demás componentes a que alude el artículo 534, serán establecidos por resolución del Ministerio de Salud, dictada en uso de sus atribuciones legales técnico normativas”*

---

<sup>2</sup> USDA Foreign Agricultural Service (2018). Chilean Demand for Healthy Food Products Continues to Grow. USA: Global Agricultural Information Network

Estas regulaciones resultan claves para el desarrollo propio de la industria, ya que, como señala la misma agrupación que reúne a los 7 actores principales del rubro, el marco regulatorio del mercado de los suplementos es muy limitado en comparación con otros países, señalando las siguientes dificultades:

- Ausencia de un procedimiento administrativo formal para actualizar el listado de compuestos permitidos.
- La nómina de compuestos permitidos sólo enumera nutrientes consistentes en vitaminas y minerales, dejando fuera a todos los demás nutrientes que si existen y se comercializan en otros países.
- Pese a que se trata de alimentos que han sido especialmente elaborados para suplementar la dieta de las personas con fines saludables, se prohíbe expresamente informar la finalidad con que dichos productos han sido elaborados.

### **I.1.2 Puntos de venta**

La empresa que ha facilitado los datos para esta investigación se encuentra constantemente analizando la posibilidad de nuevas aperturas de puntos de ventas. Dentro de la compañía definen cuando un local se encuentra en fase de crecimiento o madurez según la antigüedad de funcionamiento (además del comportamiento de la curva de ventas del propio local). En términos generales, un punto de venta se dice que alcanza su fase de madurez, cuando ya posee más de 5 años en funcionamiento.

En 2016 la empresa presentaba 25 locales repartidos en todo el territorio nacional. Ya para 2017 se realizaron 5 aperturas para completar un total de 30. En 2018 la compañía decidió abrir 4 tiendas, alcanzando un total de 34 puntos en su red. En 2019 no se realizaron aperturas de nuevas tiendas. En la tabla 1 se adjunta este comportamiento, además del crecimiento de las ventas totales, con respecto al año anterior. Se puede apreciar observando la tabla 1, que en 2019 la empresa sufrió una caída de 20 puntos porcentuales en su crecimiento, alcanzando un 12%, en contraste con los años 2016, 2017 y 2018, donde se alcanzó un aumento de las ventas en torno a un 30%.



Año	Crecimiento de ventas	Número de tiendas
2016	35%	25
2017	31%	30
2018	29%	34
2019	12%	34

Tabla 1: Número de puntos de ventas de la empresa y su crecimiento de ventas

Con el propósito de graficar las ordenes de magnitud de la demanda, se presenta el siguiente gráfico (gráfico 1), que muestra los niveles de ventas mensuales (unidades totales) de los 5 locales que se investigan durante la presente tesis, desde enero de 2015, hasta diciembre de 2020.

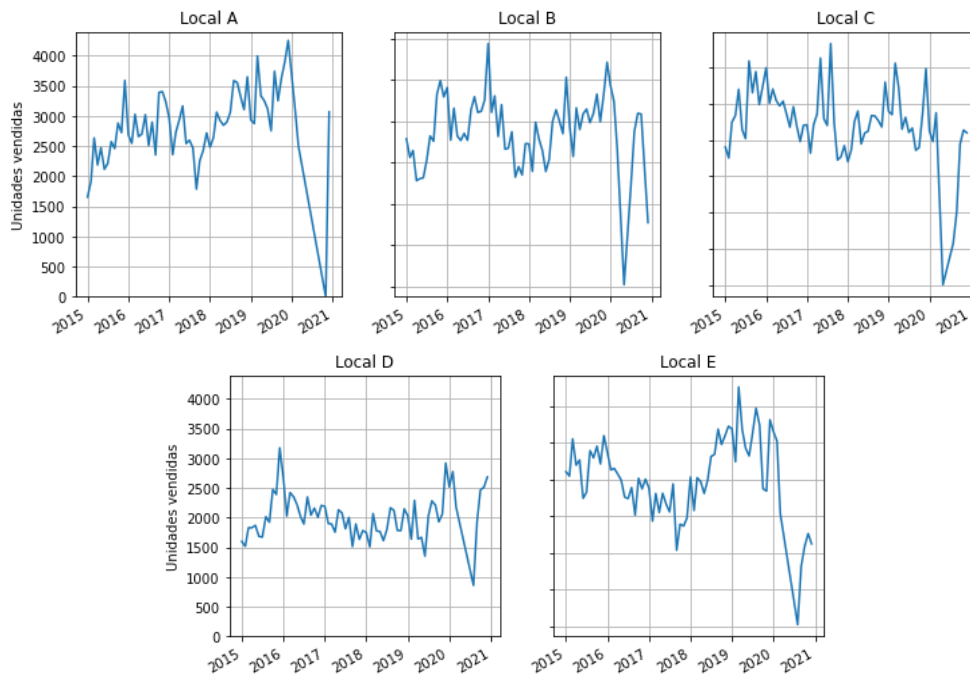


Gráfico 1: Unidades totales vendidas por mes desde enero de 2015 hasta diciembre de 2020, para cada uno de los 5 locales

En el gráfico 1 se puede apreciar el efecto de las restricciones al comercio durante 2020. Por otro lado, de los 5 locales en estudio, el que presenta un rango menor de ventas es el local

D, oscilando entre 1.500 y 3.000 unidades vendidas (a nivel mensual), mientras que el resto de los puntos oscilan entre 2.000 y 4.200 unidades vendidas.

Con el propósito de ilustrar cual ha sido el crecimiento de la demanda en cada uno de los 5 locales, se adjunta la siguiente tabla (tabla 2), donde se calcula cual es el aumento porcentual de las unidades vendidas, para cada uno de los puntos de ventas, comparando 2018 y 2019 con el año que le precede.

Local	2018/2017	2019/2018
A	20.1%	10.4%
B	-1.9%	14.8%
C	-3.3%	7.8%
D	-1.3%	7.9%
E	40.2%	14.2%

Tabla 2: Variación porcentual de las unidades vendidas en cada uno de los 5 locales

Se puede apreciar en la tabla 2 que los 5 locales lograron un crecimiento en ventas en 2019 con respecto a 2018. Por otro lado, el desempeño de 2018 con respecto a 2017 refleja que en los locales B, C y D tuvieron una contracción en torno a 1 y 3 puntos porcentuales. En el siguiente gráfico (gráfico 2), se puede observar cual es el comportamiento de la curva de demanda conjunta para los 5 puntos de ventas descritos anteriormente.

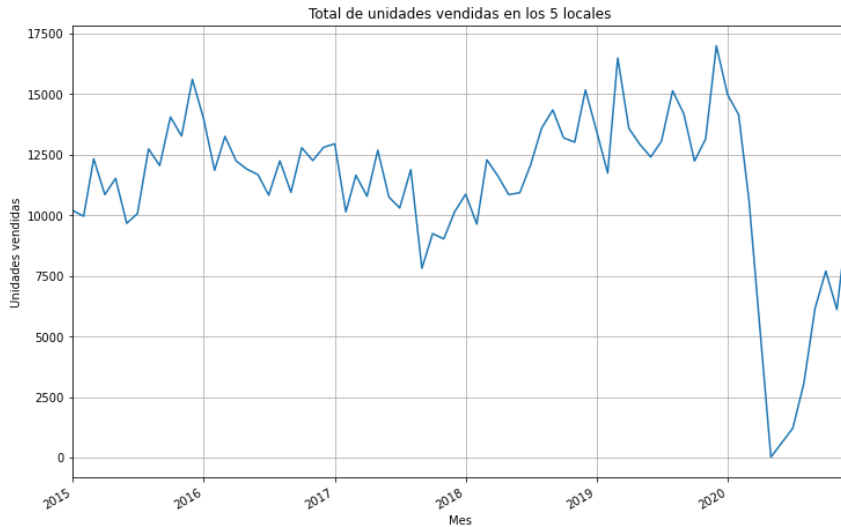


Gráfico 2: Unidades totales vendidas desde enero de 2015 hasta marzo de 2020, de los 5 locales en conjunto

Observando el gráfico 2, se aprecia que entre 2015 y finales de 2019, el rango de unidades vendidas oscila entre 7.500 y 17.500. Por otro lado, desde el segundo semestre de 2017 hasta iniciado el 2020 se puede apreciar una clara tendencia de crecimiento. Junto con lo anterior, visualmente no es posible encontrar estacionalidades claras dentro del periodo de estudio, esto puede explicarse, debido a que existen más de 1.200 productos distintos, cuyos comportamientos y estacionalidades puede presentarse en distintas etapas de un año natural, provocando que no exista una apreciación visual de estaciones marcadas. Además de lo anterior, a inicios de 2020 se puede observar una caída brusca en las ventas provocado por la irrupción de la pandemia y las restricciones aplicadas al comercio.

## I.2 Justificación

La tarea de gestionar el nivel de inventario es un fenómeno ampliamente estudiado, y donde el rol de la predicción de la demanda a la que estará sujeto un determinado punto de venta cobra un rol fundamental. Malas proyecciones pueden derivar en inventarios con exceso de abastecimiento o bien sin unidades suficientes para satisfacer la demanda de un determinado producto, afectando el nivel de competitividad de la empresa y su crecimiento (Ramos, Patrícia y Santos, Nicolau y Rebelo, Rui, 2015).

Los resultados que genera un modelo de predicción de demanda no sirven únicamente para decidir las unidades de abastecimiento. Shaohui, Kolassa, Stephan y Fildes (2018) señalan que el proceso de predicciones, tienen efectos en los 3 niveles de gestión empresarial:

- Nivel estratégico: En un ambiente competitivo y tecnológico, la definición de estrategias que abarcan el mercado, además de otros factores competitivos, dentro de un marco regulatorio cambiante (como es el caso de la industria de los suplementos alimenticios, como se explicó en la sección de Marco regulatorio), dependen en gran medida de los pronósticos, esto debido a que, si bien los modelos no tienen una alta precisión en ventanas de largo plazo (más de un año) si sirven como una fuente de alimentación en la decisión de estrategias a seguir, con el fin de percibir el crecimiento que se puede llegar a tener.
- Nivel táctico: Dentro de los principales objetivos del retail, es el de maximizar las ganancias a nivel de categorías (en lugar de las marcas). Así también es el caso de la empresa en estudio, que posee distintos niveles de categorías para los distintos productos. Para ello, es necesario fijar planes de precios y promociones, para determinar cuántas y la profundidad de estas en el horizonte de planificación. Parte importante de esto tiene que ver con la disponibilidad de los distintos productos en las góndolas, la cual depende en gran medida de la calidad de predicción que se utiliza.
- Nivel operativo: La planificación de la oferta y demanda resultan claves en toda empresa para no tener dificultades en el servicio al cliente. Junto con lo anterior, es necesario balancearlo con un inventario no tan alto, que pueda implicar muchas pérdidas en términos de mermas. Un buen nivel de estimación a nivel sku/local tiene como consecuencias una mejor satisfacción del cliente, menos pérdidas por desperdicios, mayores ingresos en términos de ventas y mejoras en la eficiencia.

Además de lo anterior, es importante señalar la existencia del “Efecto látigo” en la cadena de suministros. Aburto, Luis y Weber, Richard. (2007) apuntan a este efecto como una necesidad clave de ser más precisos en las estimaciones. Debido a que las predicciones que realiza una empresa presentan distorsiones, las proyecciones que realizaran sus proveedores basados en estas solicitudes presentarían una mayor variabilidad al utilizar información que ya acarrea distorsiones. Dentro de las alternativas (Aburto, Luis y Weber, Richard. 2007) para resolver el problema de alta variabilidad en la demanda están:

- Cada agente de la cadena de suministros puede aumentar su nivel de inventario.
- Aumentar la velocidad del flujo de información entre los participantes de la cadena de suministros.
- Mejorar los pronósticos de demanda. Esto permite operaciones más confiables y reducir los costos de inventarios asociados.

En la empresa en estudio tienen su propia metodología de estimación de la demanda, el cual consta de un proceso de 9 etapas, las cuales se describen a continuación y además se ilustran de manera sintetizada en la figura 1:

- Fase 1: En la empresa se estima cual es el crecimiento esperado a nivel global (considerando tanto locales como ventas por el canal online), en base al comportamiento histórico reciente. Lo anterior se combina con información macroeconómica como el crecimiento esperado del país.
- Fase 2: Se observa cual fue la precisión de las estimaciones del año anterior, y si es que existieron desviaciones significativas (más de 50%). En caso de existir una desviaciones grandes, se observa si es que hubo algún evento exógeno que explique dicha distorsión (como por ejemplo el cierre del local por remodelaciones), en caso de existir un evento exógeno, se corrige el valor real para ajustarlo a la predicción realizada.
- Fase 3: Existen distintos clústeres de locales, clasificados en base a la madurez y tamaño como aspectos más relevantes. Se observa su crecimiento histórico y se estima cual debería ser su aporte esperado en ventas durante el año.
- Fase 4: Se definen las ventas esperadas para cada uno de los puntos de venta.
- Fase 5: Se observa cual es el crecimiento histórico de las distintas categorías de productos que ofrece la empresa, y en base a ello se define cual será el crecimiento para cada una de las categorías durante el próximo año.
- Fase 6: Combinando los resultados de la fase 4 y 5, se definen cuáles serán los crecimientos, y con ello la demanda para cada una de las categorías en cada uno de los puntos de ventas.
- Fase 7: En base al comportamiento mensual, y los resultados obtenidos en la Fase 6, se definen las proyecciones para cada una de las categorías, por local a nivel mensual.

- Fase 8: Se procede a observar cual es la participación histórica y demanda de cada una de las marcas que ofrece la compañía. En base a esta información y la obtenida en la Fase 7, se definen proyecciones para cada una de las marcas dentro de las categorías existentes.
- Fase 9: En esta etapa se realiza una validación global del proceso y se confirma (o rechaza) las proyecciones realizadas.

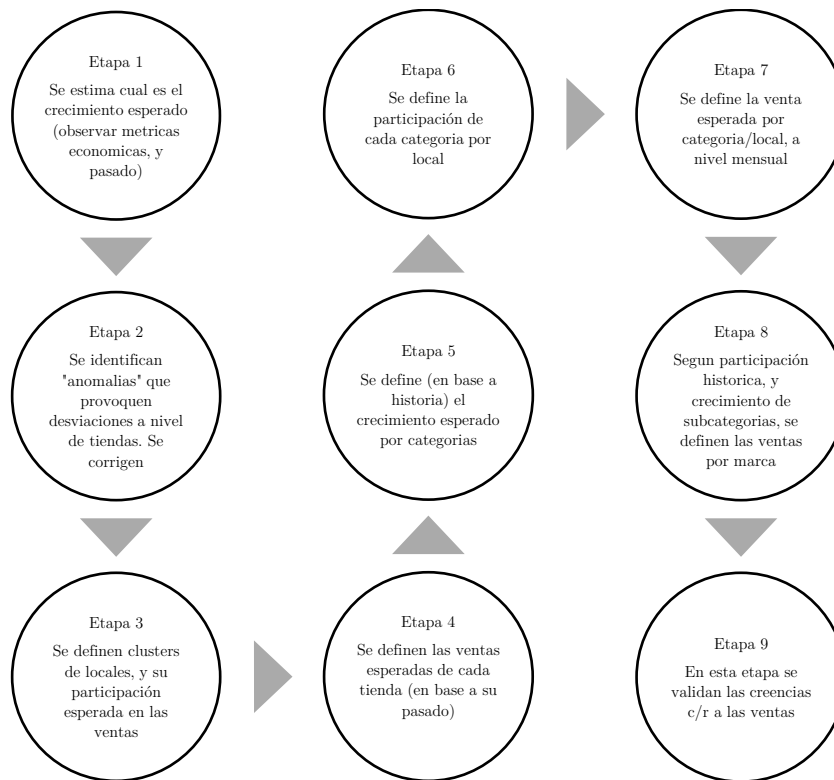


Figura 1: Fases en el proceso de estimación de demanda que realiza la empresa

Con respecto al proceso de estimación de demanda, dentro la compañía que ha facilitado los datos para esta investigación se señala que con su metodología actual ocurren entre otras, las siguientes problemáticas:<sup>3</sup>

<sup>3</sup> Gerencia general de la compañía, comunicación personal, diciembre 2020

- Muchas horas/trabajador dedicadas en la elaboración de presupuesto y estimación futura de la demanda.
- Se generan presupuestos/proyecciones por categorías y subcategorías en base a expectativas de crecimiento, pudiendo incurrir en errores de sesgo.
- Existe una precisión razonable a nivel global (desviaciones en torno al 8%), sin embargo, dentro de la empresa se observa que el nivel de precisión global se debe a que ciertos locales venden más de lo estimado, otros menos, generando un efecto de compensación (observar tabla 3).
- La metodología actual “elimina” datos que se desvíen mucho de sus estimaciones, lo que conlleva a que no exista una forma definida de proyectar el impacto en la demanda ante distintos eventos particulares.

Mes	Categoría	Desviación
Enero	Dieta	-15.45%
Febrero	Deporte	-26.15%
Febrero	Accesorio	14.34%
Junio	Deporte	-29.02%
Agosto	Salud	21.38%
Agosto	Deporte	-19.70%
Septiembre	Accesorio	21.88%

Tabla 3: Error porcentual mensual de la estimación de ventas en distintas categorías

## I.3 Objetivos

### Objetivo General

- Desarrollar modelos de estimación de demanda mensual para distintas categorías de productos, en 5 puntos de ventas de una empresa de suplementos alimenticios, con el propósito de fortalecer el proceso de toma de decisiones y abastecimiento de la compañía.

### Objetivos Específicos

- Caracterizar a las categorías según su demanda histórica y otras características como estacionalidad, tendencia y estacionariedad.
- Desarrollar modelos de suavización exponencial, ARIMA y redes neuronales recurrentes para distintos niveles de agregación de categorías de productos.
- Estimar la demanda de las categorías en una ventana de 12 meses a futuro.
- Aplicar pruebas estadísticas para validar que los resultados tengan diferencias significativas, y evaluar el desempeño de los modelos según métricas como MAPE o NRMSE.



# Capítulo II – Alcances y resultados esperados

## II.1 Alcances

Los alcances de la presente investigación se resumen en los siguientes puntos:

- Los datos que se utilizarán para realizar las proyecciones de demandas corresponden a 5 puntos de venta distintos. Estos se eligen debido a la homogeneidad de los datos que presentan dichos locales con respecto a la ventana de tiempo en el que ocurren.
- La ventana de tiempo de los datos que se utilizara abarca el espacio temporal desde enero de 2015 hasta marzo de 2020.
- Las proyecciones se realizarán para 2 niveles distintos de agrupación de los productos (estos niveles se describen con mayor detalle en el capítulo Datos), para cada uno de los 5 locales, siendo un total de 130 series de tiempo distintas.
- No se utilizarán los datos correspondientes a la devolución de productos.
- Se utilizará la información correspondiente a 1.391 skus distintos.
- Las ventanas de tiempo para las cuales se realizarán las estimaciones serán a nivel mensual.
- Se asume que siempre hay existencias de las distintas categorías durante el periodo de estudio. En el caso que no se registre ventas en un determinado mes, se asumirá que no hubo demanda de la categoría en cuestión.

## II.2 Resultados esperados

Los resultados esperados en el proceso de investigación se desglosan en los siguientes puntos:

- Predicciones de modelos de series de tiempo (suavización exponencial y ARIMA) y machine learning (redes neuronales recurrentes) para cada uno de 5 los puntos de venta, para distintos niveles de agrupación de productos (Grupos 1 y 2), a nivel mensual y por local, utilizando los datos de las boletas históricas desde 2015.
- Determinar dentro de un conjunto de hiperparametros y arquitecturas de redes neuronales cual es mejor para cada una de las 130 series en estudio
- Determinar en el caso de las redes neuronales recurrentes, cual es la ventana de tiempo pasada óptima para poder estimar la demanda futura.
- Validación estadística de los errores de los modelos mediante tests de hipótesis paramétricos y no paramétricos.

# Capítulo III – Marco teórico

## III.1 Conceptos de series de tiempo

Las definiciones y conceptos que se utilizan en esta y la sección: Modelos clásicos de series de tiempo, son extraídas de Adhikari, Ratnadip y Agrawal, R. (2013). An Introductory Study on Time series Modeling and Forecasting, Montgomery, Jennings y Kulahci (2008). Introduction to Time Series Analysis and Forecasting y de Hyndman, R. J. y Athanasopoulos, G. (2013). Forecasting: principles and practice.

### III.1.1 Definiciones (tendencia, estacionalidad, ciclos y errores)

Una serie de tiempo se define como una secuencia de datos que son medidos en ciertos intervalos en un espacio temporal definido. Sea  $t \in \{1, 2, 3 \dots T\}$ , el arreglo  $(y_1, y_2, y_3 \dots y_T)$ , donde  $y_i \in \mathbb{R}$  (variable cuyo valor es registrado en el momento  $i$ ), representa la serie de tiempo de la variable  $y$ . Esta definición no solo es válida para el caso cuando  $y_i$  es una variable de dimensión 1, sino que también aplica en el caso que pertenezca a un espacio dimensional  $n > 1$ . En este caso, se dice que  $y_i \in \mathbb{R}^n$ , es decir,  $y_i$  es un vector que posee  $n$  valores, los cuales son registrados en cada instante  $i$ . Por otro lado, se considera en ambos casos a  $y_i$  como una variable o proceso aleatorio debido a que antes de su realización no se tiene certeza de cuál es el valor exacto que esta tendrá.

Las series de tiempo en general pueden ser descritas mediante cuatro componentes distintas: tendencia, ciclos, estacionalidad y componentes irregulares o errores:

- Tendencia ( $T_t$ ): Se dice que una serie de tiempo posee una tendencia cuando existe un sostenido aumento o disminución de los valores que toma la variable.

- Estacionalidad ( $S_t$ ): Se define la estacionalidad como aquel fenómeno que ocurre y afecta la serie cada cierto periodo de tiempo definido.
- Ciclo ( $C_t$ ): Los ciclos hacen referencia a patrones en periodos de tiempo largos (dos años, por ejemplo). En general se dice que actúan en el mediano plazo de la serie.
- Componentes irregulares o errores ( $I_t$ ): Los errores capturan todo aquello que las tres componentes anteriores no captaron.

En base a las cuatro componentes mencionadas anteriormente, existen dos tipos de modelos que son utilizados con mayor frecuencia: Modelos aditivos y multiplicativos:

- Modelo aditivo: En este modelo se dice que la serie de tiempo puede ser descrita como la suma de las cuatro componentes anteriores:

$$y_t = T_t + S_t + C_t + I_t$$

- Modelo multiplicativo: En este modelo se dice que la serie de tiempo puede ser descrita como la multiplicación de las cuatro componentes anteriores:

$$y_t = T_t \times S_t \times C_t \times I_t$$

### III.1.2 Proceso estocástico y estacionariedad

Una serie de tiempo, debido a la incertidumbre que existe en la realización de los periodos que no han ocurrido, se puede definir como un proceso estocástico, es decir, es una colección de variables aleatorias que tienen un orden temporal.

Sea  $A = \{x_t: t = k, k + 1, \dots, k + T\}$ , el conjunto de variables obtenidas de la realización de  $x_t$  entre  $t = k$  y  $t = k + T$ , se dice que esta sigue una función de distribución que describe

la probabilidad de ocurrencia conjunta del conjunto  $A$ . Sea la  $f$  la función de densidad de probabilidad se puede escribir:

$$f(A) = f(x_k, x_{k+1} \dots x_{k+T})$$

Es decir, existe una forma funcional (función de densidad conjunta) que describe cual es la probabilidad de ocurrencia para  $(x_k, x_{k+1} \dots x_{k+T})$ . Cuando dicha forma funcional es igual para una misma cantidad de mediciones ( $T$ ), independiente de la ventana temporal en la que se esté realizando la medición  $(k, k + T)$ , se dice que la serie de tiempo  $x_t$  posee una estacionariedad fuerte.

Por otro lado, un proceso estocástico se dice que posee una estacionariedad débil de orden  $m$ , si los  $m$  momentos del proceso solo dependen del número de variables que se están considerando ( $T$ ), y no de la ventana temporal donde se esté realizando la medición  $(k, k + T)$ .

### III.1.3 Autocovarianza, autocorrelación y autocorrelación parcial

En series de tiempo el término “lag” hace referencia al par de valores que se encuentran separados por una ventana temporal  $k$ . Así, sea una serie de tiempo  $\{y_t: t = 1, 2 \dots T\}$ , el conjunto de tuplas  $(y_t, y_{t+k})$  representan la serie de valores separados por un lag igual a  $k$ . En base a lo anterior, se puede calcular la covarianza de este par de valores. Este término se denomina como la autocovarianza de lag  $k$ . Sea  $\mu$ , la media de  $y_t$ , la autocovarianza de lag  $k$  se define matemáticamente como:

$$\gamma_k = Cov(y_t, y_{t+k}) = E[(y_t - \mu)(y_{t+k} - \mu)]$$

Un estimador que se utiliza comúnmente para calcular la autocovarianza, para  $k = 1, 2 \dots$  es:

$$\widehat{\gamma}_k = \frac{1}{T} \sum_{t=1}^{T-k} (y_t - \bar{y})(y_{t+k} - \bar{y})$$

El conjunto de valores  $\gamma_k$ , para  $k = 1, 2, 3 \dots$  se denomina la función de autocovarianza. En base a lo anterior, se puede definir el coeficiente de autocorrelación de lag  $k$  como:

$$\rho_k = \frac{E[(y_t - \mu)(y_{t+k} - \mu)]}{\sqrt{E[(y_t - \mu)^2]E[(y_{t+k} - \mu)^2]}} = \frac{Cov(y_t, y_{t+k})}{Var(y_t)} = \frac{\gamma_k}{\gamma_0}$$

En este caso,  $\gamma_0 = \sigma^2$ , es la varianza de  $y_t$ , además  $-1 < \rho_k < 1$ . El conjunto de valores de  $\rho_k$  para  $k = 1, 2 \dots$  se denomina la función de autocorrelación (ACF).

Por otro lado, existe la autocorrelación parcial, la cual intenta medir la correlación de una variable  $y_{t-k}$  sobre  $y_t$ , controlando previamente en el conjunto de variables que se encuentran entre ambas mediciones ( $\{y_i: i \in t-k, k\}$ ). El conjunto de estos valores para  $k = 1, 2 \dots$  se denomina la función de autocorrelación parcial (PACF). Para un lag igual a 1, es posible notar que PACF(1) toma el mismo valor que ACF(1).

## III.2 Modelos clásicos de series de tiempo

### III.2.1 Suavización exponencial simple

Los métodos de suavización exponencial tienen como objetivo, otorgar distintas ponderaciones a valores pasados de una variable  $y_t$ . Matemáticamente, sea un conjunto  $\{y_t: t = 1, 2, \dots, T\}$ , y un parámetro  $\alpha$ , donde  $0 < \alpha < 1$ , se puede escribir una serie ponderada como:

$$\sum_{t=1}^{T-1} \alpha^t y_{T-t} = \alpha y_T + \alpha^2 y_{T-1} + \dots + \alpha^{T-1} y_1$$

En base a lo anterior, se busca que la suma de los ponderadores que acompañan a la variable  $y_t$ , sea igual a 1. La suma de los ponderadores como se definió en la ecuación anterior cumple que:

$$\sum_{t=1}^{T-1} \alpha^t = \frac{1 - \alpha^T}{1 - \alpha}$$

Dado que la forma de los ponderadores no garantiza que su suma sea igual a 1, es que se multiplica la igualdad anterior por un factor  $(1 - \alpha)/(1 - \alpha^T)$ , con ello se tiene que:

$$\sum_{t=1}^{T-1} \alpha^t \frac{(1 - \alpha)}{(1 - \alpha^T)} = \frac{(1 - \alpha^T)}{(1 - \alpha)} \frac{(1 - \alpha)}{(1 - \alpha^T)} = 1$$

Generalmente se suele realizar un cambio de parámetros para simplificar la posterior notación. Así, un cambio típico es el de definir  $\alpha = 1 - \alpha^*$ . Con ello, la formula cuando  $T \rightarrow \infty$  implica que  $(1 - \alpha^*)^T \rightarrow 0$ , debido a que  $0 < 1 - \alpha^* < 1$ , con lo que el modelo quedaría de la siguiente forma:

$$y_{T+1|T} = \sum_{t=0}^{T-1} \alpha^* (1 - \alpha^*)^t y_{T-t} = \alpha^* y_T + \alpha^* (1 - \alpha^*) y_{T-1} \dots + \alpha^* (1 - \alpha^*)^{T-1} y_1 + (1 - \alpha^*)^T l_0$$

En esta ecuación,  $y_{T+1|T}$  representa la estimación en  $T + 1$  usando los  $T$  pasos anteriores. Cuando el factor  $\alpha^*$  se aproxima a 0 quiere decir que la ecuación da más importancia a los datos antiguos que a los más recientes, por otro lado cuando  $\alpha^*$  se aproxima a 1, el modelo da más importancia a la información más reciente. El termino  $l_0$  se puede explicar cuando la ecuación se escribe en forma recursiva:

$$y_{t+1|t} = \alpha^* y_t + (1 - \alpha^*) y_{t|t-1}$$

Con lo que, retrocediendo en cada paso se tiene que:

$$y_{4|3} = \alpha^* y_3 + (1 - \alpha^*) y_{3|2}$$

$$y_{3|2} = \alpha^* y_2 + (1 - \alpha^*) y_1$$

$$y_{2|1} = \alpha^* y_1 + (1 - \alpha^*) y_{1|0}$$

Debido a que  $y_{1|0}$  no está definido, se fija este como un parámetro inicial,  $l_0$ . Para proyectar un valor que se encuentra en  $t + h$ , dado que se tiene información hasta el periodo  $t$ , se utiliza el ultimo nivel calculado, es decir, sea  $y_{t+h|t}$ . La proyección en  $t + h$  se define como:

$$y_{t+h|t} = y_{t+1|t}$$

Para encontrar el parámetro  $\alpha$ , se minimiza una función objetivo. En general, una forma común es minimizar el error cuadrático medio, descrito como:

$$MSE = \frac{1}{T} \sum_{t=1}^T (y_t - y_{t|t-1})^2 = \frac{1}{T} \sum_{t=1}^T [y_t - \sum_{i=1}^{t-1} \alpha^* (1 - \alpha^*)^{t-1-i} y_i]^2$$

### III.2.2 Modelo de Holt's

El método de Holt's, también conocido como Suavización exponencial doble, considera que existen dos fuentes que explican una serie de tiempo: el nivel  $L_t$  y la tendencia  $T_t$ . Para calcular estas dos componentes, se utilizan dos ecuaciones:

$$L_t = \alpha y_t + (1 - \alpha)(L_{t-1} + T_{t-1})$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$

La primera ecuación dice que el nivel en  $t$  es una combinación lineal entre el valor real de la variable  $y$ , y la suma entre el nivel y la tendencia del periodo anterior ( $t - 1$ ). Por otro



lado, la segunda ecuación señala que la tendencia en  $t$  es una combinación lineal entre la diferencia entre los periodos consecutivos  $t$  y  $t - 1$ , es decir, el crecimiento de la serie, y la tendencia en el periodo  $t - 1$ .

Como se puede apreciar, en este modelo se incorpora un parámetro adicional,  $\beta$ , el cual busca capturar la tendencia de la serie, es decir, si esta se encuentra en crecimiento o decrecimiento. Al igual que en la suavización exponencial simple, en este modelo ambos parámetros solo pueden tomar valores entre 0 y 1.

Para proyectar al periodo  $t + h$  (en su forma aditiva), dado que se utiliza información hasta  $t$ , se utiliza el nivel, y la tendencia ponderada por un factor  $h$ , es decir:

$$y_{t+h|t} = L_t + hT_t$$

### III.2.3 Modelo de Holt/Winters

El modelo de Holt's/Winters considera que una serie de tiempo puede ser descrita por 3 componentes: nivel ( $L_t$ ), tendencia ( $T_t$ ) y estacionalidad ( $S_t$ ). En ese sentido, se incorpora un parámetro adicional con respecto al modelo de suavización exponencial doble, tendiendo así  $\alpha, \beta$  y  $\gamma$ , los cuales solo pueden tomar valores entre 0 y 1.

Las ecuaciones con las que se calculan el nivel, la tendencia y la estacionalidad, en la forma aditiva del modelo, son:

$$L_t = \alpha(y_t - S_{t-s}) + (1 - \alpha)(L_t + T_{t-1})$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$

$$S_t = \gamma(y_t - L_t) + (1 - \gamma)S_{t-s}$$

Se puede observar que en la tercera ecuación aparece un término  $s$ . Este hace referencia a al tamaño de la estacionalidad. Por ejemplo, si se está trabajando con series de tiempo mensuales, y se presume que existe una estacionalidad una vez al año, entonces  $s = 12$ .

Para predecir a una distancia  $h$ , dado que se utiliza información hasta el periodo  $t$ , se utiliza la siguiente ecuación:

$$y_{t+h|t} = L_t + hT_t + S_{t-s+h_s}$$

En la ecuación anterior, el término  $h_s$ , se define como:

$$h_s = [h - 1 \text{ mod } s] + 1$$

Donde *mod* es el resto que se obtiene de dividir  $h - 1$  con  $s$ .

### III.2.4 ARIMA (Autoregressive integrated moving average)

El modelo ARIMA se divide en tres componentes esenciales: Autoregresión (AR), Medias móviles (MA) e Integración (I). La unión de estas tres componentes da origen al modelo ARIMA, el cual consta de 3 parámetros,  $p$ ,  $d$  y  $q$ , obtenido de cada una de las componentes mencionadas anteriormente.

#### III.2.4.1 Autoregresión (AR)

La componente de Autoregresión busca modelar el valor futuro de una determinada variable, en función de una combinación lineal de sus valores pasados. En ese sentido, el parámetro  $p$ , es un número entero que define cuantas observaciones anteriores deben ser utilizadas. Matemáticamente, una Autoregresión de orden  $p$ ,  $AR(p)$ , se expresa de la siguiente manera:

$$y_t = k + \sum_{i=1}^p \varphi_i y_{t-i} + \varepsilon_t = k + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} \dots + \varphi_p y_{t-p} + \varepsilon_t$$

La componente  $k$  en la ecuación anterior es una constante. El conjunto  $\{\varphi_i\}_{i=1}^p$  son los parámetros que acompañan a las observaciones anteriores de  $y_t$ . Finalmente,  $\varepsilon_t$  es una componente o shocks aleatorios que el modelo no captura.

### III.2.4.2 Medias móviles (MA)

La componente de medias móviles, de orden  $q$ , es una combinación lineal de los  $q$  errores pasados, los cuales son usados como variables explicativas. Matemáticamente se describe de la siguiente manera:

$$y_t = \mu + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \eta_t = \mu + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \eta_t$$

En la ecuación anterior,  $\mu$  representa la media de la variable  $y$ . El conjunto  $\{\theta_i\}_{i=1}^q$ , son los parámetros que acompañan a las  $q$  “observaciones” anteriores de los errores. Por otro lado, el término  $\eta_t$  representa los shocks aleatorios. En general, se le refiere como ruido blanco, por lo que se considera que el conjunto de variables  $\{\eta_t\}_{t=1}^T$ , distribuyen idénticamente de manera independiente, con media cero y varianza constante  $\sigma^2$ .

### III.2.4.3 Integración (I)

La componente de Integración de orden  $d$  se utiliza para transformar una serie en estacionaria. Así,  $d$ , hace referencia a cuantas veces es necesario diferenciar elementos consecutivos de la variable  $y$ . Por ejemplo, si  $d = 1$ , esto quiere decir que solo se diferencia la variable  $y_t$  una vez. Matemáticamente se vería como:

$$\Delta^d y_2 = y_2 - y_1$$

$$\Delta^d y_3 = y_3 - y_2$$

.

.

.

$$\Delta^d y_t = y_t - y_{t-1}$$

Luego, las componentes  $\Delta^d y_t$  son utilizadas para obtener el modelo ARIMA. Es importante remarcar, que según la cantidad de diferenciaciones que se aplique, la cardinalidad del conjunto resultante tiene  $d$  elementos menos que el original.

#### III.2.4.4 Autoregressive integrated moving average (ARIMA)

Combinando las tres componentes, AR( $p$ ), MA( $q$ ) e I( $d$ ), se obtiene el modelo ARIMA( $p, d, q$ ).

Cuando la componente de Integración es igual a cero, es decir,  $d = 0$  y por ende no es necesario realizar la diferenciación de la variable (debido a que esta ya es estacionaria), el modelo que se tiene se denomina ARMA( $p, q$ ), o Autoregression Moving Average. Matemáticamente, si los parámetros son  $p$  y  $q$ , se escribe de la siguiente manera:

$$y_t = k + \sum_{i=1}^p \varphi_i y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \phi_t$$

En la ecuación anterior,  $\phi_t$  representa el error del modelo en el periodo  $t$ ,  $k$  es una constante, y los conjuntos  $\{\varphi_i\}_{i=1}^p$  y  $\{\theta_j\}_{j=1}^q$  son los parámetros que acompañan a las observaciones pasadas de  $y_t$  y los errores pasados respectivamente.

Cuando la componente de diferenciación es igual a 1, es decir,  $d = 1$ , y los parámetros de Autoregresión y medias móviles son  $p$  y  $q$  respectivamente, se tiene un modelo ARIMA( $p, 1, q$ ), el cual se escribe de la siguiente manera:

$$\Delta^d y_t = k + \sum_{i=1}^p \varphi_i \Delta^d y_{t-i} + \sum_{j=1}^q \theta_j \Delta^d \varepsilon_{t-j} + \phi_t$$

Es importante recalcar que el modelo ARIMA, puede optar por configuraciones donde se utilizan distinta cantidad de observaciones pasadas para las componentes de Autoregresion y Medias Móviles, es decir cuando  $p \neq q$ .

## III.3 RNN (Recurrent neural networks)

La mayoría de los términos y conceptos que se desarrollan en esta sección son extraídos de Brownlee (2020). Deep Learning for Time Series Forecasting.

### III.3.1 Algoritmos supervisados, no supervisados y de reforzamiento

Según el tipo de datos que se dispone, y del propósito del estudio que se esté realizando, existen tres tipos de algoritmos de machine learning:

- Algoritmos de aprendizaje supervisados.
- Algoritmos de aprendizaje no supervisados.
- Algoritmos de aprendizaje de reforzamiento.

Los algoritmos de aprendizaje supervisados son aquellos donde existe una variable  $Y$ , la cual se intenta calcular en base a una variable (o variables)  $X$ . Se intenta encontrar una función  $f$ , que logre explicar la relación entre  $X$  e  $Y$ . En otras palabras, sea  $x \in X$  e  $y \in Y$ , el objetivo es encontrar  $f$  tal que:

$$f(x) = y + \varepsilon$$

Donde  $\varepsilon$  representa un componente de error en la estimación de  $y$ . Por otro lado, los algoritmos de aprendizaje no supervisado carecen de una variable objetivo  $Y$ , por lo que en base a determinados criterios se busca encontrar relaciones dentro de las variables disponibles  $X$ .

Finalmente, los algoritmos de aprendizaje de reforzamiento consisten en métodos de entrenamientos de un agente, el cual recibe una “recompensa” según la interpretación que este hace de la información que procesa.

### **III.3.2 RNN: Arquitectura y optimización**

#### **III.3.2.1 Forma de una RNN**

Las redes neuronales recurrentes (RNN) son un tipo de algoritmo supervisado, el cual es utilizado para datos que siguen una secuencia estructurada. En general suele usarse con datos en forma de serie de tiempo, pero también son utilizadas para problemas como el procesamiento del lenguaje, traducción de idiomas, clasificación de imágenes entre otros.

Rumelhart et al. (1986) proponen las redes neuronales recurrentes como una alternativa a las redes neuronales Feedforward tradicionales (FNN). La estructura de una red neuronal recurrente tiene forma de cadena, donde en cada etapa dentro de esta, se ingresa una realización de una variable  $X$ , la cual posee un orden secuencial propio.

La forma en la que se estructuran las RNN permiten compartir información entre realizaciones, con lo que, al momento en que la red genera un resultado, este no solo es calculado en base a la última realización  $x_T$ , sino que también existe una influencia de las  $T - 1$  variables anteriores.

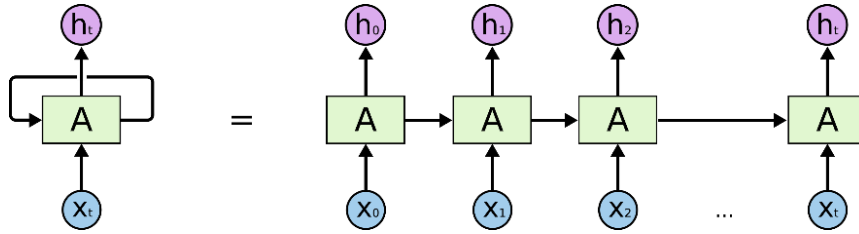


Figura 2: Dos formas de representar una red neuronal recurrente (Colah, 2015)

La figura 2 muestra dos formas distintas en que una red neuronal recurrente puede ser representada. Los círculos azules representan las variables de entrada,  $x_t$ , las cuales deben tener un orden secuencial natural. Los círculos morados representen los estados “ocultos” o hidden states ( $h_t$ ), que son los resultados que genera la red luego de realizar ciertas operaciones. Los rectángulos verdes son las unidades neuronales o capas. En ellas se realizan distintas operaciones, luego de recibir la información  $x_t$  y  $h_{t-1}$ . Es importante remarcar que en una RNN existe compartimiento de información entre variables (esta es una de las grandes diferencias con una FNN), la cual es entregada de unidad neuronal en unidad neuronal a través de los hidden states ( $h_t$ ). Así, para la capa  $l$ , la cual recibe como input  $x_l$ , recibe también información del conjunto  $\{x_t\}_1^{l-1}$ , a través del estado calculado en la unidad neuronal  $l - 1$ ,  $h_{l-1}$ .

El número de unidades neuronales que posee una RNN depende de la cantidad de “time steps” o “pasos” que se definan. Así, si para un determinado conjunto de datos  $\{x_t, y_t\}_{t=1}^T$ , si se intenta estimar el valor de  $y_n$  (con  $1 < n < T$ ), y se define que los últimos  $m$  valores de  $x_t$  ( $x_n, x_{n-1}, \dots, x_{n-m+1}$ ) son relevantes para poder calcular  $y_n$ , entonces la red neuronal tendrá  $m$  unidades neuronales.

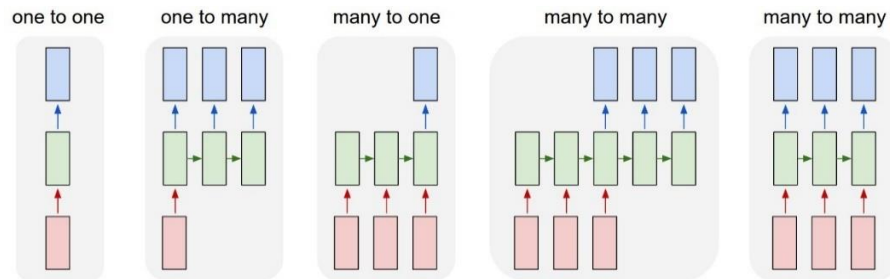


Figura 3: Distintas arquitecturas de una red neuronal recurrente (Karpathy, 2015)

Las redes neuronales recurrentes pueden ser estructuradas de muchas formas. En la figura 3 se presentan 5 de las arquitecturas simples más utilizadas en este algoritmo. Los rectángulos rojos representan los “inputs” o variables que se ingresan las unidades neuronales, las cuales son representadas por los rectángulos verdes. Los “outputs” o “hidden states” están representados por los rectángulos azules.

La arquitectura One-to-One, o Uno-a-Uno es la forma más sencilla de una RNN. En estricto rigor, al no tener una forma secuencial, se le podría considerar como una FNN. La estructura One-to-Many o Uno-a-Muchos considera solo un “input” para poder calcular un número mayor definido de “outputs”. Generalmente esta forma se utiliza para predecir secuencias de palabras con una sola imagen. Por otro lado, la arquitectura Many-to-One, o Muchos-a-Uno utiliza varios “inputs” para poder generar un único resultado. Un ejemplo del uso de esta forma es la clasificación del sentimiento de cadenas de texto. La forma Many-to-Many utiliza más de un “input” para generar más de un resultado. Un uso que se le da es el caso en que se utiliza una secuencia de audio para generar una cadena de palabras.

### III.3.2.2 Ecuaciones y unidad neuronal

Para poder describir mejor el conjunto de ecuaciones que componen una red neuronal recurrente, primero es relevante comprender que sucede dentro de una unidad neuronal.

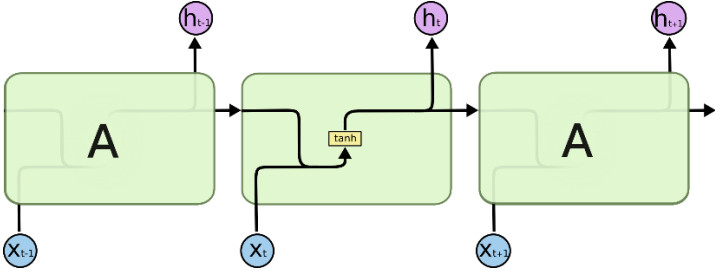


Figura 4: Dos formas de representar una red neuronal recurrente (Colah, 2015)

En la figura 4 podemos observar una RNN con 3 unidades neuronales. Al centro, se puede apreciar un recuadro amarillo con la función tangente hiperbólica ( $\tanh$ ). Esta función se denomina función de activación, y es un hiperparametro dentro de la red neuronal. Este concepto será descrito con mayor detalle en la sección hiperparametros.



A la unidad neuronal ingresa el valor  $x_t$ , el cual puede ser tanto un valor escalar como un vector de dimensión  $n$ . Al ingresar a la capa, es ponderado por una matriz  $W_{xh}$ , tal que  $W_{xh} \in M_{m \times n}(\mathbb{R})$ . En este caso, la dimensión  $m$  de la matriz es un hiperparámetro, el cual recibe el nombre de neuronas o unidades. Este hiperparámetro hace referencia a la dimensión de los “hidden states”, por lo que, en este caso,  $h_t \in M_{m \times 1}(\mathbb{R})$ . Junto con lo anterior, en algunas redes se incorpora un parámetro denominado “bias”,  $b_h$ , el cual es de dimensiones  $m \times 1$ . Además de  $W_{xh}$ , existe otra matriz,  $W_{hh}$ , tal que  $\dim W_{hh} = m \times m$ , la cual es ponderada por el estado oculto que la unidad neuronal anterior,  $h_{t-1}$ .

En algunas redes, el resultado que entrega cada una de las unidades neuronales es el mismo “hidden state”, sin embargo, este vector si bien es compartido a la siguiente unidad neuronal, es normal realizar una última transformación antes de generar el resultado final. En ese sentido, existe en la capa de salida, una matriz  $W_{hy}$  y un vector  $b_y$  (de ser especificado así), de dimensiones  $r \times m$  y  $m \times 1$  respectivamente. El término  $r$  hace referencia a la dimensión del resultado  $y_t$ .

Además de lo anterior, existen dos funciones de activación,  $G_h$  y  $G_y$ , las cuales se aplican sobre las operaciones descritas anteriormente. En base a esto, las ecuaciones que describen a una RNN son:

$$\begin{aligned} h_t &= G_h(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \\ y_t &= G_y(W_{hy}h_t + b_y) \end{aligned}$$

### III.3.3 Aprendizaje y optimización

#### III.3.3.1 Función de pérdida

La función de pérdida o función objetivo de una RNN es una ecuación que se busca minimizar. Una menor función de pérdida se traduce en un modelo que se tiende a ajustar mejor. En ese sentido es común dividir los datos disponibles en 3 subconjuntos: Entrenamiento, Validación y Testeo. El primer conjunto se utiliza para encontrar los

parámetros de la red,  $W_{xh}, W_{hh}, W_{hy}, b_h$  y  $b_y$ . El segundo se utiliza para elegir la mejor combinación de hiperparámetros para la RNN según la función de pérdida en el conjunto de validación. Finalmente, se evalúa el desempeño del modelo en el conjunto de testeo.

Sea  $\theta = [W_{xh}, W_{hh}, b_h, W_{hy}, b_y]$  el conjunto de parámetros y  $\{y_t\}_{t=r}^p$ , el conjunto de variables a los cual se busca ajustar el modelo, una función de pérdida típica, en el caso que  $y_t \in \mathbb{R}^n$ , y se posea un problema de regresión y no de clasificación, es utilizar el error cuadrático medio, o mean squared error (MSE). En base a lo anterior, la función objetivo se puede escribir como:

$$Loss(W_{xh}, W_{hh}, b_h, W_{hy}, b_y) = \frac{1}{p - r + 1} \sum_{t=r}^p \left( y_t - \widehat{y}_t(W_{xh}, W_{hh}, b_h, W_{hy}, b_y) \right)^2$$

Con lo que, en base a lo anterior, el objetivo es:

$$\arg \min_{\theta \in \Theta} Loss(W_{xh}, W_{hh}, b_h, W_{hy}, b_y)$$

El termino  $\Theta$  representa el espacio de dominio del parámetro  $\theta$ . En algunos casos se utiliza la suma de los errores cuadráticos en lugar de MSE. También se utiliza el error absoluto medio como una función objetivo.

### III.3.3.2 Gradiente descendente

Una metodología usual para poder minimizar una función es utilizar formulas cerradas. La más común es la de encontrar el conjunto de parámetros donde la pendiente de la función es cero, sin embargo, este método en muchos casos no es replicable debido a la complejidad de la función de pérdida. El método de gradiente descendente es otra metodología usada. En cada iteración de este algoritmo, se actualizan los parámetros en la dirección opuesta la gradiente de la función,  $\nabla_{\theta} L(\theta)$ . Además de lo anterior, se utiliza una tasa de aprendizaje, o “learning rate”, comúnmente especificada por la letra  $\eta$ . Este

hiperparametro indica que tan grande será el desplazamiento en la actualización de los parámetros en cada iteración.

Antes de describir el algoritmo de gradiente descendente, primero definimos  $\nabla_{\theta}L(\theta)$  para una red neuronal recurrente de una capa:

$$\nabla_{\theta}L(\theta) = \begin{bmatrix} \partial L / \partial W_{xh} \\ \partial L / \partial W_{hh} \\ \partial L / \partial b_h \\ \partial L / \partial W_{hy} \\ \partial L / \partial b_y \end{bmatrix}$$

La gradiente es un vector de dimensión  $n$ , donde  $n$  es el número de parámetros en los que se encuentra definida la función. A continuación, se procede a describir los pasos del algoritmo basado en Smola y Vishwanathan (2008):

---

#### Algoritmo de Gradiente Descendente

---

- 1: Inicializar  $\theta_o$ , y definir tolerancia para finalizar  $\epsilon$
  - 2: Definir  $i = 0$ :
  - 3: While  $\|\nabla_{\theta}L(\theta^i)\| \geq \epsilon$ :
  - 4:      $\theta^{i+1} = \theta^i - \eta \nabla_{\theta}L(\theta^i)$
  - 5:      $i = i + 1$
  - 6: end while
  - 7: Return:  $\theta^i$
-

En el algoritmo, el termino de tolerancia  $\epsilon$  hace referencia al tamaño aceptable para actualizar según la gradiente. En el caso que la gradiente calculada sea pequeña, se entiende que ya se ha alcanzado un mínimo local, y por ende ya no es necesario actualizar los parámetros. El resultado final es el conjunto de parámetros  $\theta^i$ .

### III.3.3.3 Gradiente descendente estocástica

El algoritmo Gradiente descendente estocástica es similar al de Gradiente descendente, sin embargo, en esta técnica, se elige de manera aleatoria un registro del conjunto de entrenamiento (o más de uno en caso de que sea especificado), y se replican los pasos del algoritmo de Gradiente descendente original. La gran diferencia radica en que el conjunto de parámetros  $\theta^i$  se actualiza con cada muestra de registros que el algoritmo utilice, a diferencia del proceso de Gradiente descendente, donde los parámetros se actualizan sobre la función de perdida de todo el conjunto de entrenamiento,  $L(\theta) = \sum_i^T L_i(\theta)$ , donde  $L_i(\theta)$  representa la función de perdida calculada para el registro  $i$ .

La actualización de los parámetros se hace de manera similar, buscando la orientación hacia donde la gradiente decrece. Además de lo anterior, en este algoritmo se debe especificar una función que converja a 0 para la tasa de aprendizaje o learning rate. Así, la actualización y una tasa de aprendizaje comúnmente utilizada, y el algoritmo según Smola y Vishwanathan (2008) son:

$$\theta^{i+1} = \theta^i - \eta_i \nabla_{\theta} L_i(\theta^i)$$
$$\eta_i = \sqrt{\frac{\tau}{\tau + i}}$$

---

Algoritmo de Gradiente Descendente Estocástica

---

- 
- 1: Inicializar  $\theta_o$ , definir máximo de iteraciones  $I$ , tamaño de actualización  $k$  y  $\tau$
  - 2: Definir  $i = 0$ :
  - 3: While  $i < I$ :
  - 4:     Elegir un conjunto de puntos de tamaño  $k$ , calcular  $\nabla_{\theta} L_i(\theta^i)$
  - 5:     Calcular  $\eta_i = \sqrt{\tau/\tau + i}$
  - 6:     Actualizar  $\theta^{i+1} = \theta^i - \eta_i \nabla_{\theta} L_i(\theta^i)$
  - 7:      $i = i + 1$
  - 8: Return:  $\theta^i$
- 

Es importante recalcar que la actualización se hace en base a la función de pérdida sobre el punto  $i$  de los datos, y no sobre la pérdida de todo el conjunto de entrenamiento. Es por este motivo que se necesita una tasa de crecimiento que converja a 0, ya que de lo contrario se estarían realizando actualizaciones de los parámetros muy grandes, de manera constante y sin garantías de converger a un resultado.

### III.3.3.4 Backpropagation

El algoritmo Backpropagation es una técnica que se utiliza para poder calcular los gradientes que se necesitan en el proceso de actualización de los parámetros de una red neuronal. En el caso de una red neuronal recurrente, el método es adaptado para tratar con la estructura secuencial que posee el algoritmo, denominándose Backpropagation through time (BPPTT). El objetivo es propagar mediante la regla de la cadena, los errores a través de la estructura. Werbos, Paul. (1990) describe este método como una extensión para aplicarse en sistemas dinámicos.

Las siguientes ecuaciones y cálculos son realizados en base Werbos, Paul. (1990), en una red neuronal simple, con una capa oculta,  $T$  time steps, y de forma many-to-one, es decir, posee la forma de la ilustración central de la figura 3. Además, se asume una red que posee los

parámetros  $W_{xh}$ ,  $W_{hh}$ ,  $b_h$ ,  $W_{hy}$  y  $b_y$ . Antes de ilustrar los cálculos, se definen las ecuaciones que describen a esta red:

$$h_t = f(W_{xh}x_t + W_{hh}h_{t-1} + b_h)$$

$$y = g(W_{hy}h_t + b_y)$$

Sea  $y$  el resultado que genera la red en la última unidad neuronal,  $L$  la función de pérdida, las ecuaciones que describen el método BPPTT son:

Parámetro  $b_y$ :

$$\frac{\partial L}{\partial b_y} = \frac{\partial L}{\partial y} \frac{\partial y}{\partial b_y}$$

Parámetro  $W_{hy}$ :

$$\frac{\partial L}{\partial W_{hy}} = \frac{\partial L}{\partial y} \frac{\partial y}{\partial W_{hy}}$$

Parámetro  $W_{hh}$ :

$$\frac{\partial L}{\partial W_{hh}} = \sum_{t=1}^T \frac{\partial L}{\partial y} \frac{\partial y}{\partial h_T} \frac{\partial h_T}{\partial h_t} \frac{\partial h_t}{\partial W_{hh}}$$

Notar que el termino  $\frac{\partial h_T}{\partial h_t}$  es expandible. Es decir, cuando  $t = T - 3$  se tiene que:

$$\frac{\partial h_T}{\partial h_{T-3}} = \frac{\partial h_T}{\partial h_{T-1}} \frac{\partial h_{T-1}}{\partial h_{T-2}} \frac{\partial h_{T-2}}{\partial h_{T-3}}$$

Esto se puede reescribir como:

$$\prod_{i=T-3}^{T-1} \frac{\partial h_{i+1}}{\partial h_i} = \frac{\partial h_T}{\partial h_{T-1}} \frac{\partial h_{T-1}}{\partial h_{T-2}} \frac{\partial h_{T-2}}{\partial h_{T-3}}$$

Con lo que, la forma para el parámetro  $W_{hh}$  queda como:

$$\frac{\partial L}{\partial W_{hh}} = \sum_{t=1}^T \frac{\partial L}{\partial y} \frac{\partial y}{\partial h_T} \left( \prod_{i=t}^{T-1} \frac{\partial h_{i+1}}{\partial h_i} \right) \frac{\partial h_t}{\partial W_{hh}}$$

Parámetro  $W_{xh}$ :

$$\frac{\partial L}{\partial W_{xh}} = \sum_{t=1}^T \frac{\partial L}{\partial y} \frac{\partial y}{\partial h_T} \left( \prod_{i=t}^{T-1} \frac{\partial h_{i+1}}{\partial h_i} \right) \frac{\partial h_t}{\partial W_{xh}}$$

Parámetro  $b_h$ :

$$\frac{\partial L}{\partial b_h} = \sum_{t=1}^T \frac{\partial L}{\partial y} \frac{\partial y}{\partial h_T} \left( \prod_{i=t}^{T-1} \frac{\partial h_{i+1}}{\partial h_i} \right) \frac{\partial h_t}{\partial b_h}$$

Con las ecuaciones anteriores se realiza la actualización en contra de la gradiente del parámetro. Así, sea  $l$  el índice que indica el numero en la iteración del algoritmo de gradiente descendente, las ecuaciones que resumen las actualizaciones son:

$$b_y^{l+1} = b_y^l - \eta_l \frac{\partial L}{\partial b_y}$$

$$W_{hy}^{l+1} = W_{hy}^l - \eta_l \frac{\partial L}{\partial W_{hy}}$$

$$W_{hh}^{l+1} = W_{hh}^l - \eta_l \frac{\partial L}{\partial W_{hh}}$$

$$W_{xh}^{l+1} = W_{xh}^l - \eta_l \frac{\partial L}{\partial W_{xh}}$$

$$b_h^{l+1} = b_h^l - \eta_l \frac{\partial L}{\partial b_h}$$

### III.3.4 Hiperparametros

Una red neuronal recurrente posee varios hiperparametros. A continuación, se enumeran algunos de ellos y se describe su función dentro de la red neuronal.

#### Learning rate o tasa de aprendizaje:

Este hiperparametro hace referencia al termino que acompaña a la gradiente en el proceso de actualización de los parámetros vía gradiente descendente. En el método de gradiente descendente estocástica, este término es decreciente a medida que se realizan las iteraciones, buscando que este converja a 0.

Este término es relevante para poder generar buenos resultados, un learning rate muy grande puede provocar que el algoritmo nunca converja, y una tasa de aprendizaje muy baja puede provocar que no se alcance nunca un mínimo local.

#### Función de activación:

La función de activación hace referencia a la operación que se realiza para generar un hidden state, o bien el resultado final de la red. En general, las funciones más utilizadas son: tangente hiperbólica (*tanh*), función sigmoide (*sigm*), función rectificadora (*Relu*) y función lineal. Esta ultima consiste en no aplicar ninguna transformación. A continuación, se explicita las fórmulas de las funciones anteriormente mencionadas, y el dominio y recorrido de estas:



$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}, -\infty < x < \infty, -1 < y < 1$$

$$\text{sigm}(x) = \frac{1}{1 + e^{-x}}, -\infty < x < \infty, 0 < y < 1$$

$$\text{Relu}(x) = \max(0, x), -\infty < x < \infty, 0 < y < \infty$$

### Time Steps/Unidades neuronales:

El número de unidades neuronales hace referencia a cuál será el tamaño de la secuencia de datos con los que pretende ajustar la red. Este hiperparámetro dependerá de la forma en que se estructuren los datos. En la figura 3, en la ilustración central se tiene una RNN con 3 time steps como estructura.

### Hidden Layer/Capas ocultas:

El número de capas ocultas hace referencia a la cantidad de niveles donde existen tantas unidades neuronales como time steps estén definidos.

### Unidades:

El número de unidades hace referencia a la dimensión que posee los hidden states, así, sea la RNN descrita por las siguientes ecuaciones:

$$h_t = f(W_{xh}x_t + W_{hh}h_{t-1} + b_h)$$

$$y_t = g(W_{hy}h_t + b_y)$$

Si se define el parámetro de unidades como  $m$ , con  $m \in \mathbb{Z}^+$ , entonces el  $\dim(h_t) = m \times 1$ . Además, define las dimensiones de las matrices, tal que:  $W_{xh} \in M_{m \times n}(\mathbb{R})$  y  $W_{hh} \in M_{m \times m}(\mathbb{R})$ .

Épocas:

El número de épocas hace referencia a cuantas veces se recorre completamente el conjunto de entrenamiento mientras se actualizan los parámetros internos de la red neuronal.

## III.4 Preprocesamiento de una variable

### III.4.1 Normalización

La normalización de una variable en algunos algoritmos es necesaria para mejorar su desempeño. Brownlee (2017) señala que cuando una red neuronal se ajusta a datos que no se encuentran escalados, los valores más grandes pueden distorsionar el proceso de aprendizaje del algoritmo. Una de las transformaciones más utilizadas es la normalización. Esto consiste en reescalar una variable para que su nueva escala este entre 0 y 1. La fórmula para realizar esta transformación es:

$$x_t^R = \frac{x_t - \min(x_t)}{\max(x_t) - \min(x_t)}$$

Luego que el algoritmo genere un resultado, para volver a la escala original, simplemente se hace la inversa de la transformación. Sea  $x_t^R$ , el resultado que genera el algoritmo con datos escalados, entonces para poder transformar dicho valor a la escala original de la variable se aplica la siguiente formula:

$$x_t = x_t^R \max(x_t) - \min(x_t) + \min(x_t)$$

### III.4.1 Estandarización

Otra transformación utilizada es la estandarización. Este proceso consiste en sustraer la media, y luego dividir por la desviación estándar muestral de la variable que se esté utilizando. Matemáticamente:

$$x_t^R = \frac{x_t - \mu}{\sigma}$$

Donde  $\mu$  y  $\sigma$  son la media y la desviación estándar de la variable  $x_t$ .

## III.5 Métricas de error

### III.5.1 Error cuadrático medio (Mean squared error, MSE)

Una métrica de error comúnmente utilizada es el error cuadrático medio, o mean squared error. Este método para calcular los errores computa la diferencia entre la predicción y el valor real, luego se calcula el cuadrado de esta diferencia y se suman todos los errores para cada una de las proyecciones. Luego se divide por la cantidad de puntos para obtener la media. Matemáticamente, sea  $y_t$  e  $\widehat{y}_t$  el valor real y la predicción respectivamente, y sea  $n$  el número de puntos sobre el que se calcula el error cuadrático medio, la fórmula que describe esta métrica está descrita por:

$$MSE = \frac{1}{n} \sum_{t=1}^n (y_t - \widehat{y}_t)^2$$

Es importante recalcar que la escala de esta métrica es distinta a la de la variable  $y_t$ , más específicamente, la escala del MSE está al cuadrado. Por otro lado, el rango de valores del error cuadrático medio está entre 0 y  $\infty$ .

### III.5.2 Error absoluto medio (Mean absolute error, MAE)

El error absoluto medio es otra las métricas comúnmente utilizadas. El cálculo consiste en aplicar la diferencia entre el valor real de la variable y la proyección calculada. Posteriormente se calcula el valor absoluto de la diferencia en cada punto, para luego sumar todos los errores calculados. Finalmente, al igual que en MSE, se divide por la cantidad de puntos para obtener el promedio. Matemáticamente, sea  $y_t$  e  $\widehat{y}_t$  el valor real y la predicción respectivamente, y sea  $n$  el numero de puntos sobre el que se calcula el error absoluto medio, la fórmula que describe esta métrica esta descrita por:

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \widehat{y}_t|$$

El rango de esta métrica, al igual que MSE, se encuentra entre 0 y  $\infty$ . Una gran ventaja de esta métrica con respecto al error cuadrático medio, es que el error absoluto medio se encuentra en la misma escala de la variable  $y_t$ .

### III.5.3 Raíz del error cuadrático medio (Root mean squared error, RMSE)

La raíz del error cuadrático medio (RMSE por sus siglas en inglés) es una métrica comúnmente utilizada en el análisis de problemas de regresión. Se obtiene calculando la raíz cuadrada del error cuadrático medio. Matemáticamente es escribe:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \widehat{y}_t)^2}$$

Es importante recalcar, que, a diferencia del MSE, esta métrica se encuentra en la escala original de la variable  $y_t$ . Por otro lado, el rango de la raíz del error cuadrático medio se encuentra entre 0 y  $\infty$ .

Jierula, Alipujiang, Wang, Shuhong y Oh, Tae-Min. (2021) señalan que la dependencia de estas métricas en la escala, generan dificultades al momento de comparar series de datos que se encuentran en medidas distintas, sin embargo, son muy usadas para calcular la evaluación

en algoritmos de aprendizajes de máquinas. Por otro lado, también señalan que la métrica RMSE “penaliza” más que las métricas MAE y MSE las grandes desviaciones en una serie.

### III.5.4 Raíz del error cuadrático medio normalizada (NRMSE)

Una de las principales desventajas de las tres métricas anteriores, es que son dependientes de la escala de la variable en estudio, motivo por el cual, utilizando dichas métricas, solo es posible evaluar el desempeño de diferentes modelos en una misma serie de tiempo. En el trabajo realizado por Atiya et al. (1999), utilizan redes neuronales recurrentes, midiendo el rendimiento en ciertas configuraciones utilizando la métrica RMSE normalizada, lo que permite comparar el desempeño entre distintas variables. Matemáticamente se escribe de la siguiente manera:

$$NRMSE = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n y_t^2}}$$

# Capítulo IV – Datos

## IV.1 Descripción de datos disponibles

Los datos disponibles corresponden a las boletas diarias de 38 tiendas físicas, las cuales abarcan las ventas realizadas entre enero de 2015 y diciembre de 2020, sin embargo, debido a la influencia de las cuarentenas desde 2020, es que para el presente estudio se decide trabajar con los datos correspondientes desde enero de 2015 hasta marzo de 2020, un mes antes que comenzaran las restricciones al comercio. Por otro lado, se decide trabajar con 5 locales en específico, debido a que presentan una mayor completitud de los datos en el periodo de estudio mencionado anteriormente. Estos 5 puntos de ventas representan cerca del 26% de las ventas totales. Una vez realizada la limpieza de datos la cual será descrita en la próxima sección, existen un total de 350.537 boletas en el conjunto en estudio, lo cual corresponde a la venta de 733.728 unidades. Por otro lado, en los datos que se disponen, existen 93 marcas que fueron vendidas, las cuales corresponden a 1.391 productos distintos. A continuación, se hace una breve descripción de las variables que se disponen:

- Local: Esta columna posee el id de la tienda donde se realizó la venta. Para la presente investigación estos serán identificados por las letras: A, B, C, D y E.
- Nombre de producto: Esta columna posee el nombre del artículo y generalmente la cantidad que este trae.
- Fecha: Fecha en la que se vendió el producto en formato año/mes/día.
- Precio: El valor monetario por el que se vendió el producto, en pesos.
- Folio: Id de la boleta en la que se registró la venta del producto.
- Marca: La marca del producto.
- Grupo 1: Esta columna hace referencia a que categoría pertenece el producto dentro del grupo 1, donde existen 6 posibles valores.

- Grupo 2: Esta columna hace referencia a que categoría pertenece el producto dentro del grupo 2, donde existen 30 posibles valores, es decir, es una descripción más detallada que los niveles que ofrece la variable grupo 1.
- Grupo 3: Esta columna hace referencia a que categoría pertenece el producto dentro del grupo 3, donde existen 91 posibles valores, por lo que es aún más precisa en cuanto a la naturaleza del producto.
- Hora: Esta columna hace referencia a la hora y minuto en que fue vendido el producto.

Con el propósito de ilustrar la información que se tiene, se procede a analizar más en profundidad las variables que se disponen (este análisis se realiza con los datos después de ser procesados):

### Precio:

En los 5 locales que se está estudiando, de los productos vendidos, el más caro es de \$99.990, y el más barato tiene un valor de \$490, por otro lado, el precio medio del total de artículos transados es de \$12.716.

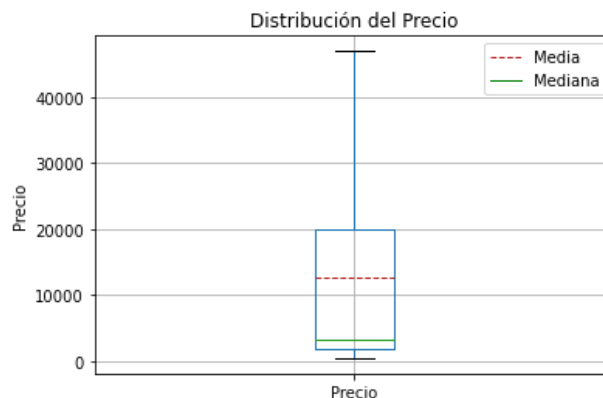


Gráfico 3: Distribución del precio de los artículos vendidos

En el gráfico 3 se puede apreciar la distribución del precio de los productos transados para los 5 locales entre enero de 2015 y marzo de 2020. Se puede apreciar que los primeros dos

cuartiles se encuentran más concentrados que los últimos dos, siendo el ultimo cuartil el que abarca el rango más amplio (entre \$19.990 y \$99.9900).

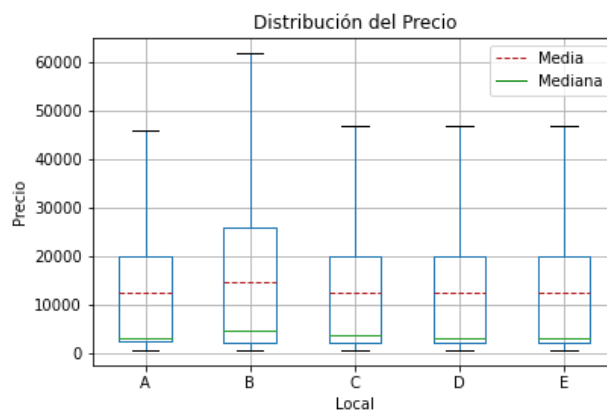


Gráfico 4: Distribución del precio de los artículos vendidos en los 5 locales en estudio

En el gráfico 4 se puede apreciar la distribución del precio de los artículos transados, para cada uno de los 5 locales. La media del precio en los 5 locales es relativamente similar, siendo el local B el que presenta un mayor valor (\$14.549). La distribución del precio de los artículos vendidos se resume en la tabla 4.

Local	Precio medio(\$)	25%	50%	75%	Max
A	12.017	2.390	2.990	19.990	84.990
B	14.370	1.990	4.490	25.990	84.990
C	12.374	1.990	3.490	19.990	84.990
D	12.193	1.990	3.000	19.990	84.990
E	12.172	1.990	3.000	19.990	99.990

Tabla 4: Distribución del precio de los productos vendidos para cada uno de los locales



## Unidades vendidas:

Desde enero de 2015 hasta marzo de 2020, el total de unidades vendidas en los 5 locales es de 754.765. Para los locales A, B, C, D y E, las unidades vendidas fueron 173.809, 116.177, 70.329, 120.425 y 259.988 respectivamente, siendo el local C el que presento un menor nivel de ventas de las 5 tiendas.

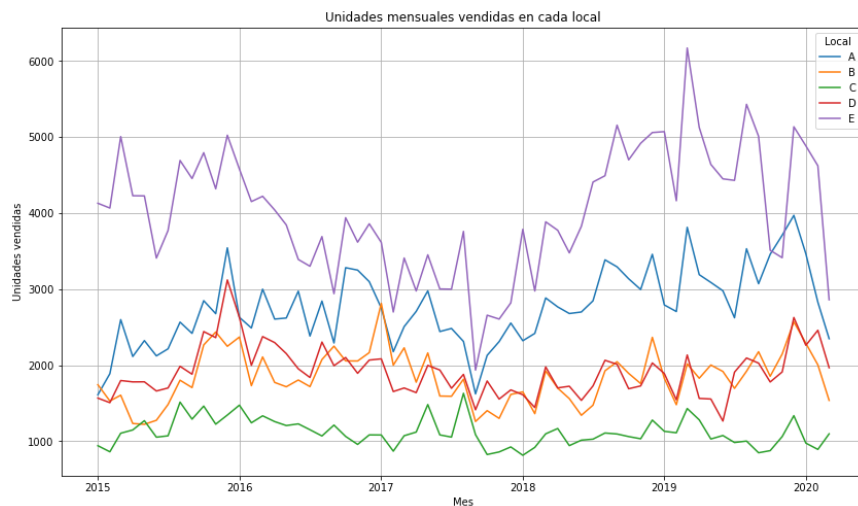


Gráfico 5: Unidades vendidas por mes, para los 5 locales, desde enero de 2015 hasta marzo de 2020

En el gráfico 5 podemos observar lo mencionado anteriormente. El local E es el que más ventas posee a lo largo del periodo de estudio, seguido por el local A. La tercera y cuarta posición se superpone dependiendo del mes los locales B y D. Por otro lado, el local C es el que presenta un menor nivel ventas comparado con el resto de los locales en casi toda la ventana temporal.

Con el propósito de comparar el comportamiento de ventas de las tiendas, se calculó la correlación de Pearson, dando como resultado que las tiendas B y D son las que presentan un mayor grado de correlación (0.63), por otro lado, los locales A y C son los que presentaron el menor valor dentro de los locales que se encuentran en el grupo de estudio (0.19). El detalle de la matriz de correlaciones se puede apreciar en el gráfico 6.

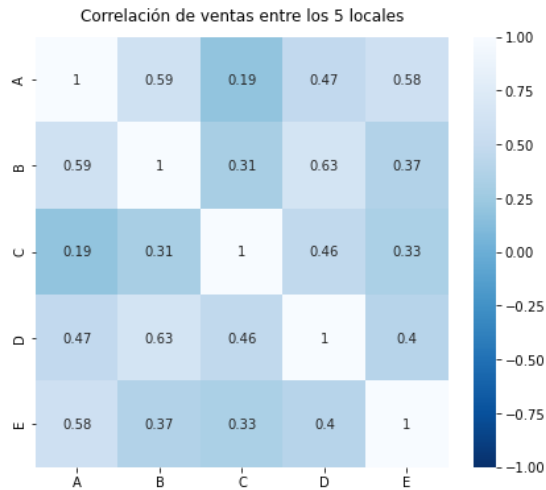


Gráfico 6: Correlación de ventas en los 5 locales en estudio

Por otro lado, en el caso de la variable Grupo 1, la cual agrupa 6 categorías distintas de productos (Accesorio, Deporte, Salud, Dieta, Consumible y Belleza y Cuidado Personal), el tipo de producto que más se vendió fueron los del tipo Consumible (365.891), seguido por los tipos Deporte (147.184), Salud (107.093), Dieta (55.113) y Accesorio (49.478). Por otro lado, los productos categorizados como de Belleza y Cuidado Personal fueron los que menos unidades vendieron dentro del conjunto de estudio (8.969). El comportamiento mensual de las ventas de estas categorías se puede observar en el gráfico 7.

Unidades mensuales vendidas por categoría del Grupo 1

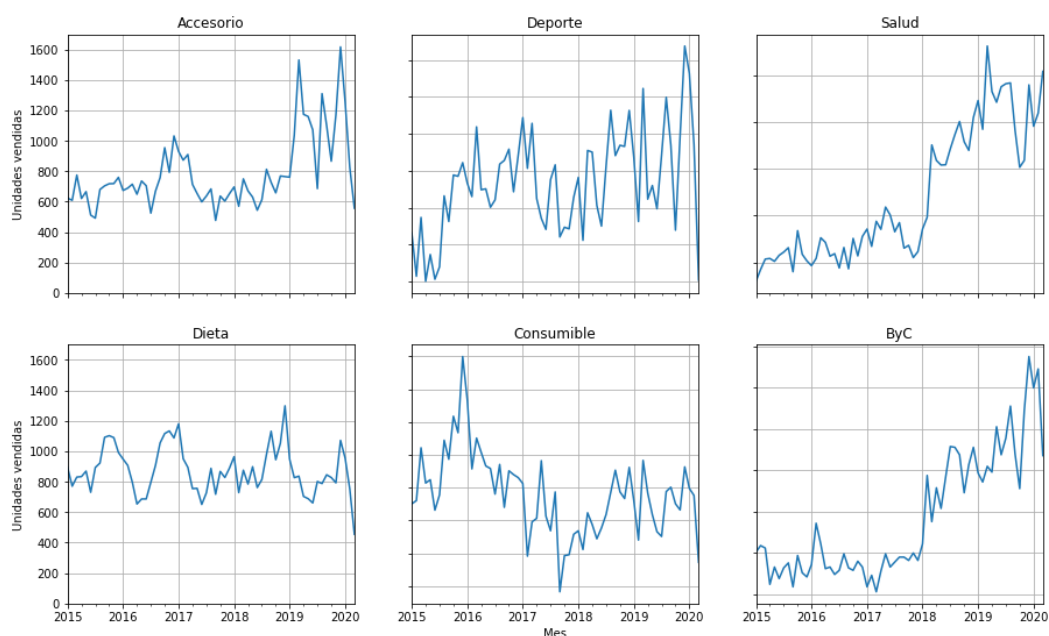


Gráfico 7: Unidades vendidas por mes, para las 6 categorías de la variable Grupo 1, desde enero de 2015 hasta marzo de 2020

En la sección Anexos se adjunta la información de las ventas totales para las categorías en las variables Grupo 2 y Grupo 3.

Al igual que en la demanda de los locales, se calculó la correlación de Pearson entre estas 6 categorías, dando como resultado que la combinación de productos categorizados como de Belleza y Cuidado Personal-Salud son aquellos que presentan un comportamiento en términos lineales, muy alto (0.87), seguido de la combinación Deporte-Accesorio (0.66). Por otro lado, las categorías que presentan un menor grado de correlación (cercano a 0), son las combinaciones Accesorio-Consumible (0.15) y Accesorio-Dieta(0.16). En el gráfico 8 se explicita de manera detallada esta información.

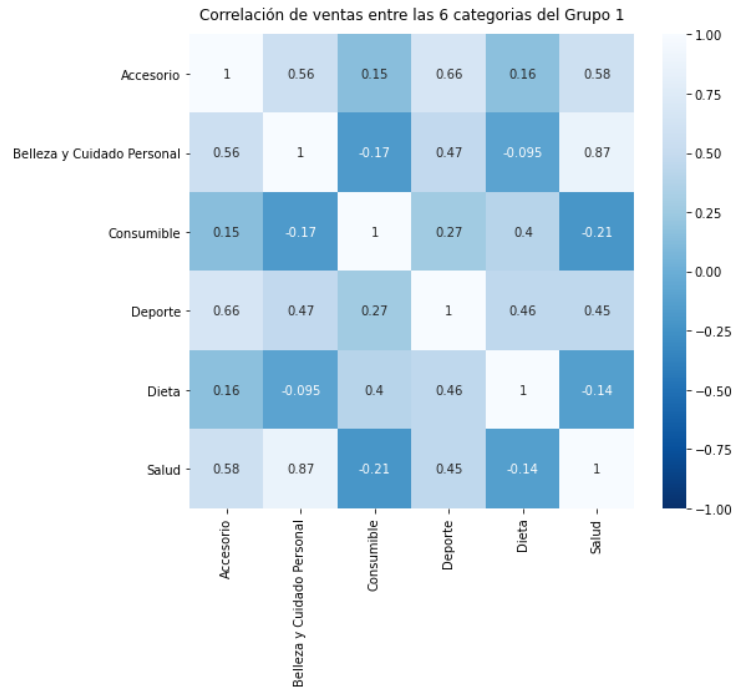


Gráfico 8: Correlación de ventas entre las 6 categorías del Grupo 1

En el caso de las 20 categorías que se estudian del Grupo 2 (en la siguiente sección se describe esta selección), existen 15 combinaciones que poseen una correlación de Pearson fuerte (en términos absolutos, mayor a 0.8). En la tabla 5 se detalla las tuplas de categorías con una correlación superior a 0.8.

Tupla	Correlación
Multivitamínicos /Nutricosmética	0.801
Proteínas Veganas/Proteínas Aisladas	0.808
Cuidado Óseo/Proteínas Aisladas	0.818
Proteínas Veganas/Defensas	0.819
Probióticos/Salud Cardiovascular	0.822
Cuidado Óseo/Defensas	0.833
Cuidado Óseo/Probióticos	0.851
Superalimento/Cuidado Óseo	0.852

Defensas/Multivitamínicos	0.861
Multivitamínicos/Proteínas Veganas	0.864
Multivitamínicos/Cuidado Óseo	0.868
Nutricosmética/Cuidado Óseo	0.892
Salud Cardiovascular/Cuidado Óseo	0.900
Proteínas Veganas/Cuidado Óseo	0.902
Proteínas Veganas/Nutricosmética	0.942

Tabla 5: Tuplas de categorías pertenecientes al Grupo 2 con una correlación superior a 0.8

En la sección Anexos se adjunta la información detallada de las correlaciones de las 20 categorías del Grupo 2. Para el caso de los 67 niveles del Grupo 3 restantes una vez que se realiza la limpieza de datos, no es posible calcular la matriz de correlación de forma completa, teniendo que reducirse a 48 categorías. Los detalles se encuentran en la sección Anexos.

### Boletas:

En el periodo que abarca desde enero de 2015 hasta marzo de 2020, en los 5 locales se transaron 350.537 boletas, donde el valor promedio de estas fue de \$26.616, y el valor medio de unidades en cada una de estas es de 2.09, siendo un total de 147.434 boletas donde se compraron dos o más productos. En el gráfico 9 y la tabla 5 se detalla esta información. En la tabla 5 además se incorpora la distribución del precio y de unidades en cada una de las boletas.

Boleta	Media	25%	50%	75%	Max
Valor (\$)	26.616	4.990	21.480	35.990	933.690
Unidades	2.09	1	1	2	132

Tabla 5: Distribución de las unidades y valor de las boletas

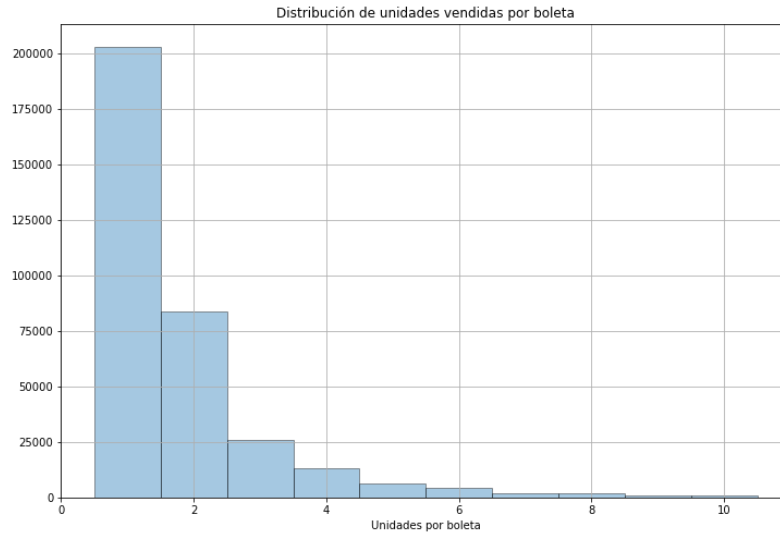


Gráfico 9: Distribución del número de unidades vendidas por boleta

## IV.2 Limpieza y preprocesamiento de datos

El enfoque que se le da a la limpieza de los datos disponibles es la de poder generar un conjunto de series de tiempo lo más homogéneo posible. En ese sentido, uno de los objetivos prioritarios es la de seleccionar series que se encuentren “completas” desde enero de 2015 hasta marzo de 2020, en termino de unidades mensuales vendidas, con foco en la selección de categorías del Grupo 2, debido a que estas presentan un equilibrio entre nivel de desagregación y completitud de datos. Es por ello que las series de tiempo que se trabajaran son aquellas correspondientes a las categorías que existen en la variable Grupo 1, y un subconjunto del total de categorías del Grupo 2, cuya selección será descrita más adelante.

La base de datos disponible posee las ventas realizadas en 5 locales, registrando originalmente un total de 806.272 unidades vendidas, correspondientes a 379.320 boletas, 1.864 productos distintos y 130 marcas diferentes. A continuación, se describen los pasos que se realizan en una primera etapa con respecto a la limpieza de datos:

- 1) Se registran un total de 2.666 productos devueltos dentro de la base de datos, representando cerca de un 0.3% del total de artículos vendidos. Debido a que

representa un porcentaje menor del total, se eliminan los registros de productos devueltos.

- 2) Dentro de la base de datos existen 117 unidades que fueron vendidas que no presentan ninguna categorización en las variables Grupo 1, Grupo 2 y Grupo 3. Estas 117 unidades corresponden a 5 productos distintos, donde el precio mínimo y máximo de estos es de \$1.500 y \$14.990. Se eliminan estos productos debido a que no existe forma de asignarlo a alguna de las categorías que se estudiarán.
- 3) Existen 636 registros que corresponden a boletas anuladas. Se eliminan de la base de datos.
- 4) Existen 5.574 unidades vendidas, que corresponden a 29 productos distintos, con un precio que oscila entre \$1 y \$150. Entre ellos se encuentran poleras, muestras y bolsas de papel. Se eliminan estos productos debido a que los registros comienzan desde abril de 2015.
- 5) La base de datos incluye información hasta septiembre de 2020. Se eliminan todos los registros que se encuentren fuera de la ventana temporal comprendida entre enero de 2015 y marzo de 2020. Esto corresponde a 33.281 unidades vendidas.
- 6) A continuación, se describen las categorías pertenecientes al Grupo 2 que son eliminadas:
  - a. Cuidado Corporal: Esta serie no posee información entre junio de 2015 y abril de 2019, motivo por el cual es eliminado de la base de datos.
  - b. Digestión: Esta categoría se decide eliminar debido a que, en los locales B y C, en algunos meses, hay niveles de ventas inferiores a las 3 unidades, y meses sin registros de ventas.
  - c. Proteínas de Caseína: En esta serie existen locales que solo poseen información hasta enero de 2020. Por otro lado, en el local C existen 10 meses sin ningún registro, motivo por el cual es eliminado.
  - d. Soporte para práctica deportiva: Esta categoría también es eliminada debido a los niveles de ventas.
  - e. Otros Accesorios: En el caso de esta categoría, para los locales B, C y D existen más de 20 meses faltantes en los registros.

- f. Alimentos enriquecidos: En el local C hay 16 meses faltantes de información, motivo por el cual se decide eliminar esta categoría.
- g. Barras y Snack Saludables: En el caso de esta categoría hay 13 meses faltantes en el local A, y 11 meses en los locales C y D, motivo por el cual se decide eliminar esta categoría.
- h. Aromaterapia: En los 5 locales no existen registros de ventas en 2015, 2016 y 2017, motivo por el cual se decide eliminar esta categoría.
- i. Té e infusiones: En los 5 locales no existe información correspondiente a los años 2015 y 2016, motivo por el cual se elimina esta categoría.
- j. Cuidado Facial: En los 5 locales no existen registros de ventas en 2015 y 2016, motivo por el cual se decide eliminar esta categoría.

En el caso de las categorías del Grupo 3, de las 455 series de tiempo que existen (considerando los 91 niveles), solo 221 poseen información completa (desde enero de 2015 hasta marzo de 2020), es por este motivo que se decide estudiar los niveles seleccionados de la categorización Grupo 2, además de los correspondientes al Grupo 1.



# Capítulo V – Metodología

## V.1 Modelos clásicos de series de tiempo

Para los modelos clásicos de series de tiempo (suavización exponencial y ARIMA), la forma en organizar los datos es idéntica. En cada una de las 130 series que se estudian, que corresponden al cruce entre categorías y locales, se estructuran los datos de manera mensual, desde enero de 2015 hasta marzo de 2020, dando como resultado un total de 63 meses en estudio.

Para cada una de las 130 series, antes de proceder a la división de los datos en conjuntos de entrenamiento y testeo, se rellena con ceros los meses en los que no existieron registros de ventas, es decir, se asume que durante todo el periodo siempre hay existencias de los productos de la categoría, y, por ende, la no compra de artículos en un determinado mes se debe a que no existió demanda. En total, 11 series de las 130 son las que necesitaron ser completadas utilizando esta metodología, siendo la categoría Probióticos perteneciente al local B donde se hace un mayor número de reemplazos (7 en total). En la tabla 6 se detalla la información con los locales y categorías que necesitaron de esta metodología.

Tienda	Categoría	#Reemplazos
A	Probióticos	1
A	Prot. Vegana	2
B	Probióticos	7
B	Prot. Vegana	4
C	Probióticos	2
C	Prot. Vegana	3
D	Prot. Vegana	3
D	Bebida Funcional	2

E	Probióticos	2
E	Geles	1
E	Prot. Vegana	3

Tabla 6: Numero de meses en los que no hubo demanda en cada serie de tiempo

Una vez estructurados las 130 series de tiempo, se procede a dividir los datos en grupos de entrenamiento y de testeo. El primer grupo, está conformado por los primeros 51 meses de los datos, es decir, desde enero de 2015 hasta marzo de 2019, incluyendo el último mes. Por otro lado, el conjunto de testeo consta de los últimos 12 meses del conjunto de estudio, los cuales abarcan la ventana temporal entre abril de 2019 y marzo de 2020, ambos meses inclusive (el grupo de entrenamiento representa cerca del 80% de los datos, mientras que el de testeo aproximadamente el 20%).

El conjunto de entrenamiento se utilizará para ajustar los parámetros de los modelos, y el de testeo para evaluar el desempeño de los pronósticos de cada algoritmo. Es importante señalar que la selección posterior del mejor modelo para cada serie de tiempo debe realizarse en base al desempeño en datos fuera de la muestra, es decir, el conjunto de testeo. Con respecto al tamaño de división de cada conjunto seleccionado, es decir, 80-20, Hyndman y G.Athanasopoulos (2013) señalan que no existe una división exacta recomendable para estos problemas, además que la división debe realizarse en virtud de la cantidad de datos disponibles, y que el tamaño del conjunto de testeo debe ser idealmente no más grande que el horizonte de tiempo que se pretende pronosticar.

División de conjunto de entrenamiento y testeo para distintas series

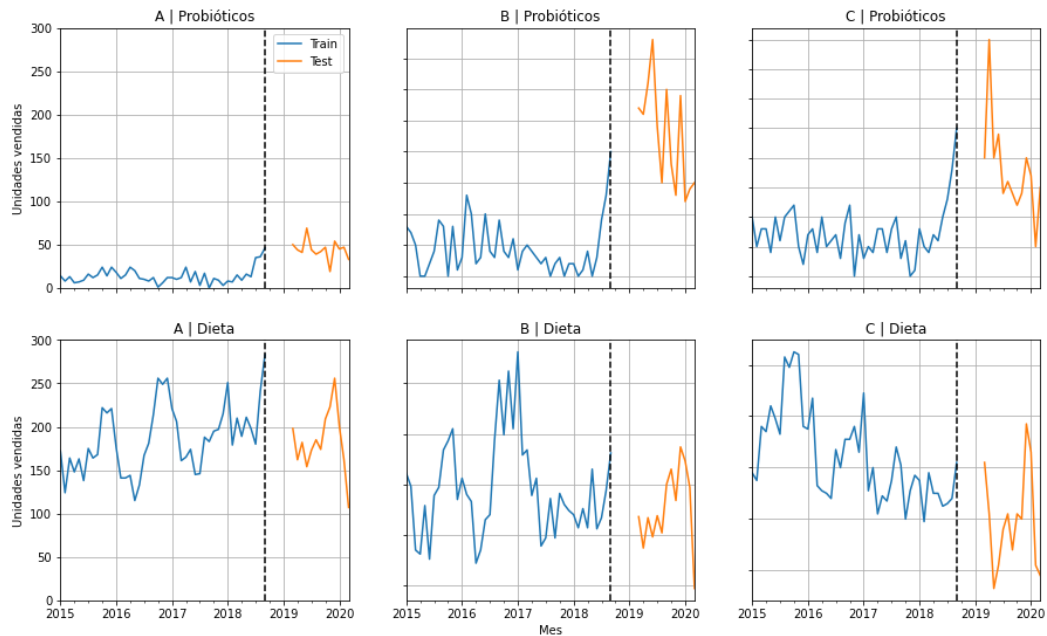


Gráfico 10: División en grupos de entrenamiento y testeo para 6 series de tiempo

En el gráfico 10 se puede apreciar 6 series de tiempo, donde cada una representa a la categoría Probióticos o Dieta, y a los locales A, B o C. En esta ilustración se puede observar de manera visual como es la división en conjuntos de entrenamiento y testeo.

### V.1.1 Suavización exponencial simple

La metodología usada en el modelo de suavización exponencial simple consiste en utilizar los primeros 51 meses correspondientes al conjunto de entrenamiento, donde se ajusta el modelo, es decir, se encuentra el parámetro  $\alpha$  descrito en el Marco Teórico, mediante la minimización de la suma de los errores cuadráticos en el conjunto de entrenamiento. Para ello se utiliza el paquete estadístico statsmodel en el lenguaje de programación Python.

Una vez encontrado el parámetro  $\alpha$ , se proyecta a 12 meses, es decir, desde abril de 2019 hasta marzo de 2020. Es importante recordar que, en un modelo de suavización exponencial

simple, las proyecciones son una curva “plana”. Es decir, sea  $T$  el último periodo dentro del conjunto de entrenamiento, la proyección en  $T + 1$  es la misma que en  $T + h$ .



Gráfico 11: Proyección de suavización exponencial simple en 3 series de tiempo distintas

En el gráfico 11 se puede apreciar 3 series de tiempo denotadas por las siglas  $X|Y$ , donde  $X$  hace referencia al nombre del local, e  $Y$  es la categoría que se está estudiando. Se eligen tres series con el propósito de ilustrar la división que se propone en la metodología (en la sección de Resultados se encuentran los resultados para todas las series en estudio). Las series representadas en la ilustración son:  $A|GanadoresDePeso$ ,  $B|ProteinasAisladas$  y  $D|Dieta$ . En el gráfico se muestra, dividido por una línea vertical discontinua, los conjuntos de entrenamiento y testeo, además de la proyección que se realiza utilizando la suavización exponencial simple. Una vez realizada la proyección, se procede a calcular las métricas en el conjunto de testeo: MAPE y RMSE, además del error según el mes que se está proyectando a futuro. Estas serán analizadas con mayor profundidad en el siguiente capítulo. En el gráfico también se desprende el valor del parámetro  $\alpha$  encontrado para cada una de las 3 series: 0.77, 0.33 y 0.63. Esto quiere decir, que en las series  $A|GanadoresDePeso$  y  $D|Dieta$ , el método de suavización exponencial simple le otorga mayor importancia a la información más reciente con respecto a la serie  $B|ProteinasAisladas$ .

### V.1.2 Suavización exponencial doble (Holt's)

La metodología utilizada en el método de Holt's es idéntica a la usada en el modelo de suavización exponencial simple, es decir, se utiliza la información desde enero de 2015 hasta marzo de 2019 para ajustar el modelo, en este caso, hallar los parámetros  $\alpha$  y  $\beta$  que son descritos en el Marco Teórico, y luego se proyecta desde abril de 2019 hasta marzo de 2020. Al igual que en la metodología anterior, se utiliza el paquete estadístico statsmodel en Python, donde se minimiza la suma de los errores cuadráticos, es decir, sea  $y_t$  el valor real,  $y_{t|t-1}$  la estimación realizada por el modelo, entonces el paquete estadístico minimiza  $\sum_{t=1}^T (y_t - y_{t|t-1})^2$ .

Con el propósito de ilustrar el comportamiento del algoritmo, en el gráfico 12 se adjuntan las 3 series presentadas en el método de suavización exponencial simple, con la proyección realizada por esta metodología.



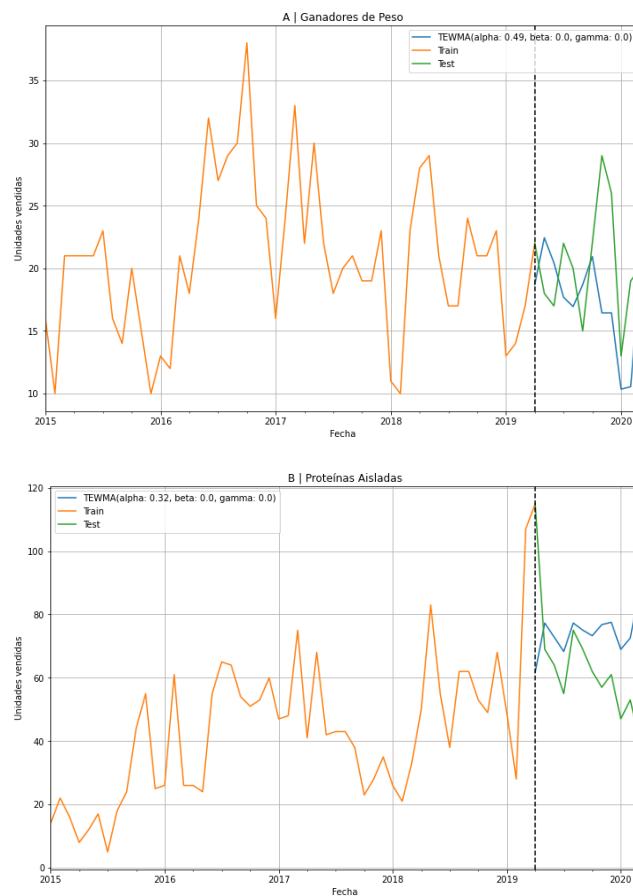
Gráfico 12: Proyección de suavización exponencial doble en 3 series de tiempo distintas

Es importante recalcar que el segundo parámetro que posee el método de Holt's intenta capturar si es que la serie posee una tendencia, ya sea creciente o decreciente. Esto se puede apreciar en la serie *B|Proteínas Aisladas*, donde el método interpreta que existe una tendencia creciente. Por otro lado, se utiliza la configuración lineal del método de Holt's, es

decir, este método asume que la tendencia de las series en cuestión no posee un comportamiento exponencial.

### V.1.3 Suavización exponencial triple (Holt's/Winters)

La metodología utilizada en el método de Holt's y Winters es idéntica a la utilizada en los dos métodos anteriores. Se utilizan los primeros 51 meses de cada una de las series para encontrar los parámetros  $\alpha$ ,  $\beta$  y  $\gamma$ , para luego proyectar sobre los siguientes 12 meses y así generar las métricas MAE y RMSE, además de los errores según la distancia mensual a la que se esté proyectando. En el gráfico 13 se puede apreciar el comportamiento de este modelo con respecto a las mismas 3 series presentadas anteriormente.



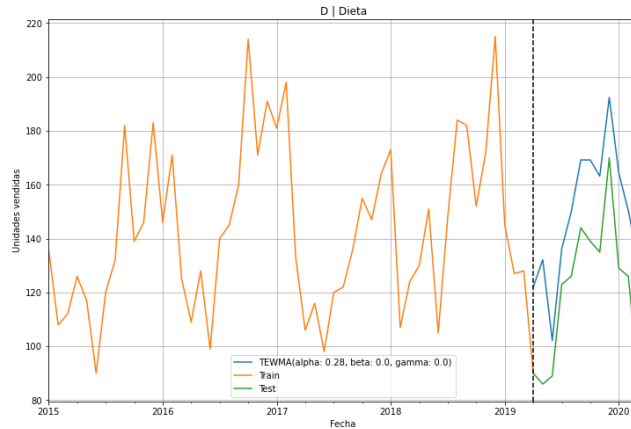


Gráfico 13: Proyección de suavización exponencial triple en 3 series de tiempo distintas

Este método consta de 3 parámetros  $\alpha$ ,  $\beta$  y  $\gamma$ , los cuales capturan el nivel, la tendencia y estacionalidad respectivamente. En este método se fija la frecuencia de estacionalidad en 12, que es lo que se suele utilizar para datos estructurados en meses (Hyndman y G.Athanasopoulos, 2013). Por otro lado, para todas las series en estudio, se define utilizar la componente aditiva para la estacionalidad, es decir, se asume que las variaciones estacionales son relativamente constantes durante toda la serie, y la componente aditiva para la tendencia, es decir, que la serie posee una tendencia de crecimiento lineal.

En el gráfico 13 se puede apreciar que, a diferencia de los dos métodos anteriores, el modelo de suavización exponencial triple, a priori, posee un mejor ajuste. Esto se analizará con mayor profundidad en la siguiente sección.

## V.1.4 ARIMA

Una de las metodologías que se suele usar en el análisis de series de tiempo con el propósito de encontrar  $p$ ,  $d$  y  $q$  (los tres parámetros que posee el modelo ARIMA), es a través de una inspección visual en los gráficos correspondientes a la autocorrelación y autocorrelación parcial de la serie en estudio.



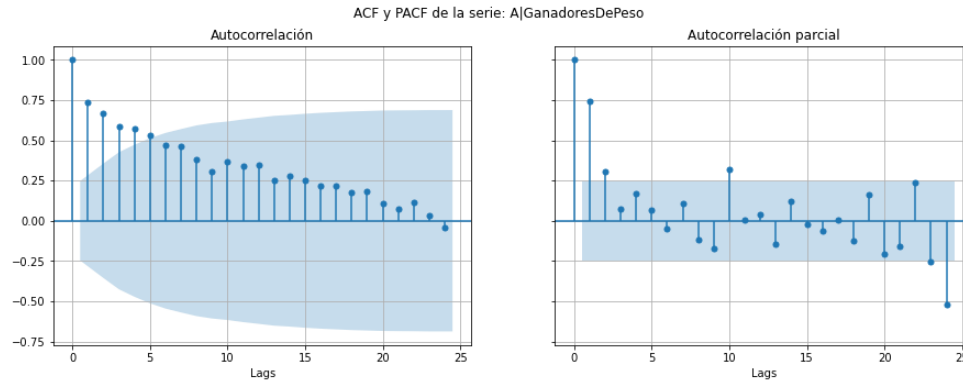


Gráfico 14: Autocorrelación y Autocorrelación parcial para la serie de los productos Ganadores de Peso del local A

El gráfico 14 corresponde a los mapas de autocorrelación y autocorrelación parcial para la serie de productos pertenecientes a la categoría “Ganadores de peso” en el local A. Los límites en azul hacen referencia al intervalo de confianza (en este caso al 95%) con el que se estimaron los datos de correlación.

Hyndman y G.Athanasopoulos (2013), señalan que los gráficos de ACF y PACF deben ser analizados en conjunto. Un patrón que detectaría un posible modelo  $ARIMA(p, d, 0)$ , es el de observar un decaimiento exponencial o sinusoidal en el gráfico de autocorrelación, y que al mismo tiempo en el de PACF exista un valor alto en el lag  $p$ , pero ninguno significativo después de este. En el caso de la serie del gráfico 14, el PACF muestra un decaimiento brusco después de  $p = 2$  (y con ello un modelo  $ARIMA(2, d, 0)$ ). Por otro lado, para detectar potenciales modelos  $ARIMA(0, d, q)$ , se debe cumplir los siguientes patrones: el gráfico de PACF decae exponencial o sinusoidalmente, y, existe un valor significativamente alto en el lag  $q$  del gráfico de ACF, pero ninguno después de este. En base al gráfico 14, la ACF no presenta valores claros que cumplan con la segunda condición, pero una alternativa podría ser definir  $q = 1$ , dando así un modelo de tipo  $ARIMA(0, d, 1)$ .

Hyndman y G.Athanasopoulos (2013) señalan que en el caso de que el mejor modelo que describe los datos fuera de la forma  $ARIMA(p, d, q)$ , con  $p, q > 0$ , entonces la inspección visual de los gráficos de ACF y PACF no es recomendable. Otra alternativa es utilizar la función `auto_arima` que se incluye en el paquete estadístico `statsmodel`. Esta función realiza tests de diferenciación (Dickey-Fuller, Phillips-Perron entre otros) para encontrar el parámetro  $d$ , y además criterios basados en la verosimilitud de los datos (AIC, BIC entre otros). Es esta última herramienta la que se decide utilizar en la presente investigación,

dando como resultado para el caso de la serie correspondiente en el gráfico 14, un modelo ARIMA(1, 0, 2). La metodología con respecto a la división y proyección de resultados es la misma descrita en los modelos de suavización exponencial.

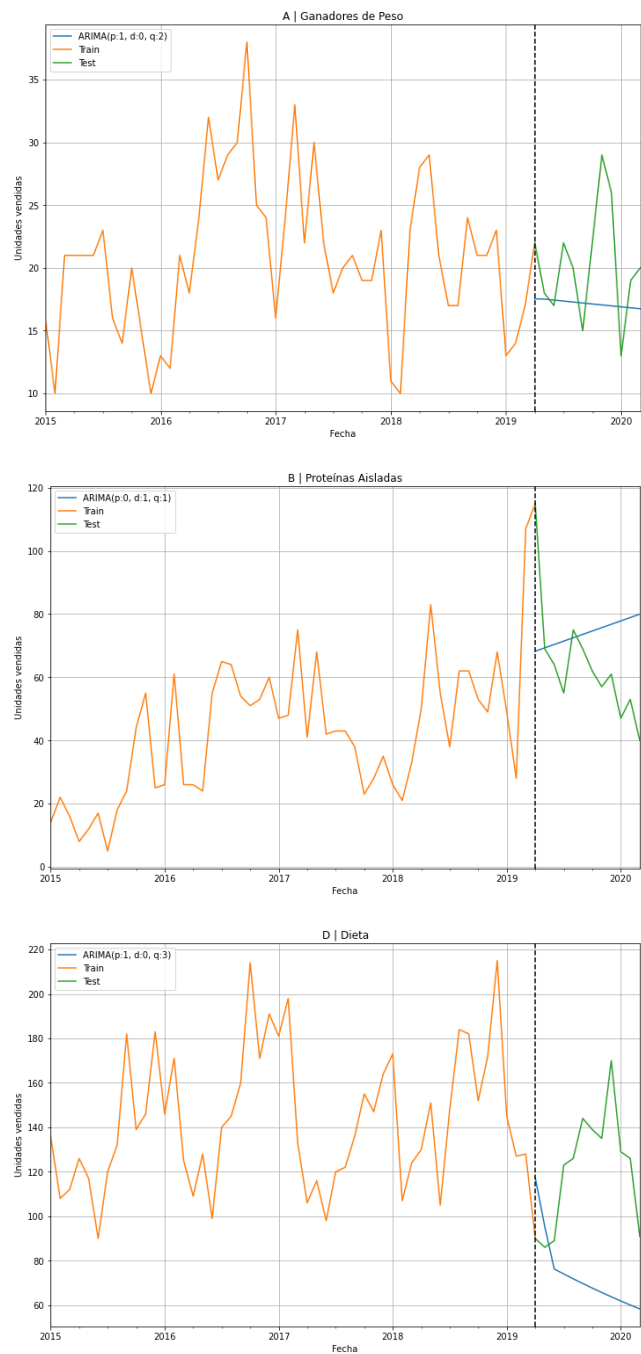


Gráfico 15: Proyección del modelo ARIMA en 3 series distintas

En el gráfico 15 se puede apreciar las proyecciones del modelo ARIMA para las series *A*|*GanadoresDePeso*, *B*|*ProteinasAisladas* y *D*|*Dieta*, además de los modelos finales correspondientes: ARIMA(1, 0, 2), ARIMA(0, 1, 1) y ARIMA(1, 0, 3) respectivamente.

## V.2 Redes neuronales recurrentes

La metodología utilizada en el modelamiento a través de redes neuronales difiere en algunos aspectos a la utilizada con los modelos clásicos de series de tiempo. En el caso de las RNN, no solo es necesario tener un conjunto de entrenamiento y testeo, sino que también es recomendable la existencia de un conjunto de validación, el cual se utiliza para la selección de hiperparámetros de la red neuronal. Bengio (2012) señala que si bien es posible utilizar el mismo conjunto de entrenamiento para seleccionar los hiperparámetros, esta no es una práctica muy recomendable. En ese sentido, el mismo autor señala que es una mejor práctica realizar dicha selección en base al desempeño del modelo en una muestra fuera del conjunto de entrenamiento, ya sea, en un conjunto de validación, o bien utilizando la técnica de validación cruzada.

Si bien no existe una división recomendada de los datos, la que se suele utilizar es una partición 70/15/15, donde el conjunto de entrenamiento posee el 70% de los datos, mientras que los conjuntos de validación y testeo poseen cada uno un 15%. Para la presente investigación se decide utilizar un ratio 70/10/20, principalmente por qué parte del objetivo de la investigación es la de evaluar el desempeño de los modelos en una proyección de 12 meses al futuro, condición que se cumple otorgando un 20% de los datos al conjunto de testeo. Es con ello que los conjuntos de entrenamiento, validación y testeo poseen cada uno un total de 45, 6 y 12 meses respectivamente. La figura 5 ilustra la división de datos que se utilizara para el entrenamiento y evaluación de las redes neuronales recurrentes.

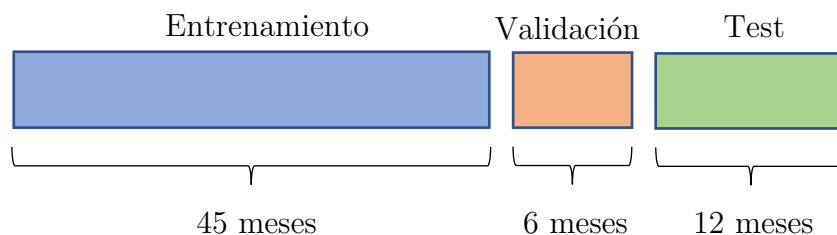


Figura 5: División de conjuntos para el entrenamiento de RNN

Es importante remarcar, que tanto en la división que se realiza en los modelos clásicos de series de tiempo, como en las redes neuronales recurrentes, los datos siguen el orden cronológico original, es decir, el conjunto de entrenamiento posee los primeros 45 meses (enero de 2015 hasta septiembre de 2018), el de validación tiene los siguientes 6 meses (octubre de 2018 hasta marzo de 2019) y el de testeo los últimos 12 meses de la base de datos (abril de 2019 hasta marzo de 2020). Con el propósito de ilustrar como se realiza esta división, se eligen de manera arbitraria 3 locales y dos categorías (gráfico 16). Los resultados para el total de las series se encuentran en el apartado de Resultados y Análisis.

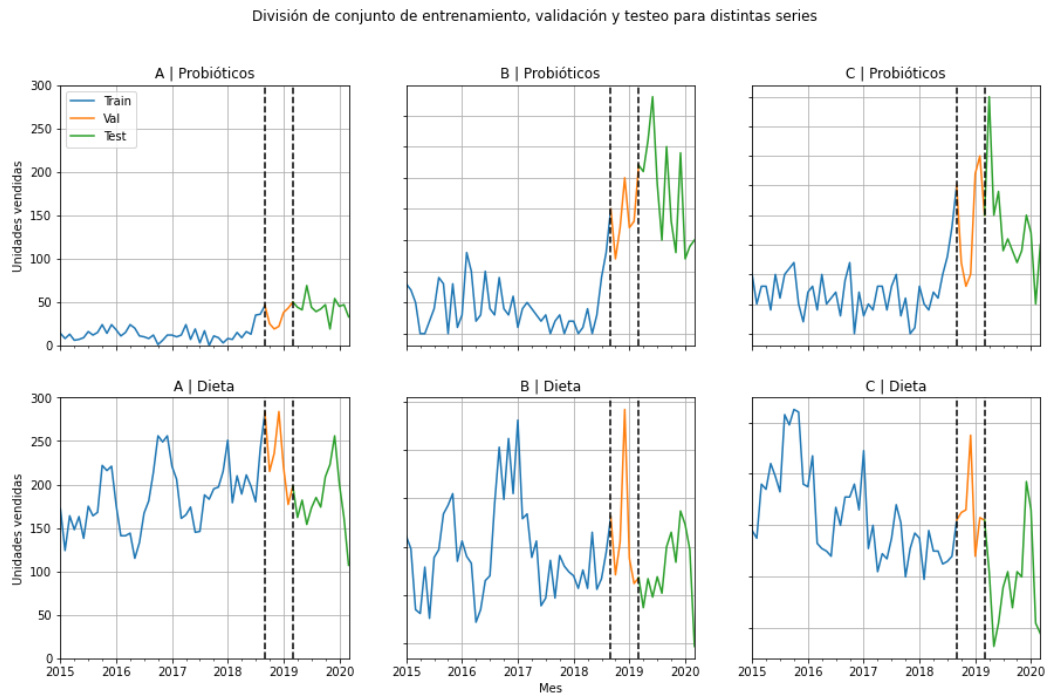


Gráfico 16: División en conjuntos de entrenamiento, validación y test para 6 series distintas

Una vez realizada la división de los datos, se procede a aplicar una estandarización de estos con el propósito de que su nueva escala este entre 0 y 1. La transformación seleccionada es la de normalización, o también conocida como transformación  $min - max$ , descrita en el apartado de Marco Teórico. Es importante recalcar que la normalización se debe aplicar utilizando solo la información del conjunto de entrenamiento, ya que, en caso contrario, se estaría utilizando información del futuro que se pretende proyectar, y por ende sería una metodología no replicable en la realidad.

El proceso de estandarización de los datos es relevante para el comportamiento y entrenamiento de una red neuronal. Brownlee (2017) señala que, en el caso de las redes neuronales recurrentes, el uso de información que no se encuentre estandarizada, puede provocar que la fase de aprendizaje de la red sea distorsionada por los valores excesivamente alejados del resto.

Una vez estandarizados los datos, se procede a estructurarlos según la arquitectura que se le asigne a la red neuronal. Por ejemplo, en el caso de una RNN donde se utilicen 12 meses de información para predecir el siguiente mes, se crean varios subconjuntos, de tal manera que en cada uno de ellos existen dos “listas”, la primera conteniendo la información de los primeros 12 meses, y la segunda un único valor haciendo referencia al mes siguiente que es el que se busca predecir. Este proceso se repite avanzando de mes en mes hasta llegar al último mes del conjunto de validación. Esto da como resultado (para el caso de una red neuronal que utilice 12 meses para predecir el siguiente) 39 subconjuntos que son los utilizados para entrenar y seleccionar los hiperparámetros de la red neuronal. Matemáticamente, sea  $x_i$  el número de unidades vendidas en el mes  $i$ , esto se puede ver de la siguiente manera:

$$\begin{aligned}
 x_1, x_2 \dots x_{12} &\rightarrow \{x_{13}\} \\
 x_2, x_3 \dots x_{13} &\rightarrow \{x_{14}\} \\
 &\cdot \\
 &\cdot \\
 &\cdot \\
 x_{39}, x_{40} \dots x_{50} &\rightarrow \{x_{51}\}
 \end{aligned}$$

Ya con los datos estructurados, se procede a ajustar la red neuronal utilizando el conjunto de entrenamiento. Dentro de este proceso, la red recibe como información de partida, un conjunto de hiperparámetros, los cuales definen en cierto grado la arquitectura de la RNN. Con estos hiperparámetros, se evalúa el desempeño de la red (ya entrenada), en el conjunto de validación (esta evaluación se hace en base al error cuadrático medio). Este proceso se repite tantas veces como configuraciones de hiperparámetros existan (esta metodología conoce como Grid-Search o búsqueda de rejilla), siendo la combinación que posea el mejor desempeño la seleccionada. Una vez seleccionado el conjunto de hiperparámetros, se procede

a reentrenar la red, esta vez, incorporando el conjunto de validación al conjunto de entrenamiento, con el propósito de actualizar con la información más reciente disponible, los parámetros de la red neuronal.

En la fase de predicción, existen dos aproximaciones para los algoritmos de redes neuronales. Una de ellas se conoce como Método directo. Yunpeng et al. (2017) describen que esta metodología requiere la construcción de más de un modelo, donde cada uno de ellos se utilizara para predecir un determinado valor de futuro. Por otro lado, existe la metodología basada en iteración, descrita por Atiya et al. (1999), como el uso de las predicciones que generan las redes como inputs para las siguientes proyecciones. También se puede utilizar el resultado generado para otras series de tiempo que tengan algún grado de correlación causal con la serie que se busca predecir. La principal desventaja de este último método es que se estaría generando una acumulación de errores a medida que se utilizan más estimaciones como inputs para la red, sin embargo, a diferencia del Método directo, requiere un menor costo computacional (Yunpeng et al., 2017). Es por este último motivo que se decide utilizar el método basado en iteración, principalmente por la gran cantidad de series de tiempo que se está estudiando (130).

Todas las arquitecturas utilizadas tienen la forma Many-to-One descrita en la sección de Marco Teórico. Además de lo anterior, en todas las formas que se trabajan se utiliza el algoritmo de Gradiente Descendente Estocástica como método de optimización. Con respecto a la selección de hiperparámetros, Bengio (2012) señala que la tasa de aprendizaje es probablemente el más importante y relevante de todos, motivo por el cual, siempre debiera ser ajustado dentro de la selección en base al rendimiento en el conjunto de validación. Además de lo anterior, el autor señala que en el caso de redes neuronales donde los inputs se encuentren en una escala entre 0 y 1, es recomendable probar con tasas que se encuentren entre  $10^{-6}$  y 1. Es por este motivo que finalmente se decide trabajar con tres posibles valores para este hiperparámetro: 0.001, 0.01 y 0.1. Con respecto a la función de activación, el autor señala que, en general, para la capa oculta, las funciones más utilizadas y recomendadas son la tangente hiperbólica, función sigmoide, la función rectificadora (*Relu*) y la función tangente hiperbólica fuerte. En el presente estudio se decide trabajar con las primeras tres funciones, y no las cuatro con el objetivo de reducir el costo computacional en el proceso de la selección. Por otro lado, las unidades será el tercer hiperparámetro que se optimizará. Debido al amplio costo computacional que implica añadir un valor más en el espectro de búsqueda de hiperparámetros, se decide asignar tres posibles valores para la dimensión de los hidden states: 20, 60 y 100. Para finalizar, con respecto a la profundidad de las redes, se

utilizarán dos tipos de modelos, unos con una sola capa, y otros con dos capas de profundidad. A continuación, se describen las 4 arquitecturas que se utilizaran en el estudio.

### V.2.1 RNN-12 Time steps, one layer (RNN-12-1)

Esta red utiliza 12 meses como información para predecir el mes siguiente. Esta selección está basada en el trabajo realizado por Atiya et al. (1999), donde utilizan combinaciones de meses recientes y el mismo valor a una distancia de un año o dos, con el propósito de capturar estacionalidades anuales que pueda existir en la información. Esta red solo tendrá una capa oculta, donde además se determina que la función de activación de la capa de salida es la función lineal, es decir, no se realiza ninguna transformación al valor generado en la fase anterior. Junto con lo anterior, en la fase de validación, el espectro de posibles valores de hiperparametros son los siguientes:

- Unidades: 20, 60 o 100
- Tasa de aprendizaje: 0.001, 0.01, 0.1
- Función de activación: *tanh*, *Relu* y *Sigmoide*

Si bien existen más hiperparametros, se decide trabajar con estos 3, y no más, debido al costo computacional que se incurre al procesar más valores. La combinación de estos 3 significa que en la fase de validación se estimen 27 modelos distintos (para cada serie de tiempo) donde, como se mencionó anteriormente, se elige la combinación que posea un menor error cuadrático medio en el conjunto de validación.

### V.2.2 RNN-24 Time steps, one layer (RNN-24-1)

La estructura de esta red es similar a la anterior. Solo posee una capa oculta, pero en lugar de utilizar 12 meses como información de ajuste, se utilizarán 24. El propósito de este número es el de lograr capturar posibles estacionalidades anuales de los productos más de una vez. Los Hiperparametros que se ajustan en esta red son los mismos descritos en la estructura anterior:

- Unidades: 20, 60, 100
- Tasa de aprendizaje: 0.001, 0.01, 0.1
- Función de activación: *tanh*, *Relu*, *Sigmoide*

### V.2.3 RNN-12 Time steps, two layers (RNN-12-2)

Esta red neuronal posee dos capas ocultas, una más que las dos redes anteriores. Por otro lado, utiliza 12 meses para predecir el siguiente mes. Es importante remarcar que, debido a la presencia de una capa extra, esta red posee dos grupos de hiperparámetros adicionales con respecto a los modelos anteriores: una segunda función de activación, y un segundo grupo de unidades, ambas pertenecientes a la nueva capa oculta. A continuación, se detalla el espectro de combinaciones donde se realiza el proceso de validación:

- Unidades (1era capa): 20, 60, 100
- Unidades (2da capa): 20, 60, 100
- Tasa de aprendizaje: 0.001, 0.01, 0.1
- Función de activación (1era capa): *tanh*, *Relu*, *Sigmoide*
- Función de activación (2da capa): *tanh*, *Relu*, *Sigmoide*

Es importante remarcar que el costo computacional en esta arquitectura es mayor que en los dos modelos anteriores. La incorporación de estos dos nuevos hiperparámetros genera que durante el proceso de validación se evalúa el desempeño de 243 combinaciones distintas, en lugar de las 27 que se realiza en las redes con una sola capa oculta.



## V.2.4 RNN-24 Time steps, two layers (RNN-24-2)

Esta red neuronal posee dos capas ocultas al igual que el modelo anterior, pero utiliza 24 meses en lugar de 12 para predecir el mes siguiente. El conjunto de hiperparametros donde se elige es el mismo que el modelo RNN-12-2.

# Capítulo VI – Resultados y análisis

El análisis de los resultados se realizará desde dos perspectivas distintas. Primero se realizará un análisis global de los modelos, es decir, cuál fue su desempeño en la predicción de los 12 meses en su conjunto en los datos correspondientes para testear. Después se realizará un análisis mensual, donde se evaluará el desempeño de los algoritmos en cada uno de los 12 meses que se realizan las estimaciones.

## VI.1 Distribución de los errores

Las 3 principales métricas que se utilizaran para evaluar el desempeño de los modelos son el valor absoluto medio (MAE), la raíz del error cuadrático medio normalizada (NRMSE) y el error porcentual medio (MAPE). Se decide utilizar estas tres métricas debido a que tienen características distintas que permiten obtener conclusiones diferentes. La métrica NRMSE al elevar al cuadrado los errores, penaliza más fuerte las grandes desviaciones, por lo que esta métrica es útil cuando se quiere dar énfasis al ajuste en valores atípicos. Además de lo anterior, la métrica se encuentra normalizada, lo que permite, a diferencia de RMSE, MAE y MAPE, comparar el desempeño de un modelo inter-series. Por otro lado, se decide utilizar la métrica MAE para tener otra referencia, principalmente debido a que esta no penaliza fuertemente las grandes desviaciones. Además de las dos métricas anteriores, se calcula el MAPE, con el propósito de tener una perspectiva de la proporción del error en cada una de las series.

Modelo/Métrica	NRMSE	MAE	MAPE
ARIMA	3	2	2
SEWMA	2	3	2
DEWMA	3	2	1
TEWMA	4	10	4
RNN-12-1	1	2	2
RNN-24-1	3	6	3
RNN-12-2	6	1	5

RNN-24-2	8	4	11
----------	---	---	----

Tabla 7: Numero de series donde cada modelo es mejor según NRMSE, MAE y MAPE, para el Grupo 1

En la tabla 7 se resume el desempeño de los 8 modelos utilizados, en base al número de series donde el algoritmo fue mejor, según las métricas NRMSE, MAE y MAPE, en las categorías pertenecientes al Grupo 1. En ella se puede apreciar que, con respecto a NRMSE, las arquitecturas RNN-12-2 y RNN-24-2, son las que presentaron un mejor desempeño, obteniendo el mejor resultado en 6 y 8 series respectivamente. A priori, este resultado es un indicio de que la incorporación de dos capas ocultas, generan un mejor resultado para grandes desviaciones, a diferencia de las redes que solo incorporan una capa. Por otro lado, con respecto a la métrica MAE, el modelo de suavización exponencial triple es que presenta una mayor cantidad de series donde obtiene el mejor desempeño (10 de 30). Con respecto a la métrica MAPE, la red neuronal de 2 capas que utiliza 24 puntos pasados para predecir es la que presenta un mejor desempeño en más series que los otros modelos (11). Esto último indica que, en términos proporcionales, la capa extra, y el uso de dos años de información son relevantes para predecir el futuro de las series del Grupo 1.

Modelo/Métrica	NRMSE	MAE	MAPE
ARIMA	16	16	13
SEWMA	16	20	20
DEWMA	13	13	6
TEWMA	17	19	12
RNN-12-1	8	7	10
RNN-24-1	11	14	14
RNN-12-2	8	7	5
RNN-24-2	11	4	20

Tabla 8: Numero de series donde cada modelo es mejor según NRMSE, MAE o MAPE, para el Grupo 2

En la tabla 8 se adjuntan los mismos resultados que la tabla 7, pero con respecto a las categorías pertenecientes al Grupo 2 (100 series en total). En términos generales, no existe en las tres métricas un modelo que presente un desempeño muy superior al resto. Para poder precisar la selección de los modelos se procede a analizar la distribución.

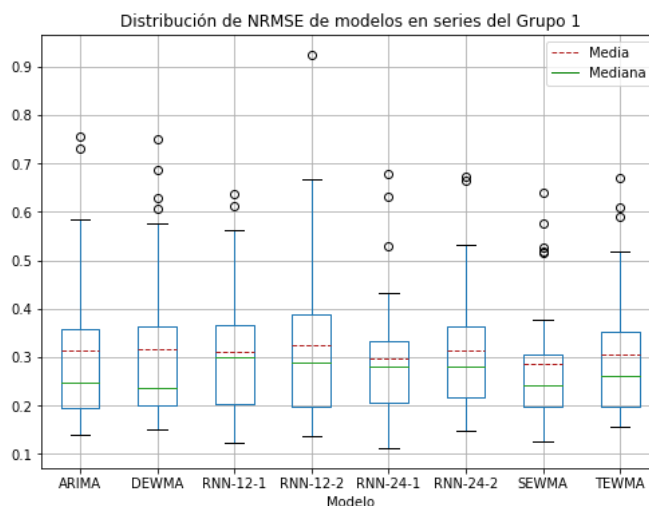


Gráfico 17: Distribución de la métrica NRMSE en cada uno de los modelos para series del Grupo 1

	Media	Std	Min	25%	50%	75%	Max
ARIMA	0.31	0.17	0.14	0.19	0.25	0.36	0.76
SEWMA	0.29	0.14	0.13	0.2	0.24	0.3	0.64
TEWMA	0.31	0.15	0.16	0.2	0.26	0.35	0.67
DEWMA	0.32	0.17	0.15	0.2	0.24	0.36	0.75
RNN-12-1	0.31	0.14	0.12	0.2	0.3	0.37	0.64
RNN-12-2	0.32	0.17	0.14	0.2	0.29	0.39	0.93
RNN-24-1	0.30	0.13	0.11	0.21	0.28	0.33	0.68
RNN-24-2	0.31	0.13	0.15	0.22	0.28	0.36	0.67

Tabla 9: Detalle de distribución de la métrica NRMSE de los modelos en las series del Grupo 1

Observando el gráfico 17, el cual ilustra la distribución de la métrica NRMSE en las series perteneciente al Grupo 1 para cada uno de los modelos, se puede observar a priori que el modelo RNN-12-2 es el que presenta una mayor dispersión en cuanto a la métrica de desempeño. Por otro lado, el modelo de suavización exponencial simple es el que, visualmente presenta una menor dispersión, sin embargo, observando la información detallada en la tabla 9, son los modelos RNN-24-1 y RNN-24-2 los que presentan una menor desviación estándar. A nivel de media, el modelo que mejor se desempeña es el de suavización exponencial simple, seguido por la red RNN-24-1. Bajo esta perspectiva, en caso de tener que elegir un único algoritmo, el modelo de suavización exponencial de orden 1 es el que presenta una mayor certeza de rendimiento para el total de las series del Grupo 1.

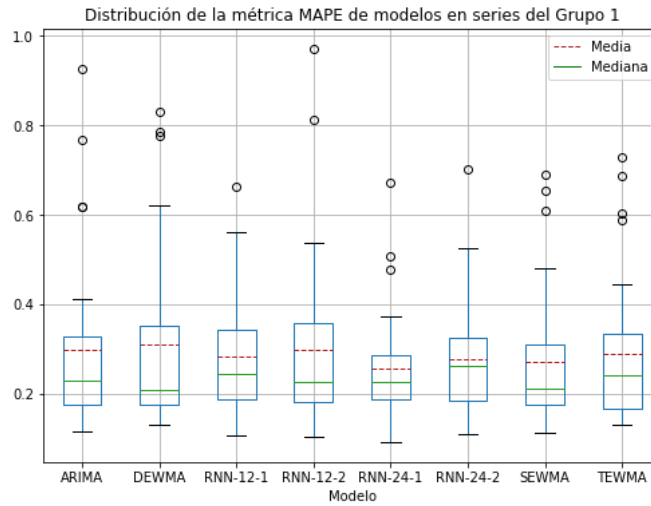


Gráfico 18: Distribución de la métrica MAPE en cada uno de los modelos para series del Grupo 1

	Media	Std	Min	25%	50%	75%	Max
ARIMA	0.30	0.20	0.12	0.18	0.23	0.33	0.93
SEWMA	0.31	0.21	0.13	0.18	0.21	0.35	0.83
TEWMA	0.28	0.14	0.11	0.19	0.25	0.34	0.66
DEWMA	0.30	0.19	0.10	0.18	0.23	0.36	0.97
RNN-12-1	0.26	0.12	0.09	0.19	0.23	0.28	0.67
RNN-12-2	0.28	0.13	0.11	0.19	0.26	0.32	0.70
RNN-24-1	0.27	0.15	0.11	0.17	0.21	0.31	0.69
RNN-24-2	0.29	0.17	0.13	0.17	0.24	0.33	0.73

Tabla 10: Detalle de distribución de la métrica MAPE de los modelos en las series del Grupo 1

El gráfico 18 y la tabla 10 adjuntan la distribución de la métrica MAPE sobre las series de las categorías del Grupo 1. Se puede apreciar que, con respecto a la media, las configuraciones RNN-12-1 y RNN-24-1 presentan mejores resultados, siendo la primera de estas estructuras la que posee una menor desviación en sus errores, y un error medio del 26%. Por otro lado, los modelos de suavización exponencial simple, doble y ARIMA son los que presentan una peor media en los resultados, teniendo un error promedio en torno al 30%.

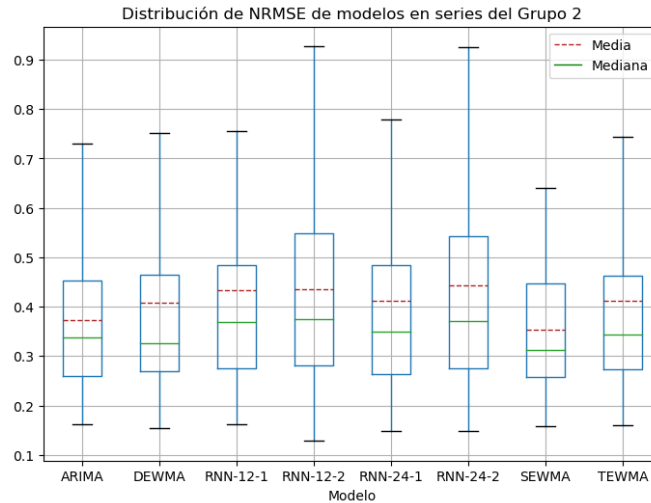


Gráfico 19: Distribución de la métrica NRMSE en cada uno de los modelos para series del Grupo 2

	Media	Std	Min	25%	50%	75%	Max
ARIMA	0.37	0.15	0.16	0.26	0.34	0.45	0.95
DEWMA	0.41	0.22	0.15	0.27	0.33	0.47	1.22
SEWMA	0.35	0.14	0.16	0.26	0.31	0.45	0.85
TEWMA	0.41	0.23	0.16	0.27	0.34	0.46	1.49
RNN-12-1	0.43	0.25	0.16	0.27	0.37	0.48	1.45
RNN-12-2	0.44	0.23	0.13	0.28	0.37	0.55	1.74
RNN-24-1	0.41	0.22	0.15	0.26	0.35	0.48	1.28
RNN-24-2	0.44	0.26	0.15	0.27	0.37	0.54	1.72

Tabla 11: Detalle de distribución de la métrica NRMSE de los modelos en las series del Grupo 2

Con respecto a las series pertenecientes a las categorías del Grupo 2 (100 en total), el gráfico 19 ilustra un comportamiento relativamente similar al que ocurre con las series del Grupo 1. El modelo de suavización exponencial simple es el que a priori muestra una menor dispersión de su precisión (con respecto a NRMSE), información que se corrobora en la tabla 11, donde se puede apreciar que este modelo es el que presenta una menor desviación estándar de los ocho (0.14). Por otro lado, el modelo que presenta una mayor desviación (0.26), y por lo que, a priori, es el modelo menos generalizable para el total de las series es la red neuronal que posee 2 capas ocultas y utiliza 24 time steps en su proceso de predicción (RNN-24-2). Por otro lado, en base a la media de los resultados, los modelos de suavización exponencial simple y ARIMA son los que presentan un mejor rendimiento. Además de lo anterior, es importante destacar como la media de los errores aumenta con respecto a las

series del Grupo 1, lo que tiene sentido debido a la mayor desviación que presentan las series del Grupo 2 por ser más desagregadas.

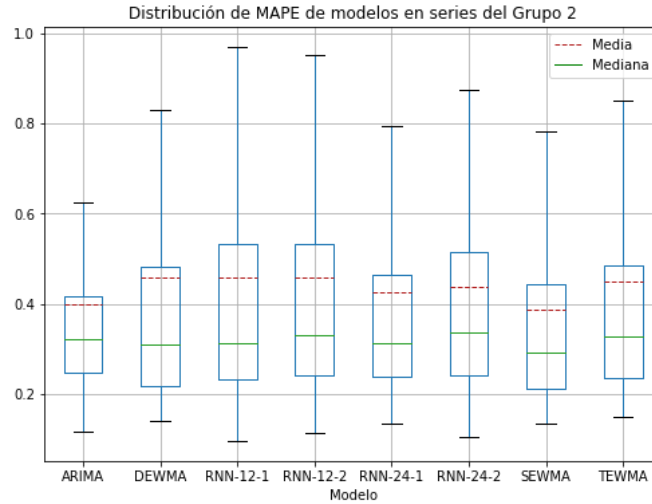


Gráfico 20: Distribución de la métrica MAPE en cada uno de los modelos para series del Grupo 2

	Media	Std	Min	25%	50%	75%	Max
ARIMA	0.40	0.26	0.12	0.25	0.32	0.42	1.46
DEWMA	0.46	0.38	0.14	0.22	0.31	0.48	2.04
SEWMA	0.39	0.30	0.13	0.21	0.29	0.44	1.95
TEWMA	0.45	0.36	0.15	0.23	0.33	0.48	1.87
RNN-12-1	0.46	0.37	0.1	0.23	0.31	0.53	1.8
RNN-12-2	0.46	0.34	0.11	0.24	0.33	0.53	1.79
RNN-24-1	0.42	0.31	0.13	0.24	0.31	0.47	1.56
RNN-24-2	0.44	0.35	0.11	0.24	0.34	0.51	2.72

Tabla 12: Detalle de distribución de la métrica MAPE de los modelos en las series del Grupo 2

La tabla 12 posee la distribución de la métrica MAPE para cada uno de los modelos en las series del Grupo 2. Se puede apreciar que, en términos de la media, el mejor modelo es el de suavización exponencial simple, presentando un error medio del 39%. Por otro lado, el modelo de suavización exponencial doble y las configuraciones RNN-12-1 y RNN-12-2 son las que presentaron un peor desempeño (46%).

Otra perspectiva al momento de analizar el desempeño de los modelos es estudiar su comportamiento para cada uno de los locales.

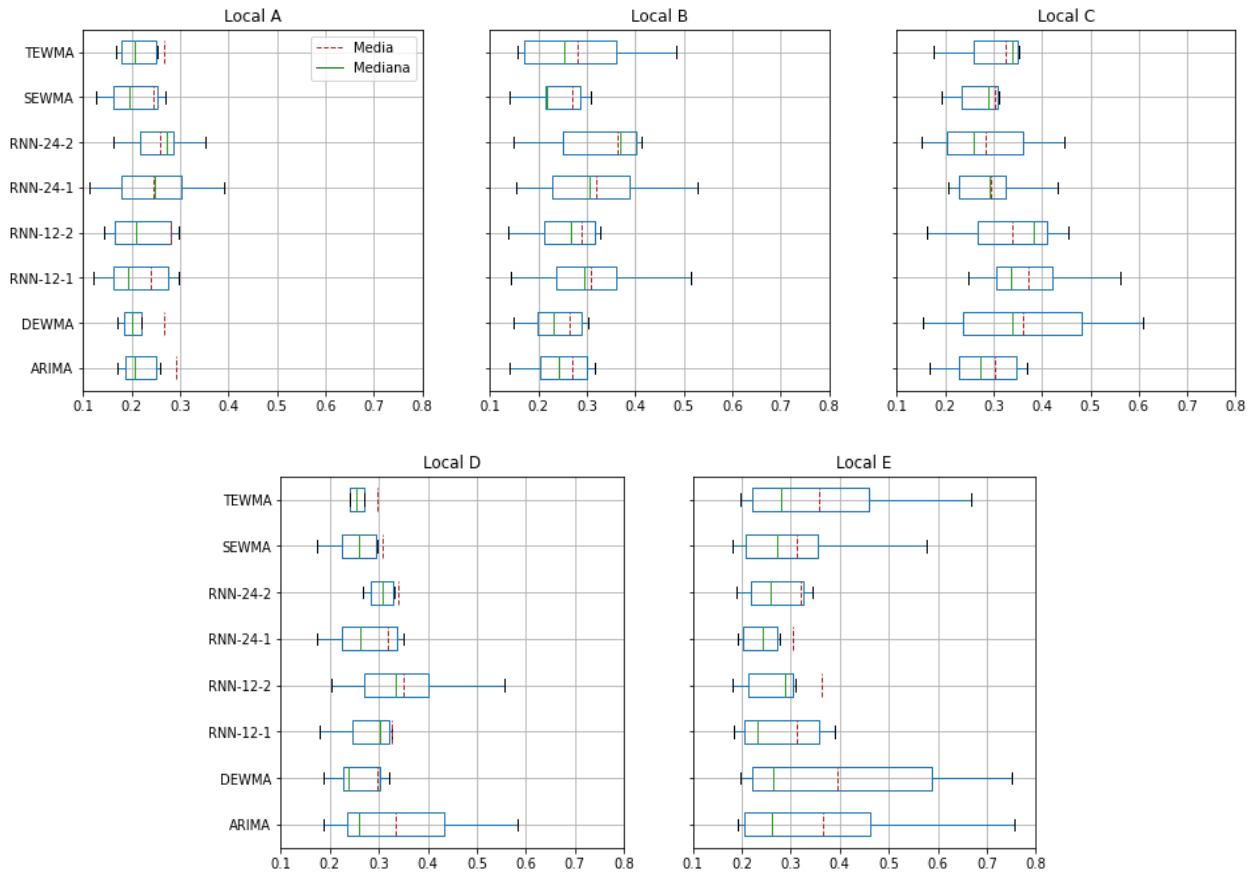


Gráfico 21: Distribución de la métrica NRMSE en cada uno de los modelos para series del Grupo 1, en cada local

El gráfico 21 muestra la dispersión de la métrica NRMSE de cada uno de los modelos, en todos los locales para las categorías del Grupo 1. En él se puede apreciar que, en términos generales, los 8 modelos presentan un mejor rendimiento en el local A, al tener la distribución de NRMSE más concentrada en valores en torno 0.1 y 0.4. Por otro lado, es el local E donde los modelos clásicos de series de tiempo, es decir, suavización exponencial simple, doble y triple, además de ARIMA, presentan en términos de la distribución de su precisión, un peor rendimiento en comparación con los modelos de redes neuronales, por lo que, al menos para dicho local, es preferible a priori, los modelos de redes neuronales recurrentes en lugar de los modelos clásicos, en las categorías pertenecientes al Grupo 1.



	A	B	C	D	E
Salud	(156.04, 0.39)	(82.43, 0.35)	(73.25, 0.33)	(101.63, 0.45)	(350.94, 0.56)
Deporte	(101.45, 0.2)	(75.66, 0.17)	(30.47, 0.15)	(69.58, 0.17)	(95.89, 0.12)
Dieta	(39.49, 0.21)	(48.98, 0.25)	(20.48, 0.26)	(31.03, 0.22)	(51.8, 0.19)
Consumible	(326.25, 0.22)	(209.35, 0.26)	(155.64, 0.29)	(311.91, 0.32)	(526.44, 0.26)
ByC	(17.14, 0.52)	(8.35, 0.47)	(7.91, 0.42)	(16.73, 0.84)	(34.74, 0.66)
Accesorio	(83.51, 0.51)	(49.34, 0.31)	(12.98, 0.24)	(40.01, 0.27)	(93.01, 0.36)

Tabla 13: Desviación estándar y coeficiente de variación para las 30 series del Grupo 1

En la tabla 13 se puede apreciar la desviación estándar y el coeficiente de variación para las 30 series del Grupo 1 (en la sección de anexos se adjunta la misma tabla, pero con respecto a las categorías del Grupo 2). En ella se puede observar que en 5 de las 6 categorías del Grupo 1, el local E es el que presenta una mayor desviación estándar, no así, el coeficiente de variación. Esto podría explicar el mal desempeño principalmente de los modelos de suavización exponencial simple y doble en el local E, debido a que ambos modelos son proyecciones lineales en el futuro, por lo que, aquellas series con una mayor varianza se traducirá en una mayor cantidad de errores.

	A	B	C	D	E
Media	2758.87	1844.08	1116.33	1911.51	4015.68
Std	501.45	347.69	182.4	331.9	829.83
Cv	0.18	0.19	0.16	0.17	0.21

Tabla 14: Media, desviación estándar y coeficiente de variación de las ventas de los locales

La tabla 14 muestra la media, la desviación estándar y el coeficiente de variación de las ventas totales de los grupos en estudio para los 5 locales. Se puede apreciar que en las 3 métricas el local E es superior al resto, pudiendo llevar a la conclusión antes expuesta, que, al menos para las categorías del Grupo 1, que son más agregadas que el Grupo 2, los modelos de redes neuronales tienen un mejor comportamiento que los modelos clásicos de series de tiempo, debido principalmente a la gran variación que existe en las ventas con respecto al resto.

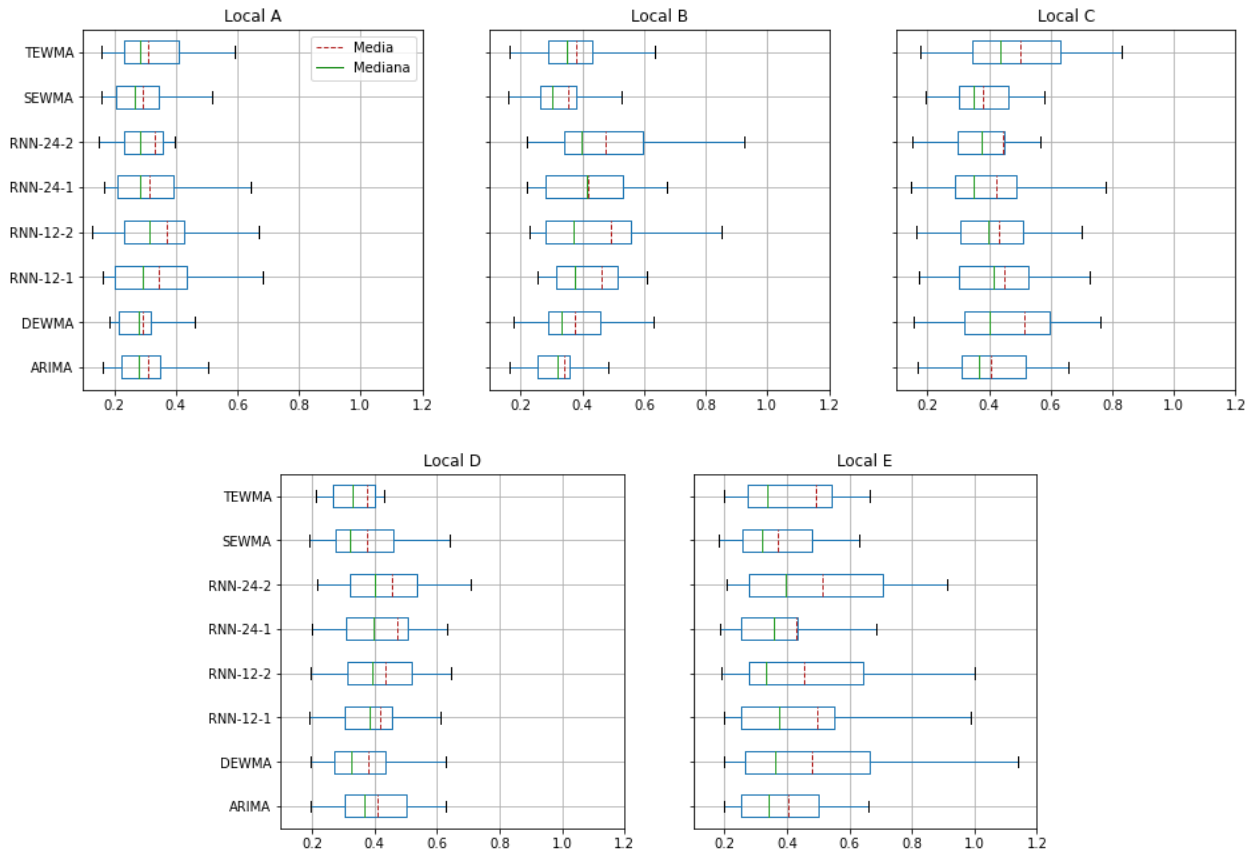


Gráfico 22: Distribución de la métrica NRMSE en cada uno de los modelos para series del Grupo 2, en cada local

El gráfico 22 ilustra la misma información que el gráfico 21, pero con respecto a las categorías del Grupo 2. De esta información, a priori, no es posible obtener conclusiones relevantes. Los únicos modelos que se ven superiores al resto son la red RNN-24-2 en el local A, y el de suavización exponencial triple en el local D. En el resto de los locales no se observa un algoritmo dominante en base a la distribución de la métrica NRMSE.

En la sección Anexos, se encuentran las tablas con los detalles para cada una de las series y los modelos. El rendimiento utilizando las métricas NRMSE, MAE y MAPE, tanto para las 30 series del Grupo 1, como para las 100 series del Grupo 2.

## VI.2 Test-t Student's pareado

Es importante recalcar que utilizar únicamente la media como único índice para calificar los distintos modelos puede llevar a cometer errores al momento de elegir el mejor algoritmo. Puede ocurrir que, las diferencias existentes en las medias se deban únicamente al azar, o la forma en que se eligieron los conjuntos de entrenamiento y testeo en los modelos (Brownlee, 2020), es por ello que se decide aplicar el test de hipótesis Student's pareado. El propósito de aplicar esta prueba es la de comparar las distribuciones de los resultados de los modelos. En total, se realizarán 28 tests distintos, comparando pares de modelos.

La prueba test-t Student's pareado compara las medias de dos distribuciones. Las hipótesis son las siguientes:

$h_0$ : *Las medias de las dos distribuciones son iguales* ( $\mu_1 = \mu_2$ )

$h_1$ : *Existe una diferencia significativa entre ambas medias* ( $\mu_1 \neq \mu_2$ )

Debido al costo computacional, cada modelo solo fue entrenado y testado (una vez realizado el proceso de validación en las redes neuronales) una única vez en cada una de las series. Es por este motivo, que para aplicar el test se utilizó la métrica NRMSE como variable aleatoria. La raíz del error cuadrático medio (al estar normalizada permite la comparación inter-series). Por otro lado, uno de los supuestos del test de hipótesis que se realizara, es la de que cada distribución sea Gaussiana, es por este motivo que se decide realizar en la métrica NRMSE y su transformación logarítmica.

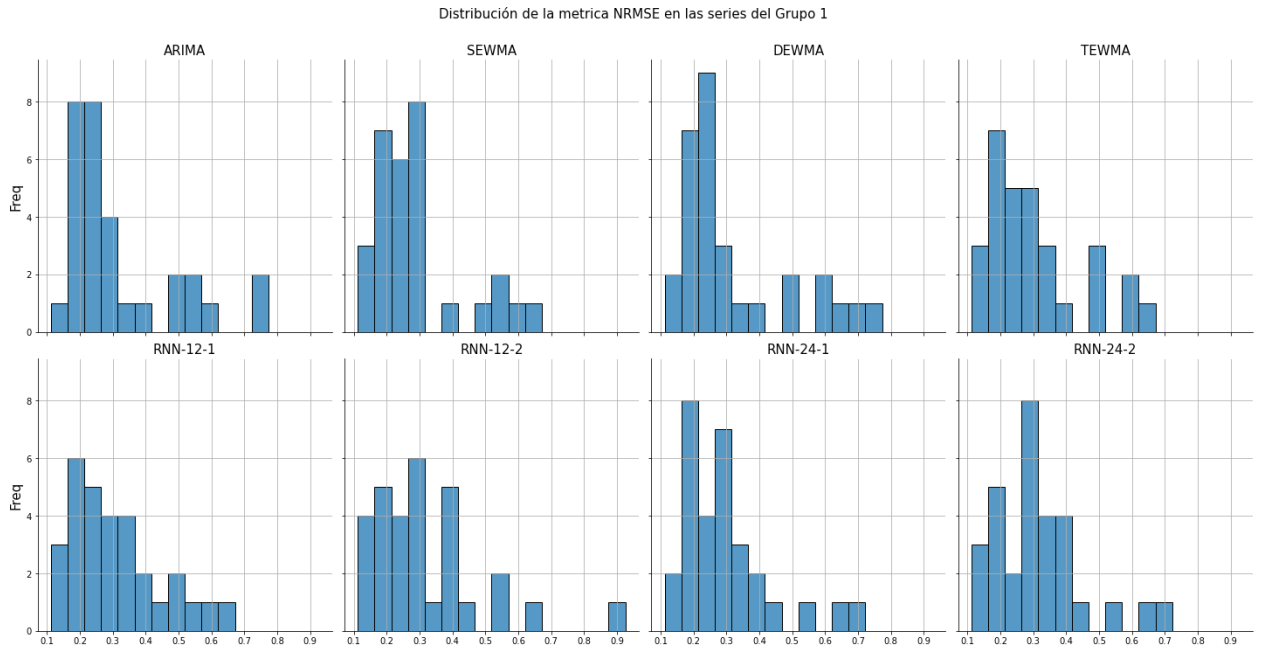


Gráfico 23: Distribución de la métrica NRMSE en cada uno de los modelos para series del Grupo 1

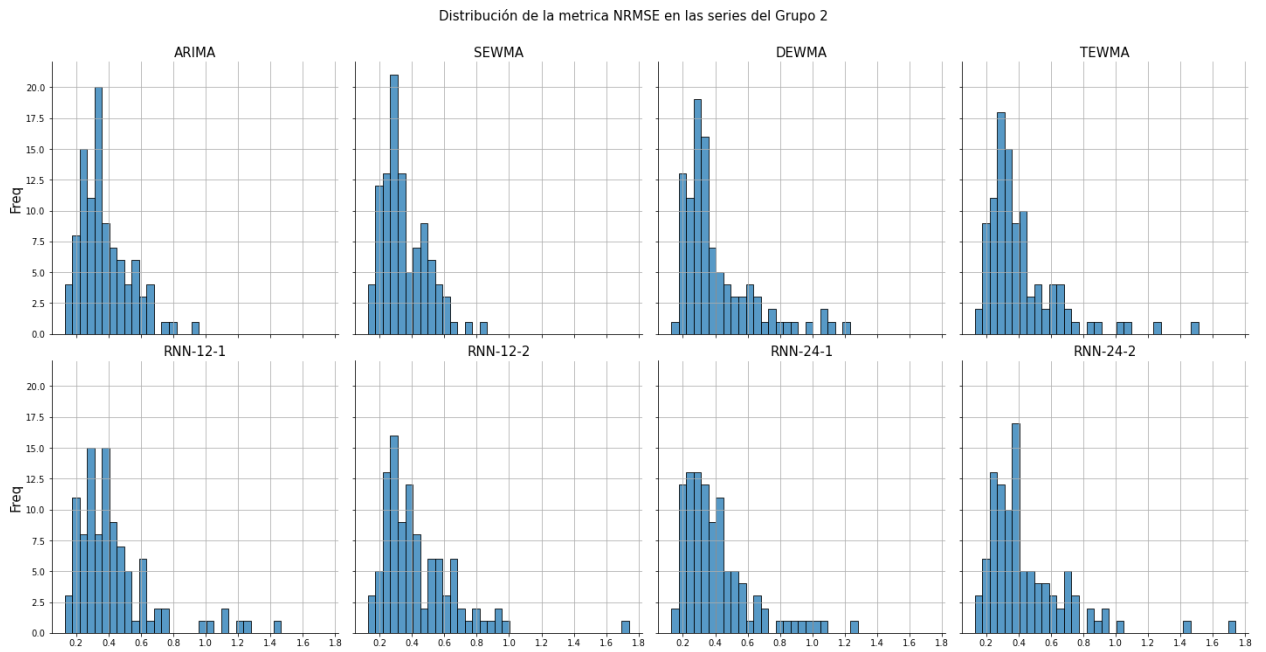


Gráfico 24: Distribución de la métrica NRMSE en cada uno de los modelos para series del Grupo 2

Los gráficos 23 y 24 muestra las distribución de la métrica NRMSE para cada uno de los modelos, en las series del Grupo 1 y Grupo 2. Si bien, visualmente no poseen una forma de

distribución Gaussiana, el test se aplicará sobre dichas distribuciones. Los resultados se adjuntan en la tabla 12.

	ARIMA	SEWMA	DEWMA	TEWMA	RNN-12-1	RNN-12-2	RNN-24-1	RNN-24-2
ARIMA	-	0.15	0.76	0.57	0.94	0.42	0.39	0.99
SEWMA	0.15	-	0.09	0.14	0.13	0.11	0.61	0.23
DEWMA	0.76	0.09	-	0.37	0.77	0.70	0.32	0.86
TEWMA	0.57	0.14	0.37	-	0.70	0.26	0.54	0.74
RNN-12-1	0.94	0.13	0.77	0.70	-	0.41	0.21	0.95
RNN-12-2	0.42	0.11	0.70	0.26	0.41	-	0.09	0.54
RNN-24-1	0.39	0.61	0.32	0.54	0.21	0.09	-	0.19
RNN-24-2	0.99	0.23	0.86	0.74	0.95	0.54	0.19	-

Tabla 15: P valores para test-t Student's pareado para cada par de modelos, según el NRMSE de las series del Grupo 1

Observando la tabla 15, no existe, para un nivel de significancia del 5%, ningún p-valor que lleve a la conclusión que las medias de la métrica NRMSE (para los ocho modelos aplicados en las series del Grupo 1) sean significativamente distintas. En otras palabras, se falla en rechazar la hipótesis nula (la distribución de los errores son generadas por un mismo proceso).

	ARIMA	SEWMA	DEWMA	TEWMA	RNN-12-1	RNN-12-2	RNN-24-1	RNN-24-2
ARIMA	-	0.06	0.03	0.02	0.00	0.00	0.02	0.00
SEWMA	0.06	-	0.00	0.00	0.00	0.00	0.00	0.00
DEWMA	0.03	0.00	-	0.82	0.17	0.19	0.87	0.09
TEWMA	0.02	0.00	0.82	-	0.28	0.26	0.98	0.17
RNN-12-1	0.00	0.00	0.17	0.28	-	0.90	0.28	0.62
RNN-12-2	0.00	0.00	0.19	0.26	0.90	-	0.19	0.70
RNN-24-1	0.02	0.00	0.87	0.98	0.28	0.19	-	0.09
RNN-24-2	0.00	0.00	0.09	0.17	0.62	0.70	0.09	-

Tabla 16: P valores para test-t Student's pareado para cada par de modelos, según el NRMSE de las series del Grupo 2

Para las series del Grupo 2, se realizó el test en los 8 modelos y los p-valores respectivos se encuentran en la tabla 16. Combinando dicha información y la de la tabla 11, la cual indica que, con respecto a la media, los modelos de suavización exponencial simple (0.35) y ARIMA (0.37) son los que presentan un mejor desempeño, se puede concluir (a un nivel de significancia del 5%) que la distribución de los errores son significativamente distintas a los otros 6. Es importante recalcar, que son los únicos algoritmos que poseen una diferencia significativa con respecto al resto, lo que lleva a concluir que son los mejores modelos en términos globales.

Otro enfoque se realizará es el de aplicar la transformación logarítmica sobre la métrica NRMSE con el propósito de generar una distribución más cercana a la de Gauss. La forma de la distribución para las series del Grupo 1 y Grupo 2 se puede apreciar en los gráficos 25 y 26.

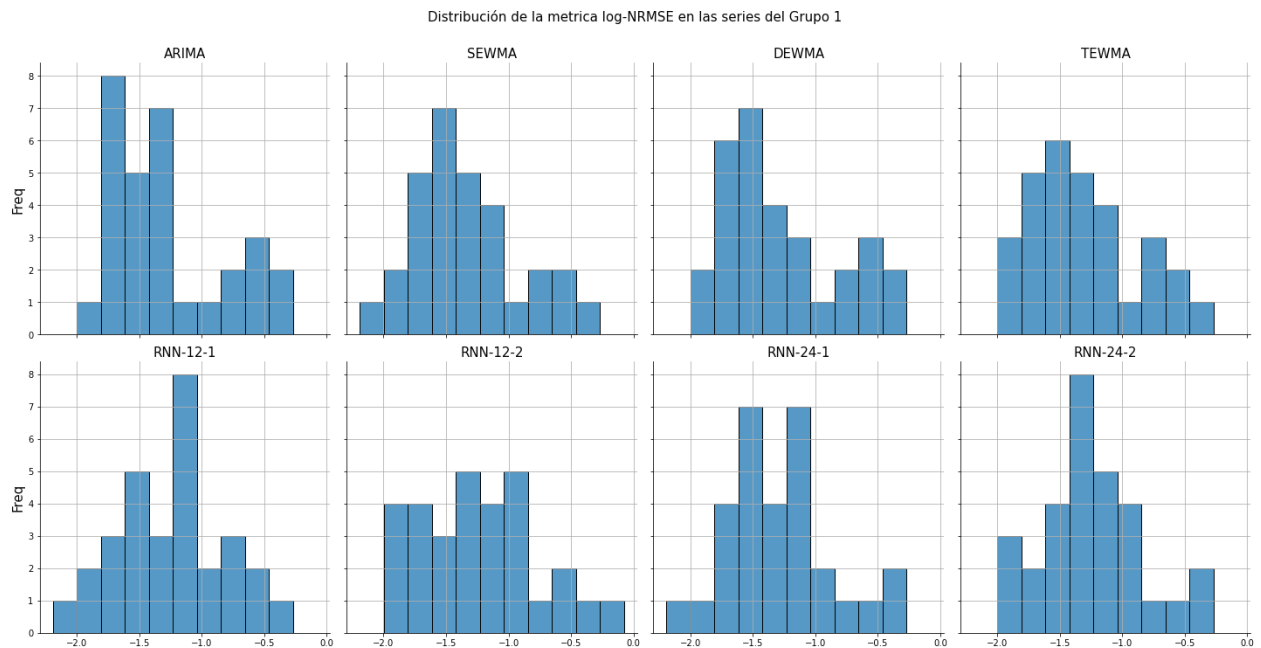


Gráfico 25: Distribución de la métrica log-NRMSE en cada uno de los modelos para series del Grupo 1

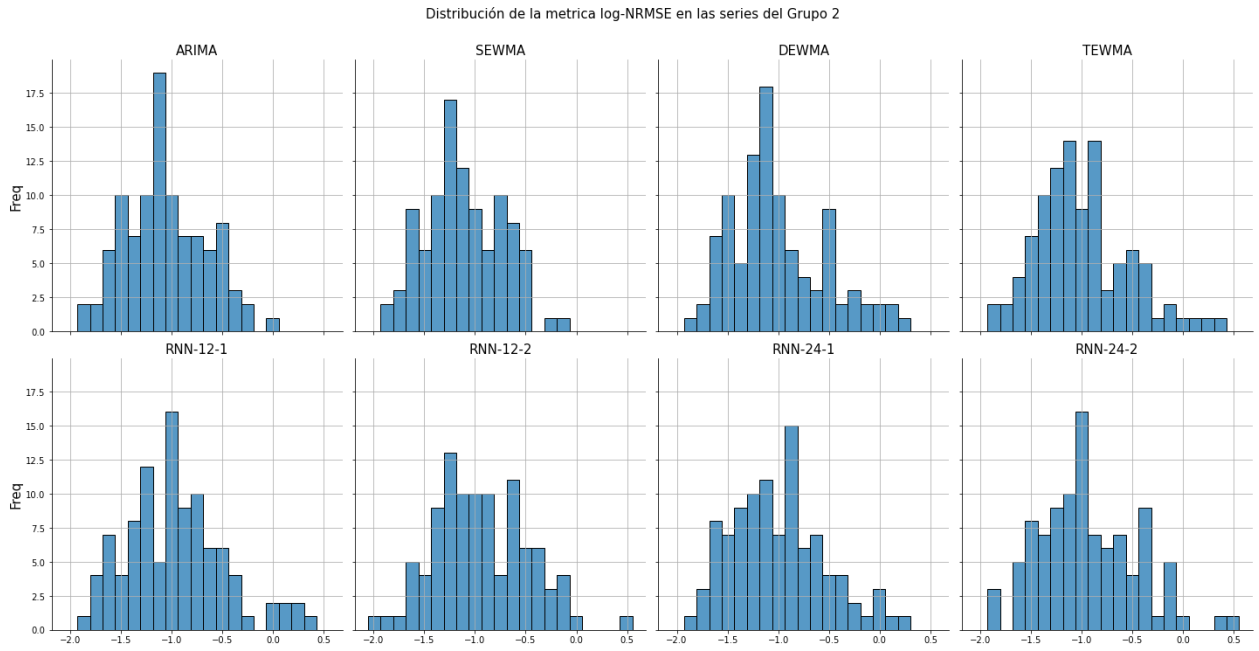


Gráfico 26: Distribución de la métrica log-NRMSE en cada uno de los modelos para series del Grupo 2

Una vez aplicada la transformación, se procede a aplicar el test. Los resultados de adjuntan en las tablas 17 y 18.

	ARIMA	SEWMA	DEWMA	TEWMA	RNN-12-1	RNN-12-2	RNN-24-1	RNN-24-2
ARIMA	-	0.12	0.90	0.81	0.77	0.38	0.61	0.57
SEWMA	0.12	-	0.08	0.10	0.05	0.05	0.41	0.10
DEWMA	0.90	0.08	-	0.68	0.84	0.51	0.56	0.66
TEWMA	0.81	0.10	0.68	-	0.61	0.37	0.71	0.50
RNN-12-1	0.77	0.05	0.84	0.61	-	0.64	0.23	0.76
RNN-12-2	0.38	0.05	0.51	0.37	0.64	-	0.18	0.99
RNN-24-1	0.61	0.41	0.56	0.71	0.23	0.18	-	0.18
RNN-24-2	0.57	0.10	0.66	0.50	0.76	0.99	0.18	-

Tabla 17: P valores para test-t Student's pareado para cada par de modelos, según el log-NRMSE de las series del Grupo 1

	ARIMA	SEWMA	DEWMA	TEWMA	RNN-12-1	RNN-12-2	RNN-24-1	RNN-24-2
ARIMA	-	0.03	0.08	0.02	0.01	0.00	0.07	0.00
SEWMA	0.03	-	0.00	0.00	0.00	0.00	0.00	0.00
DEWMA	0.08	0.00	-	0.49	0.13	0.05	0.69	0.06
TEWMA	0.02	0.00	0.49	-	0.29	0.11	0.97	0.16
RNN-12-1	0.01	0.00	0.13	0.29	-	0.58	0.26	0.56
RNN-12-2	0.00	0.00	0.05	0.11	0.58	-	0.07	0.97
RNN-24-1	0.07	0.00	0.69	0.97	0.26	0.07	-	0.08
RNN-24-2	0.00	0.00	0.06	0.16	0.56	0.97	0.08	-

Tabla 18: P valores para test-t Student's pareado para cada par de modelos, según el log-NRMSE de las series del Grupo 2

Observando los resultados de las tablas 17 y 18, los p-valores en ambos casos cambiaron, sin embargo, para un mismo nivel de significancia, las conclusiones siguen siendo las mismas que cuando se aplicó la metodología sobre la métrica NRMSE.

## VI.3 Test de Wilcoxon signed rank

En el apartado anterior se utilizó el test-t Student's pareado para evaluar si es que el rendimiento promedio de cada uno de los modelos presentaba diferencias significativas, sin embargo, en dicha prueba se realiza un supuesto fuerte: la distribución de la variable sigue una forma Gaussiana. Brownlee (2018) apunta que una alternativa, en el caso de no poder garantizar que la distribución es normal, es utilizar tests de hipótesis no paramétricos.

Brownlee (2018) señala que el test de Wilcoxon es la versión no paramétrica equivalente del test-t Student's pareado, y un contexto en el que puede ser utilizado es para evaluar el desempeño de dos o más modelos entrenados y testeados en un mismo conjunto de datos, es por ello que se decide realizar esta prueba en los modelos evaluados anteriormente.

Las hipótesis nula y alternativa son las mismas que las descritas en el test anterior. Se procede a aplicar la prueba tanto para los errores en las series del Grupo 1, como los errores



de las series del Grupo 2 en etapas separadas. Los resultados se presentan en las tablas 16 y 17.

	ARIMA	SEWMA	DEWMA	TEWMA	RNN-12-1	RNN-12-2	RNN-24-1	RNN-24-2
ARIMA	-	0.59	0.63	0.93	0.89	0.49	0.93	0.60
SEWMA	0.59	-	0.12	0.28	0.08	0.04	0.27	0.19
DEWMA	0.63	0.12	-	0.78	0.83	0.12	0.60	0.56
TEWMA	0.93	0.28	0.78	-	0.86	0.17	0.67	0.63
RNN-12-1	0.89	0.08	0.83	0.86	-	0.60	0.44	0.77
RNN-12-2	0.49	0.04	0.12	0.17	0.60	-	0.12	0.75
RNN-24-1	0.93	0.27	0.60	0.67	0.44	0.12	-	0.70
RNN-24-2	0.60	0.19	0.56	0.63	0.77	0.75	0.70	-

Tabla 19: P valores para test de Wilcoxon signed-rank para cada par de modelos, según NRMSE de las series del Grupo 1

	ARIMA	SEWMA	DEWMA	TEWMA	RNN-12-1	RNN-12-2	RNN-24-1	RNN-24-2
ARIMA	-	0.16	0.32	0.02	0.00	0.00	0.02	0.00
SEWMA	0.16	-	0.01	0.00	0.00	0.00	0.00	0.00
DEWMA	0.32	0.01	-	0.06	0.01	0.01	0.18	0.01
TEWMA	0.02	0.00	0.06	-	0.10	0.04	0.64	0.07
RNN-12-1	0.00	0.00	0.01	0.10	-	0.40	0.18	0.53
RNN-12-2	0.00	0.00	0.01	0.04	0.40	-	0.06	0.53
RNN-24-1	0.02	0.00	0.18	0.64	0.18	0.06	-	0.05
RNN-24-2	0.00	0.00	0.01	0.07	0.53	0.53	0.05	-

Tabla 20: P valores para test de Wilcoxon signed-rank para cada par de modelos, según NRMSE de las series del Grupo 2

Observando la tabla 19, se puede apreciar que las conclusiones son iguales que las obtenidas por el test-t Student's. A un nivel de significancia del 5%, para las series del Grupo 1, se falla en rechazar la hipótesis nula, por lo que no es posible asegurar que las diferencias existentes en las medias de los resultados sean significativas. Por otro lado, utilizando la información de la tabla 20, existen resultados distintos a los obtenidos por el test-t. El rendimiento medio del modelo SEWMA (0.35) es significativamente distinto al obtenido por el resto de los modelos, a excepción de ARIMA (0.37). Por otro lado, este último no presenta

diferencias estadísticamente significativas con el modelo de suavización exponencial doble (0.41).

Una conclusión que se puede obtener en base a los resultados de la tabla 20, es que el test de Wilcoxon, según las distribuciones disponibles, encuentra diferencias significativas entre modelos, a excepción de aquellos con la media más cercana. En términos generales, este resultado se puede explicar debido a la metodología de ranking que se utiliza en esta prueba. En base a esta información, no es posible asegurar que exista un modelo significativamente distinto al resto. En caso de tener que seleccionar un algoritmo según esta información, los modelos de suavización exponencial simple y ARIMA son los mejores candidatos.

## VI.4 Análisis mensual

Una análisis directo que se puede realizar sobre el desempeño de cada modelo en cada uno de los 12 meses proyectados es, para cada serie y cada uno de los meses proyectados, elegir al modelo con el menor error absoluto. Utilizando dicha metodología se obtiene la tabla 21 y 22. La primera hace referencia a las 30 series pertenecientes al Grupo 1, y la segunda las 100 series pertenecientes al Grupo 2.

	ARIMA	SEWMA	DEWMA	TEWMA	RNN-12-1	RNN-12-2	RNN-24-1	RNN-24-2
1	2	2	1	9	2	8	5	1
2	6	0	0	3	7	4	6	4
3	3	0	2	4	6	9	2	4
4	3	4	2	5	1	2	3	10
5	7	1	2	9	4	1	2	4
6	2	4	5	3	4	3	5	4
7	2	1	1	3	5	3	9	6
8	4	2	4	2	5	3	4	6
9	1	2	5	11	3	2	2	4
10	1	3	6	5	2	3	2	8
11	2	1	1	7	2	6	1	10
12	2	1	1	3	4	2	8	9

Tabla 21: Número de veces en cada mes que un modelo obtuvo el menor error absoluto en las series del Grupo 1

	ARIMA	SEWMA	DEWMA	TEWMA	RNN-12-1	RNN-12-2	RNN-24-1	RNN-24-2
1	11	13	8	25	7	17	8	11
2	14	9	6	20	14	8	13	16
3	5	11	16	17	15	9	9	18
4	9	17	7	20	8	13	10	16
5	11	9	13	19	8	14	14	12
6	8	17	13	14	15	6	12	15
7	12	13	8	16	11	10	17	13
8	10	13	11	12	15	17	9	13
9	16	11	15	16	10	16	10	6
10	7	13	15	13	10	15	12	15
11	14	10	9	12	15	10	14	16
12	23	9	6	9	11	8	14	20

Tabla 22: Número de veces en cada mes que un modelo obtuvo el menor error absoluto en las series del Grupo 2

Según la tabla 21, para las proyecciones cercanas (1 a 3 meses en el futuro), para las series del Grupo 1, los modelos de suavización exponencial triple (TEWMA) y las redes neuronales que usan 12 pasos (con una y dos capas), son las que presentan un mejor desempeño en términos del número de series. Por otro lado, para las proyecciones más distantes (10 a 12 meses) la red neuronal que utiliza 24 meses y dos capas ocultas es la que posee el mejor rendimiento (según el error absoluto) en más series que el resto.

La información de la tabla 22 muestra que el modelo de suavización exponencial simple es el que presenta un mejor desempeño en un mayor número de series para las proyecciones más próximas (1-3 meses). Por otro lado, la red neuronal con 24 meses y dos capas ocultas, seguido del modelo ARIMA son las que obtuvieron el mejor rendimiento en un mayor número de series en comparación con el resto de los modelos para las proyecciones a meses distantes.

Según ambas tablas, para las estimaciones más lejanas, la red de 24 meses (y dos capas) es la que a priori presenta mejores resultados en más series que el resto. Esta diferencia puede explicarse en parte por el número de proyecciones que utiliza como información para predecir el siguiente mes. Para las redes de 12 pasos, por ejemplo, al momento de proyectar el quinto

mes, utilizan 4 estimaciones y 8 valores reales, lo que podría generar un mayor sesgo al estar utilizando información que ya acarrea errores para realizar estimaciones.

Con el propósito de poder evaluar el desempeño de los modelos se propone utilizar una métrica similar al NRMSE. En términos simples, para cada serie y modelo, en cada uno de los 12 meses proyectados se calculará el error cuadrático, y luego se normalizará por el mismo factor de NRMSE. En adelante se referirá a esta métrica por las siglas MRSEE, que hacen referencia a ser la raíz del error cuadrático mensual escalado. Matemáticamente, sea  $y_m$  el valor real en el mes  $m$ , y  $\hat{y}_m$  la estimación para dicho mes, entonces la métrica se calcula de la siguiente manera:

$$MRSEE_m = \sqrt{\frac{(y_m - \hat{y}_m)^2}{\sum_{t=1}^{12} y_t^2}}$$

Esta métrica se calculará para cada uno de los 12 meses. El propósito es poder generar un valor que permita de cierta manera realizar comparaciones inter-series. Una vez realizada la transformación, se procede a promediar, para cada mes y modelo, el valor de la métrica propuesta. Los resultados se adjuntan en las tablas 23 y 24.

	1	2	3	4	5	6	7	8	9	10	11	12
ARIMA	0.040	0.054	0.065	0.066	0.047	0.075	0.084	0.076	0.095	0.084	0.095	0.117
SEWMA	0.042	0.059	0.065	0.062	0.045	0.061	0.069	0.065	0.090	0.068	0.084	0.109
DEWMA	0.042	0.061	0.071	0.067	0.047	0.069	0.088	0.074	0.088	0.079	0.101	0.130
TEWMA	0.034	0.062	0.054	0.055	0.052	0.073	0.083	0.075	0.070	0.073	0.092	0.134
RNN-12-1	0.042	0.055	0.054	0.059	0.059	0.071	0.070	0.070	0.097	0.085	0.096	0.116
RNN-12-2	0.035	0.062	0.055	0.066	0.073	0.076	0.075	0.076	0.107	0.081	0.093	0.115
RNN-24-1	0.037	0.061	0.052	0.052	0.067	0.067	0.054	0.072	0.116	0.092	0.079	0.088
RNN-24-2	0.046	0.066	0.060	0.050	0.076	0.080	0.074	0.074	0.110	0.085	0.081	0.089

Tabla 23: Media de MRSEE para cada mes y modelo en las series del Grupo 1

	1	2	3	4	5	6	7	8	9	10	11	12
ARIMA	0.072	0.071	0.087	0.079	0.075	0.082	0.092	0.098	0.092	0.099	0.099	0.121
SEWMA	0.067	0.068	0.084	0.078	0.069	0.072	0.081	0.089	0.083	0.077	0.090	0.132
DEWMA	0.069	0.074	0.088	0.085	0.079	0.086	0.106	0.107	0.100	0.110	0.118	0.159
TEWMA	0.066	0.070	0.083	0.076	0.089	0.093	0.100	0.106	0.098	0.109	0.121	0.169
RNN-12-1	0.076	0.073	0.089	0.089	0.090	0.098	0.104	0.110	0.114	0.113	0.118	0.156
RNN-12-2	0.070	0.085	0.110	0.091	0.085	0.116	0.108	0.093	0.111	0.108	0.113	0.151
RNN-24-1	0.069	0.077	0.087	0.084	0.094	0.092	0.097	0.110	0.114	0.105	0.102	0.140
RNN-24-2	0.074	0.085	0.096	0.092	0.106	0.117	0.106	0.103	0.125	0.118	0.100	0.133

Tabla 24: Media de MRSEE para cada mes y modelo en las series del Grupo 2

Según la información de la tabla 23, y en base a la media de MRSEE, los mejores modelos para cada uno de los 12 meses (en las series del Grupo 1), en orden cronológico (del mes 1 al 12) son: TEWMA, ARIMA, RNN-24-1, RNN-24-2, SEWMA, SEMWA, RNN-24-1, SEWMA, TEWMA, SEMWA, RNN-24-1 y RNN-24-1.

Realizando el análisis anterior, y con la información de la tabla 21, los mejores modelos (según la media) en orden cronológico (mes 1 al 12) son: TWEMA, SEWMA, TEWMA, TEWMA, SEWMA, SEWMA, SEWMA, SEWMA, SEWMA, SEWMA y ARIMA.

Se procede a realizar el test-t Student's sobre cada uno de los meses para las series del Grupo 2, para cada uno de los modelos. Los resultados se adjuntan en la sección de Anexos. Observando dicha información se puede obtener las siguientes conclusiones para cada mes:

- Mes 1: No existe un modelo que sea significativamente distinto al resto. Solo existen 2 casos donde se presentan p-valores por debajo del umbral, sin embargo, no es suficiente para poder generar afirmaciones de que un modelo es diferente en rendimiento al resto.
- Mes 2: En este mes, la distribución de los errores de la redes RNN-12-2 y RNN-24-2 presentan una diferencia significativa con el resto de los modelos, sin embargo, la media de la métrica MRSEE es mayor al resto en ambos casos, motivo que podría llevar a concluir que, para este mes, no es recomendable utilizar ambas configuraciones.

- Mes 3: El análisis es similar al del mes 2, solo que en este caso es solo la red RNN-12-2 la que es significativamente distinta al resto, por lo que a priori, según la información, no es recomendable utilizar este modelo.
- Mes 4: No existen diferencias concluyentes, por lo que cualquier modelo podría ser apto para utilizar en este mes.
- Mes 5: Para este mes, los modelos SEWMA y ARIMA presentan diferencias significantes con casi todos los modelos. A priori para este nivel de proyección serian algoritmos recomendables.
- Mes 6: El modelo SEWMA es significativamente distinto al resto, y el con menor media.
- Mes 7: Mismo resultado que en el mes 6.
- Mes 8: El modelo de SEWMA presenta diferencias significativas con 4 modelos, pero no con las configuraciones RNN-12-2, RNN-24-2 y ARIMA, por lo que no se pueden obtener resultados concluyentes.
- Mes 9: El modelo SEWMA presenta diferencias significativas con todos los modelos (menos con ARIMA). según esta información, es el mejor modelo para esta proyección.
- Mes 10: SEWMA es significativamente distinto al resto y con mejor desempeño.
- Mes 11: No se puede extraer conclusiones.
- Mes 12: El modelo ARIMA es significativamente distinto y con menor media al resto de todos los modelos (salvo SEWMA).

## VI.5 Conclusiones y futuros trabajos

Las conclusiones más relevantes al momento de realizar el estudio y observar los resultados son las siguientes:

- Con respecto al número de series del Grupo 1 donde cada modelo es mejor, los algoritmos de suavización exponencial triple, RNN-12-1 y RNN-12-2 tienen mejores resultados para los meses más cercanos (1 a 3 meses), mientras que la red RNN-24-2 es la mejor para proyecciones distantes (10 a 12 meses).
- Para las series del Grupo 2, y según el número de series donde cada modelo es mejor, el modelo de suavización exponencial simple es mejor, mientras que para proyecciones

largas RNN-24-2 y ARIMA tuvieron mejores resultados. En base a estas conclusiones, la red RNN-24-2 es un buen modelo para proyectar demandas lejanas.

- Antes de seleccionar un modelo por sobre otro, utilizando comparación de medias de errores, es relevante realizar una comparación de la distribución de estos mediante tests estadísticos. A nivel global, es el caso para las series del Grupo 1, donde, si bien existían diferencias entre las medias de los errores en los modelos, estas no eran estadísticamente significativas.
- A nivel global si bien no hay diferencias significativas para las series del Grupo 1, si existen diferencias en el local E en términos de la distribución de los errores. En dicho local, las redes neuronales tienen mejor rendimiento que los modelos clásicos de serie de tiempo.
- A nivel global, si existen diferencias significativas en las series del Grupo 2, siendo los modelos SEWMA y ARIMA los que presentan un mejor rendimiento en término de la media de NRMSE.
- Con respecto al nivel de significancia cuando se desagrega a nivel mensual, en términos generales no existen un patrón marcado de cual modelo es mejor para cada uno de los 12 meses que se proyectó.

Para futuros trabajos se recomienda utilizar el método de validación cruzada, con el propósito de generar varios registros de métricas de errores en cada una de las series en estudio, y así poder dar un mayor nivel de validez al proceso de test de hipótesis en la distribución de los errores. El uso global de cada una de las series para estos análisis puede generar distorsiones al momento de calcular los estadísticos necesarios.

Además de lo anterior, se recomienda probar con diferentes arquitecturas de redes neuronales recurrentes, entre ellas usar la forma Many-to-Many y así evitar el uso de estimaciones como inputs para las redes. Junto con lo anterior, se recomienda utilizar la red LSTM (long short-term memory) debido a que esta es capaz de conservar mejor la información histórica pasada en comparación con un red neuronal recurrente tradicional.

Para finalizar, se recomienda identificar o clasificar aquellos locales que presenten una mayor desviación o varianza en el comportamiento de venta, así también como las categorías mas “conflictivas”. Para aquellos casos se recomienda usar redes neuronales recurrentes por sobre los modelos tradicionales de series de tiempo como lo son la suavización exponencial o ARIMA.





# Bibliografía

- Ramos, Patrícia y Santos, Nicolau y Rebelo, Rui. (2015). Performance of state space and ARIMA models for consumer retail sales forecasting. *Robotics and Computer-Integrated Manufacturing*. 34. 10.1016/j.rcim.2014.12.015.
- ma, Shaohui y Kolassa, Stephan y Fildes, Robert. (2018). Retail forecasting: research and practice.
- D. Lambert, J. Stock, L. Elbron, *Fundamentals of Logistics Management*, Irwin McGraw-Hill, 1998.
- Gerencia general de la compañía, comunicación personal, diciembre 2020.
- Adhikari, Ratnadip y Agrawal, R. (2013). *An Introductory Study on Time series Modeling and Forecasting*.
- Montgomery, Jennings y Kulahci (2008). *Introduction to Time Series Analysis and Forecasting*.
- Hyndman, R.J., y Athanasopoulos, G. (2018) *Forecasting: principles and practice*, 2da edición, OTexts: Melbourne, Australia. [OTexts.com/fpp2](http://OTexts.com/fpp2). Acceso en Agosto de 2020.
- Colah, Christopher (27 de agosto de 2015). Understanding LSTM Networks. Recuperado de: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- Karpathy, Andrej. (21 de mayo de 2015). The unreasonable effectiveness of Recurrent Neural Networks. Recuperado de: <http://karpathy.github.io/2015/05/21/rnn-effectiveness>.
- Smola y Vishwanathan (2008). *Introduction to Machine Learning*.
- Werbos, Paul. (1990). *Backpropagation through time: what it does and how to do it*.
- Brownlee, Jason (7 de julio 2017). How to Scale Data for Long Short-Term Memory Networks in Python. Recuperado de: <https://machinelearningmastery.com/how-to-scale-data-for-long-short-term-memory-networks-in-python>.
- Jierula, Alipujiang y Wang, Shuhong y Oh, Tae-Min. (2021). Study on Accuracy Metrics for Evaluating the Predictions of Damage Locations in Deep Piles Using

Artificial Neural Networks with Acoustic Emission Data. Applied Sciences. 11. 2314. 10.3390/app11052314.

- Atiya, Amir y El-Shoura, Suzan y Shaheen, Samir y Mohamed, Soha. (1999). A Comparison Between Neural-Network Forecasting Techniques-Case Study: River Flow Forecasting, Neural Networks, IEEE Transactions on neural networks.
- Yunpeng, Liu y Hou, Di y Junpeng, Bao y Yong, Qi. (2017). Multi-step Ahead Time Series Forecasting for Different Data Patterns Based on LSTM Recurrent Neural Network. 305-310. 10.1109/WISA.2017.25.
- Bengio, Yoshua. (2012). Practical Recommendations for Gradient Based Training of Deep Architectures.
- Brownlee, Jason (1 de septiembre de 2020). Hypothesis Test for Comparing Machine Learning Algorithms. Recuperado de: <https://machinelearningmastery.com/hypothesis-test-for-comparing-machine-learning-algorithms/>
- Brownlee, Jason (16 de mayo de 2018). How to Calculate Nonparametric Statistical Hypothesis Tests in Python. Recuperado de: <https://machinelearningmastery.com/nonparametric-statistical-significance-tests-in-python/>

# Anexos

## Anexo A-Descripción de Variables

Grupo2	Unidades vendidas
Barras y Snack Proteicos	293.207
Bebidas Funcionales	6.507
Contenedores	49.478
Cuidado Óseo y Articular	21.719
Defensas	17.118
Dieta	55.113
Energía, Potencia y Resistencia	32.387
Ganadores de Peso	6.820
Geles	52.275
Multivitamínicos y Energía	28.634
Nutricosmética	8.969
Probióticos	5.242
Proteínas Aisladas	15.344
Proteínas Líquidas	13.902
Proteínas Vegana	5.824
Proteínas de Carne	7.469
Recuperación	25.391
Salud Cardiovascular	16.455
Superalimentos	17.279
Whey Protein	54.595

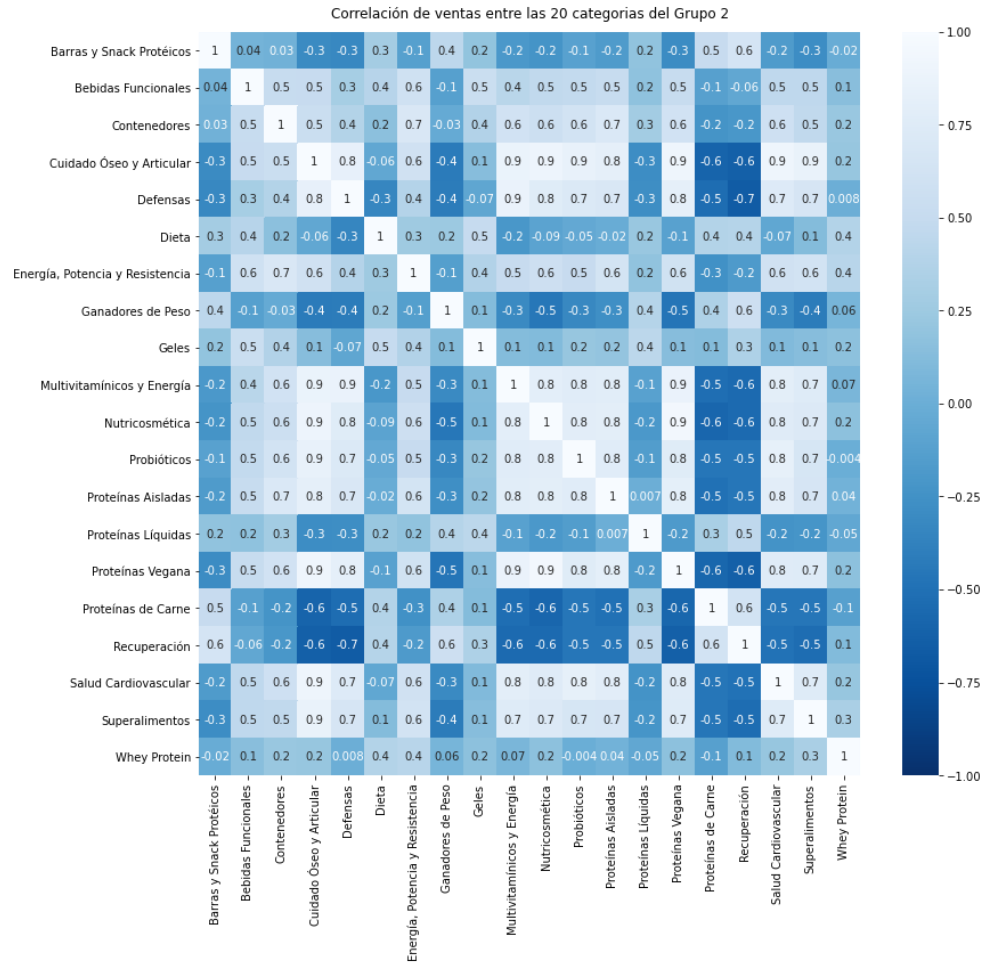
Anexo A.1: Unidades vendidas, para las 20 categorías de la variable Grupo 2

Grupo3	Unidades vendidas
Acai	285
Alimentos Proteicos	646
Alimentos de uso médico	274
Aloe Vera	225
Aminoácidos	10.449
Arginina	2.881
B Complex	6.289

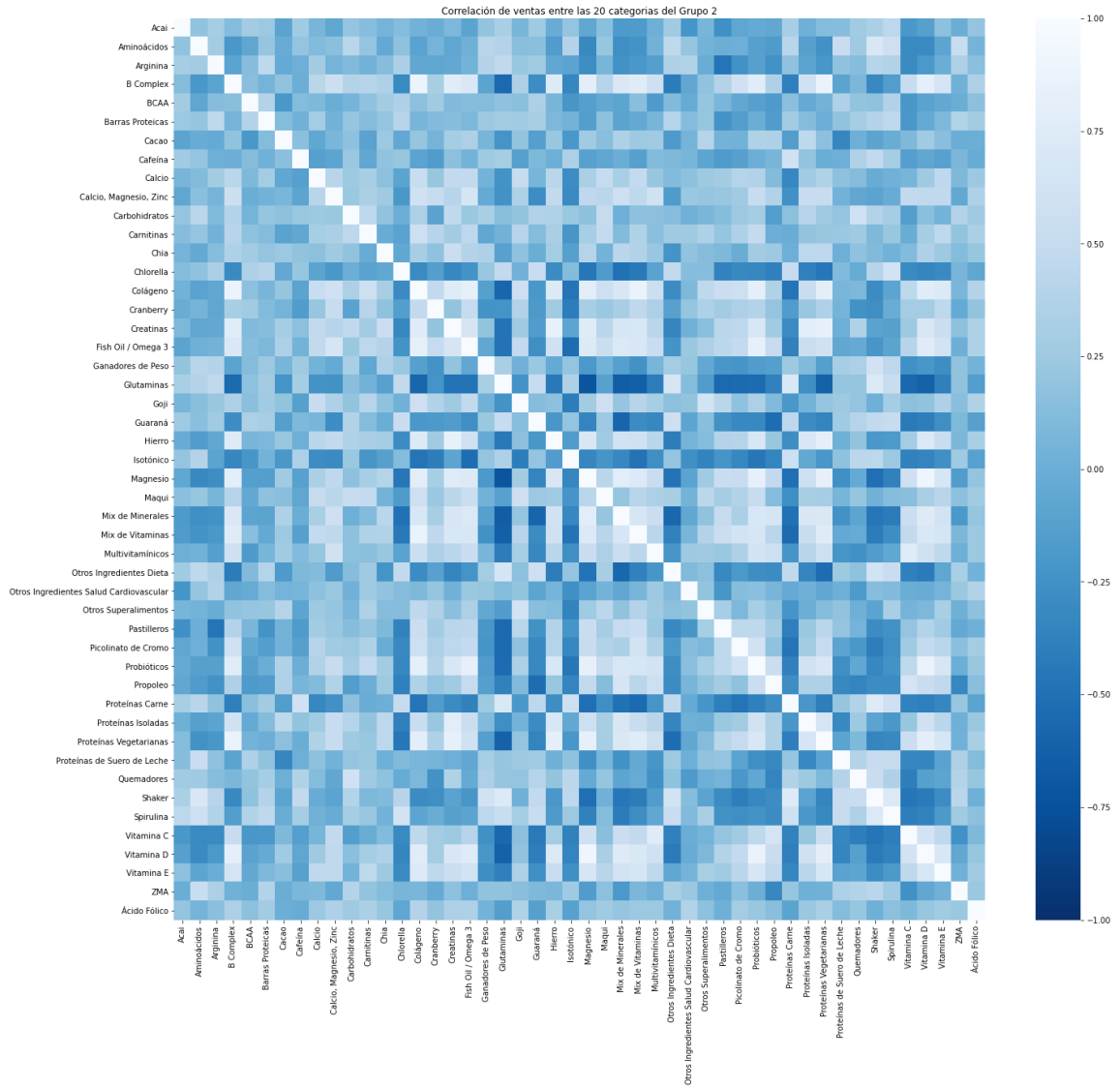
BCAA	8.353
Barras Proteicas	283.949
Berries	184
Bolsos	7.609
Botella	94
Cacao	518
Cafeína	9.554
Calcio	2.326
Calcio, Magnesio, Zinc	3.410
Camu Camu	218
Carbohidratos	52.966
Carnitinas	6.635
Chia	1.391
Chlorella	1.494
Clorofila	189
Coco	1.064
Colágeno	11.833
Cranberry	2.831
Creatinas	13.657
Curcuma	229
Fish Oil / Omega 3	14.910
Ganadores de Peso	6.820
Gingseng	1.159
Glutaminas	5.236
Goji	541
Guaraná	4.778
Hierro	572
Isotónico	4.519
Jengibre	141
Lucuma	32
Magnesio	7.325
Maqui	1.939
Mix de Minerales	3.241
Mix de Vitaminas	3.935
Multivitamínicos	14.795
Otros Accesorios	10
Otros Ingredientes Dieta	10.394

Otros Ingredientes Salud Cardiovascular	1.314
Otros Superalimentos	1.934
Pastilleros	611
Picolinato de Cromo	1.362
Probióticos	5.104
Propóleo	997
Proteínas Carne	9.863
Proteínas Isoladas	22.368
Proteínas Mezcla	1.102
Proteínas Vegetarianas	5.178
Proteínas de Suero de Leche	57.977
Quemadores	36.665
Shaker	41.154
Snack Proteico	9.258
Spirulina	7.533
Stevia	689
Té	51
Vitamina C	9.041
Vitamina D	3.607
Vitamina E	3.233
ZMA	3.791
Zinc	666
Ácido Fólico	330

Anexo A.2: Unidades vendidas, para las 67 categorías de la variable Grupo 3



Anexo A.3: Correlación de las 20 categorías del Grupo 2



Anexo A.4: Correlación de 48 categorías del Grupo 3

## Anexo B-Resultados y análisis

	ARIMA	SEWMA	DEWMA	TEWMA	RNN-12-1	RNN-12-2	RNN-24-1	RNN-24-2
A Salud	0.19	0.13	0.19	0.18	0.12	0.18	0.11	0.29
A Deporte	0.17	0.16	0.17	0.17	0.15	0.14	0.17	0.16
A Dieta	0.26	0.21	0.22	0.24	0.20	0.24	0.20	0.20
A Consumible	0.18	0.18	0.18	0.18	0.18	0.16	0.31	0.27
A ByC	0.22	0.27	0.22	0.25	0.30	0.30	0.29	0.27
A Accesorio	0.73	0.52	0.63	0.59	0.48	0.67	0.39	0.35
B Salud	0.19	0.22	0.19	0.20	0.33	0.25	0.33	0.41
B Deporte	0.14	0.14	0.15	0.16	0.14	0.14	0.15	0.15
B Dieta	0.24	0.21	0.21	0.16	0.25	0.28	0.28	0.37
B Consumible	0.25	0.22	0.25	0.38	0.23	0.20	0.21	0.21
B ByC	0.48	0.53	0.47	0.49	0.52	0.54	0.53	0.67
B Accesorio	0.32	0.31	0.30	0.31	0.37	0.33	0.41	0.37
C Salud	0.27	0.29	0.38	0.33	0.56	0.42	0.30	0.24
C Deporte	0.22	0.22	0.22	0.24	0.25	0.23	0.21	0.19
C Dieta	0.52	0.51	0.52	0.51	0.45	0.45	0.43	0.45
C Consumible	0.37	0.31	0.61	0.35	0.32	0.37	0.33	0.28
C ByC	0.17	0.19	0.15	0.18	0.30	0.16	0.21	0.15
C Accesorio	0.28	0.28	0.30	0.35	0.35	0.39	0.28	0.39
D Salud	0.25	0.30	0.25	0.24	0.33	0.29	0.30	0.28
D Deporte	0.19	0.17	0.19	0.16	0.18	0.26	0.17	0.27
D Dieta	0.49	0.23	0.23	0.24	0.30	0.38	0.22	0.29
D Consumible	0.23	0.22	0.23	0.27	0.23	0.20	0.23	0.32
D ByC	0.58	0.64	0.58	0.61	0.61	0.56	0.63	0.53
D Accesorio	0.28	0.29	0.32	0.27	0.30	0.41	0.35	0.33
E Salud	0.76	0.38	0.75	0.67	0.64	0.93	0.68	0.66
E Deporte	0.19	0.19	0.21	0.20	0.18	0.18	0.19	0.19
E Dieta	0.24	0.26	0.25	0.28	0.26	0.29	0.25	0.25
E Consumible	0.29	0.29	0.28	0.29	0.20	0.29	0.23	0.34
E ByC	0.20	0.18	0.20	0.20	0.21	0.19	0.19	0.21
E Accesorio	0.52	0.58	0.69	0.52	0.39	0.31	0.28	0.27

Anexo B.1: NRMSE para las 30 series del Grupo 1



	ARIMA	SEWMA	DEWMA	TEWMA	RNN-12-1	RNN-12-2	RNN-24-1	RNN-24-2
A Salud	98.70	63.82	95.98	90.82	56.72	84.30	52.96	157.69
A Deporte	82.04	76.02	84.70	68.36	81.51	76.47	73.62	77.90
A Dieta	36.77	31.19	31.96	37.01	30.32	38.65	32.14	35.60
A Consumible	229.53	228.58	229.25	209.14	230.16	242.39	379.95	350.72
A ByC	9.38	11.14	9.37	10.99	12.70	12.61	12.43	10.93
A Accesorio	185.04	126.49	151.84	138.61	107.06	163.43	100.28	103.28
B Salud	55.86	59.16	55.91	55.18	100.95	71.08	96.52	130.73
B Deporte	55.97	55.08	53.86	55.93	57.43	63.77	61.69	66.76
B Dieta	35.90	32.17	32.93	21.40	38.95	39.98	37.08	47.93
B Consumible	159.87	145.54	158.17	264.87	152.40	153.32	140.09	200.23
B ByC	9.78	11.18	9.61	9.49	10.34	11.18	11.06	14.71
B Accesorio	58.63	57.55	55.91	55.26	71.94	65.16	75.99	69.24
C Salud	65.87	72.88	96.89	87.52	143.10	103.61	71.54	60.23
C Deporte	36.20	36.13	36.06	40.28	42.48	40.19	34.79	35.17
C Dieta	28.42	28.08	28.20	27.15	25.14	25.47	24.11	24.51
C Consumible	133.34	113.18	229.78	112.72	119.48	134.49	121.00	103.03
C ByC	3.40	3.96	3.81	4.15	6.79	3.60	4.47	3.39
C Accesorio	13.06	13.63	14.48	17.40	16.44	18.47	13.73	18.77
D Salud	64.09	74.42	64.14	66.20	82.52	73.62	76.47	88.93
D Deporte	66.71	60.06	66.88	55.09	61.09	78.48	60.87	83.68
D Dieta	53.30	21.57	21.78	27.98	37.33	46.33	31.35	34.62
D Consumible	159.64	158.94	159.25	192.87	157.10	166.13	157.00	191.70
D ByC	22.48	25.56	21.99	24.14	23.64	20.73	25.34	19.25
D Accesorio	41.41	45.79	52.21	36.72	39.28	54.42	49.96	51.33
E Salud	708.66	352.38	715.94	632.94	602.51	901.85	607.99	651.64
E Deporte	125.18	124.92	127.31	124.02	124.92	124.94	128.49	129.29
E Dieta	46.66	48.95	47.91	44.55	49.36	54.50	47.67	55.72
E Consumible	439.60	438.05	431.94	403.66	313.08	442.58	441.21	679.37
E ByC	15.62	14.84	15.60	15.40	16.13	15.44	15.89	18.17
E Accesorio	169.69	190.66	232.18	155.70	112.23	79.14	91.65	86.99

Anexo B.2: MAE para las 30 series del Grupo 1

	ARIMA	SEWMA	DEWMA	TEWMA	RNN-12-1	RNN-12-2	RNN-24-1	RNN-24-2
A Salud	0.18	0.11	0.17	0.16	0.11	0.17	0.09	0.27
A Deporte	0.15	0.13	0.16	0.13	0.14	0.10	0.12	0.12
A Dieta	0.25	0.20	0.21	0.23	0.19	0.22	0.19	0.19
A Consumible	0.17	0.17	0.17	0.16	0.17	0.14	0.22	0.18
A ByC	0.16	0.18	0.16	0.18	0.20	0.20	0.19	0.18
A Accesorio	0.93	0.65	0.79	0.73	0.56	0.81	0.32	0.30

B Salud	0.16	0.16	0.16	0.15	0.27	0.17	0.25	0.35
B Deporte	0.13	0.13	0.13	0.13	0.13	0.12	0.13	0.14
B Dieta	0.21	0.21	0.21	0.15	0.25	0.25	0.28	0.35
B Consumible	0.19	0.19	0.19	0.33	0.19	0.19	0.19	0.18
B ByC	0.30	0.35	0.30	0.29	0.29	0.37	0.34	0.48
B Accesorio	0.34	0.35	0.36	0.36	0.35	0.30	0.37	0.36
C Salud	0.26	0.29	0.38	0.34	0.55	0.40	0.29	0.23
C Deporte	0.21	0.21	0.21	0.23	0.25	0.22	0.20	0.18
C Dieta	0.62	0.61	0.61	0.59	0.51	0.54	0.51	0.53
C Consumible	0.37	0.31	0.62	0.32	0.32	0.36	0.32	0.28
C ByC	0.12	0.13	0.15	0.15	0.24	0.12	0.15	0.11
C Accesorio	0.26	0.27	0.31	0.38	0.38	0.43	0.28	0.39
D Salud	0.18	0.19	0.18	0.19	0.20	0.22	0.23	0.27
D Deporte	0.17	0.15	0.17	0.14	0.16	0.23	0.17	0.24
D Dieta	0.41	0.21	0.21	0.25	0.31	0.40	0.21	0.25
D Consumible	0.21	0.21	0.21	0.25	0.20	0.19	0.20	0.27
D ByC	0.41	0.48	0.40	0.45	0.43	0.36	0.48	0.33
D Accesorio	0.25	0.28	0.33	0.23	0.19	0.23	0.25	0.29
E Salud	0.77	0.40	0.78	0.69	0.66	0.97	0.67	0.70
E Deporte	0.18	0.18	0.20	0.19	0.17	0.18	0.18	0.17
E Dieta	0.25	0.26	0.26	0.25	0.25	0.27	0.25	0.24
E Consumible	0.30	0.29	0.29	0.27	0.20	0.30	0.22	0.32
E ByC	0.18	0.16	0.18	0.17	0.18	0.17	0.16	0.20
E Accesorio	0.62	0.69	0.83	0.60	0.44	0.33	0.26	0.25

Anexo B.3: MAPE para las 30 series del Grupo 1

	ARIMA	SEWMA	DEWMA	TEWMA	RNN-12-1	RNN-12-2	RNN-24-1	RNN-24-2
A Multivitamínicos y Energía	0.16	0.17	0.20	0.22	0.28	0.13	0.33	0.15
A Proteínas Aisladas	0.29	0.31	0.29	0.32	0.37	0.34	0.28	0.38
A Recuperación	0.20	0.20	0.21	0.22	0.18	0.27	0.26	0.23
A Dieta	0.26	0.21	0.22	0.24	0.20	0.24	0.20	0.20
A Proteínas Líquidas	0.41	0.31	0.31	0.40	0.45	0.39	0.30	0.34
A Superalimentos	0.34	0.27	0.42	0.43	0.18	0.24	0.29	0.31
A Energía, Potencia y Resistencia	0.23	0.21	0.23	0.22	0.26	0.21	0.24	0.22
A Probióticos	0.29	0.25	0.29	0.29	0.29	0.61	0.64	0.35
A Barras y Snack Protéicos	0.17	0.17	0.19	0.16	0.16	0.22	0.18	0.30
A Proteínas de Carne	0.34	0.52	0.29	0.32	0.36	0.57	0.39	0.68
A Cuidado Óseo y Articular	0.20	0.16	0.19	0.24	0.16	0.17	0.17	0.19
A Salud Cardiovascular	0.33	0.26	0.31	0.42	0.62	0.26	0.26	0.25
A Nutricosmética	0.22	0.27	0.22	0.25	0.30	0.30	0.29	0.27
A Defensas	0.36	0.38	0.36	0.33	0.43	0.37	0.39	0.40

A Geles	0.41	0.43	0.46	0.42	0.68	0.56	0.54	0.70
A Ganadores de Peso	0.25	0.28	0.28	0.28	0.25	0.38	0.20	0.24
A Proteínas Vegana	0.27	0.33	0.23	0.25	0.38	0.33	0.21	0.24
A Contenedores	0.73	0.52	0.63	0.59	0.48	0.67	0.39	0.35
A Whey Protein	0.24	0.20	0.18	0.18	0.20	0.20	0.19	0.19
A Bebidas Funcionales	0.51	0.39	0.36	0.42	0.63	0.92	0.56	0.65
B Multivitamínicos y Energía	0.37	0.34	0.31	0.36	0.49	0.55	0.53	0.55
B Proteínas Aisladas	0.35	0.27	0.35	0.37	0.33	0.29	0.33	0.25
B Recuperación	0.19	0.22	0.19	0.21	0.38	0.26	0.26	0.22
B Dieta	0.24	0.21	0.21	0.16	0.25	0.28	0.28	0.37
B Proteínas Líquidas	0.35	0.57	0.63	0.58	1.13	0.85	0.48	0.89
B Superalimentos	0.32	0.28	0.32	0.33	0.27	0.36	0.40	0.38
B Energía, Potencia y Resistencia	0.16	0.16	0.18	0.21	0.27	0.23	0.32	0.24
B Probióticos	0.32	0.35	0.87	0.41	1.10	0.68	0.42	0.73
B Barras y Snack Protéicos	0.25	0.26	0.30	0.28	0.26	0.28	0.27	0.27
B Proteínas de Carne	0.48	0.74	0.45	0.74	0.61	0.52	0.54	0.45
B Cuidado Óseo y Articular	0.20	0.21	0.21	0.21	0.34	0.24	0.22	0.62
B Salud Cardiovascular	0.32	0.36	0.34	0.34	0.38	0.37	0.48	0.38
B Nutricosmética	0.48	0.53	0.47	0.49	0.52	0.54	0.53	0.67
B Defensas	0.57	0.59	0.57	0.57	0.60	0.59	0.67	0.59
B Geles	0.34	0.30	0.31	0.33	0.32	0.43	0.27	0.56
B Ganadores de Peso	0.35	0.36	0.36	0.36	0.51	0.68	0.45	0.41
B Proteínas Vegana	0.29	0.27	0.38	0.40	0.36	0.37	0.65	0.39
B Contenedores	0.32	0.31	0.30	0.31	0.37	0.33	0.41	0.37
B Whey Protein	0.25	0.26	0.26	0.29	0.30	0.25	0.26	0.26
B Bebidas Funcionales	0.67	0.44	0.51	0.63	0.39	1.74	0.57	0.92
C Multivitamínicos y Energía	0.45	0.35	0.32	0.28	0.39	0.40	0.34	0.36
C Proteínas Aisladas	0.29	0.28	0.64	0.71	0.27	0.39	0.58	0.29
C Recuperación	0.32	0.43	0.28	0.42	0.45	0.42	0.39	0.47
C Dieta	0.52	0.51	0.52	0.51	0.45	0.45	0.43	0.45
C Proteínas Líquidas	0.51	0.54	0.51	0.72	0.53	0.70	0.54	0.57
C Superalimentos	0.32	0.33	0.36	0.41	0.21	0.25	0.23	0.27
C Energía, Potencia y Resistencia	0.29	0.29	0.29	0.32	0.30	0.29	0.29	0.30
C Probióticos	0.56	0.47	1.05	1.24	1.23	0.40	0.78	0.86
C Barras y Snack Protéicos	0.32	0.30	0.38	0.35	0.30	0.29	0.26	0.38
C Proteínas de Carne	0.63	0.46	0.76	0.83	0.73	0.56	0.47	0.42
C Cuidado Óseo y Articular	0.41	0.30	0.58	0.65	0.55	0.64	0.90	0.22
C Salud Cardiovascular	0.58	0.35	0.43	0.49	0.17	0.25	0.15	0.15
C Nutricosmética	0.17	0.19	0.15	0.18	0.30	0.16	0.21	0.15
C Defensas	0.30	0.30	0.32	0.27	0.31	0.31	0.33	0.38
C Geles	0.45	0.42	1.22	0.51	0.53	0.52	0.35	0.35
C Ganadores de Peso	0.31	0.32	0.32	0.40	0.39	0.36	0.31	0.31
C Proteínas Vegana	0.37	0.42	0.36	0.29	0.44	0.44	0.33	0.37
C Contenedores	0.28	0.28	0.30	0.35	0.35	0.39	0.28	0.39

C Whey Protein	0.37	0.47	0.43	0.45	0.44	0.51	0.44	0.46
C Bebidas Funcionales	0.66	0.58	1.06	0.62	0.65	0.88	0.82	1.72
D Multivitamínicos y Energía	0.36	0.30	0.27	0.24	0.37	0.30	0.31	0.32
D Proteínas Aisladas	0.37	0.22	0.60	0.26	0.28	0.34	0.94	0.70
D Recuperación	0.36	0.38	0.38	0.35	0.43	0.61	0.39	0.37
D Dieta	0.49	0.23	0.23	0.24	0.30	0.38	0.22	0.29
D Proteínas Líquidas	0.63	0.56	0.63	0.68	0.76	0.37	0.55	0.48
D Superalimentos	0.39	0.41	0.35	0.37	0.45	0.43	0.41	0.40
D Energía, Potencia y Resistencia	0.21	0.20	0.21	0.23	0.22	0.32	0.21	0.32
D Probióticos	0.34	0.32	0.32	0.31	0.33	0.65	0.34	0.32
D Barras y Snack Protéicos	0.22	0.20	0.23	0.21	0.21	0.19	0.20	0.22
D Proteínas de Carne	0.81	0.85	0.80	0.90	1.01	0.93	1.28	1.03
D Cuidado Óseo y Articular	0.31	0.39	0.31	0.31	0.36	0.35	0.40	0.40
D Salud Cardiovascular	0.26	0.30	0.27	0.30	0.36	0.30	0.34	0.30
D Nutricosmética	0.58	0.64	0.58	0.61	0.61	0.56	0.63	0.53
D Defensas	0.42	0.46	0.43	0.40	0.48	0.47	0.49	0.54
D Geles	0.61	0.46	0.46	0.43	0.48	0.51	0.49	0.53
D Ganadores de Peso	0.33	0.30	0.31	0.35	0.40	0.30	0.30	0.43
D Proteínas Vegana	0.42	0.49	0.34	0.42	0.43	0.44	0.42	0.71
D Contenedores	0.28	0.29	0.32	0.27	0.30	0.41	0.35	0.33
D Whey Protein	0.19	0.19	0.20	0.29	0.19	0.26	0.20	0.22
D Bebidas Funcionales	0.56	0.32	0.33	0.39	0.40	0.62	0.97	0.62
E Multivitamínicos y Energía	0.42	0.30	0.40	0.30	1.27	0.53	0.37	0.91
E Proteínas Aisladas	0.35	0.37	0.27	0.27	0.47	0.45	0.39	0.43
E Recuperación	0.26	0.32	0.26	0.27	0.30	0.33	0.35	0.36
E Dieta	0.24	0.26	0.25	0.28	0.26	0.29	0.25	0.25
E Proteínas Líquidas	0.32	0.32	0.34	0.42	0.35	0.34	0.42	0.52
E Superalimentos	0.66	0.49	0.66	0.61	0.71	0.78	0.69	0.86
E Energía, Potencia y Resistencia	0.23	0.23	0.24	0.24	0.23	0.25	0.24	0.24
E Probióticos	0.59	0.45	1.14	1.04	1.45	0.81	1.03	0.76
E Barras y Snack Protéicos	0.26	0.26	0.42	0.33	0.20	0.21	0.19	0.31
E Proteínas de Carne	0.40	0.53	0.38	0.38	0.48	0.66	0.44	0.73
E Cuidado Óseo y Articular	0.95	0.41	0.84	0.66	0.61	0.74	0.73	0.60
E Salud Cardiovascular	0.64	0.48	0.98	1.07	0.99	1.00	1.07	1.44
E Nutricosmética	0.20	0.18	0.20	0.20	0.21	0.19	0.19	0.21
E Defensas	0.33	0.28	0.34	0.34	0.36	0.28	0.41	0.36
E Geles	0.44	0.60	0.64	0.52	0.43	0.43	0.43	0.48
E Ganadores de Peso	0.24	0.20	0.25	0.29	0.20	0.31	0.21	0.21
E Proteínas Vegana	0.23	0.25	0.27	0.26	0.23	0.24	0.25	0.28
E Contenedores	0.52	0.58	0.69	0.52	0.39	0.31	0.28	0.27
E Whey Protein	0.29	0.26	0.30	0.33	0.26	0.27	0.32	0.32
E Bebidas Funcionales	0.49	0.63	0.75	1.49	0.53	0.64	0.30	0.70

Anexo B.4: NRMSE para las 100 series del Grupo 2

	ARIMA	SEWMA	DEWMA	TEWMA	RNN-12-1	RNN-12-2	RNN-24-1	RNN-24-2
A Multivitamínicos y Energía	18.86	20.37	25.03	27.25	36.79	19.25	43.27	19.98
A Proteínas Aisladas	26.45	25.15	24.55	26.27	29.70	28.86	24.10	32.27
A Recuperación	13.37	12.69	14.59	13.62	10.69	19.30	16.52	16.05
A Dieta	36.77	31.19	31.96	37.01	30.32	38.65	32.14	35.60
A Proteínas Líquidas	19.87	15.23	15.16	21.72	21.04	18.98	14.38	15.92
A Superalimentos	23.84	17.72	29.48	29.25	9.77	13.73	19.85	21.24
A Energía, Potencia y Resistencia	23.68	23.59	23.06	20.82	27.32	23.67	26.01	27.53
A Probióticos	9.13	7.43	9.46	9.70	8.94	24.23	22.22	10.93
A Barras y Snack Protéicos	158.83	163.40	171.53	166.75	159.79	216.47	170.92	242.09
A Proteínas de Carne	4.89	7.81	4.25	4.68	5.12	7.94	5.51	9.44
A Cuidado Óseo y Articular	22.70	18.89	21.54	26.00	18.79	18.31	21.44	20.07
A Salud Cardiovascular	22.26	17.49	21.27	29.50	40.43	16.40	16.09	16.43
A Nutricosmética	9.38	11.14	9.37	10.99	12.70	12.61	12.43	10.93
A Defensas	29.18	24.42	27.18	23.85	27.25	24.67	26.23	25.15
A Geles	80.58	83.42	88.74	84.53	131.26	109.86	104.66	139.67
A Ganadores de Peso	4.18	4.69	4.63	4.75	4.19	6.95	3.37	4.40
A Proteínas Vegana	9.54	12.00	8.90	9.72	14.77	12.39	13.25	11.80
A Contenedores	185.04	126.49	151.84	138.61	107.06	163.43	100.28	103.28
A Whey Protein	40.22	31.80	29.56	27.19	32.61	34.45	30.39	32.09
A Bebidas Funcionales	17.25	11.89	11.59	13.55	22.41	33.43	19.50	23.69
B Multivitamínicos y Energía	34.08	30.28	26.83	34.38	51.26	58.52	55.20	56.25
B Proteínas Aisladas	18.40	12.85	17.95	19.08	16.30	15.01	17.99	19.09
B Recuperación	8.19	9.88	8.21	7.94	17.72	11.90	11.11	9.59
B Dieta	35.90	32.17	32.93	21.40	38.95	39.98	37.08	47.93
B Proteínas Líquidas	4.00	6.92	7.66	6.90	14.67	10.73	5.93	9.45
B Superalimentos	9.68	8.05	9.78	9.44	7.76	13.59	12.68	14.71
B Energía, Potencia y Resistencia	17.63	18.33	19.24	22.10	28.05	21.36	36.15	25.83
B Probióticos	6.70	7.58	16.85	8.50	24.17	15.60	11.04	15.81
B Barras y Snack Protéicos	113.03	119.02	133.04	130.59	121.59	128.68	123.57	128.33
B Proteínas de Carne	3.80	6.23	3.54	5.70	5.11	4.45	4.65	3.78
B Cuidado Óseo y Articular	10.95	11.40	12.64	13.63	20.17	13.60	11.20	41.45
B Salud Cardiovascular	13.37	13.89	13.58	13.62	15.34	14.65	18.00	14.51
B Nutricosmética	9.78	11.18	9.61	9.49	10.34	11.18	11.06	14.71
B Defensas	19.72	18.26	19.56	19.24	18.04	18.38	22.41	18.49
B Geles	56.48	47.75	50.19	54.03	49.39	80.63	47.35	81.81
B Ganadores de Peso	6.53	6.83	6.94	6.82	7.87	11.63	7.30	7.21
B Proteínas Vegana	6.78	6.75	9.53	10.04	9.17	9.60	16.74	9.94
B Contenedores	58.63	57.55	55.91	55.26	71.94	65.16	75.99	69.24
B Whey Protein	39.44	38.75	40.69	44.72	44.71	40.05	38.75	39.47
B Bebidas Funcionales	9.95	5.93	7.37	9.54	4.94	26.11	11.66	12.86
C Multivitamínicos y Energía	28.96	15.83	15.83	14.39	17.53	18.30	21.05	20.96
C Proteínas Aisladas	8.97	8.71	17.93	19.85	8.88	10.17	14.69	8.33
C Recuperación	6.43	8.71	5.79	8.99	9.17	8.64	8.02	9.95

C Dieta	28.42	28.08	28.20	27.15	25.14	25.47	24.11	24.51
C Proteínas Líquidas	5.18	5.25	4.96	6.20	5.25	6.33	5.25	5.25
C Superalimentos	9.40	9.49	10.89	13.61	7.10	7.75	6.07	7.57
C Energía, Potencia y Resistencia	11.87	11.86	12.36	13.88	11.59	11.85	11.85	12.09
C Probióticos	8.89	7.42	18.52	21.80	19.03	5.75	12.47	11.99
C Barras y Snack Protéicos	80.30	75.93	95.49	77.33	78.80	80.03	68.81	96.45
C Proteínas de Carne	2.45	1.58	2.96	2.87	2.57	1.88	1.90	1.60
C Cuidado Óseo y Articular	24.54	17.80	35.33	39.97	34.00	40.82	56.80	14.67
C Salud Cardiovascular	20.31	11.94	14.57	17.30	9.68	8.67	9.44	9.51
C Nutricosmética	3.40	3.96	3.81	4.15	6.79	3.60	4.47	3.39
C Defensas	16.75	15.67	16.81	15.11	15.86	17.10	16.33	19.59
C Geles	32.28	28.29	110.65	40.62	38.50	39.73	24.98	34.07
C Ganadores de Peso	1.69	1.72	1.70	2.46	2.18	2.01	1.67	1.67
C Proteínas Vegana	6.47	7.35	6.37	5.55	8.09	7.60	6.96	7.51
C Contenedores	13.06	13.63	14.48	17.40	16.44	18.47	13.73	18.77
C Whey Protein	19.07	24.59	22.49	23.54	23.07	27.00	23.30	23.76
C Bebidas Funcionales	7.24	6.19	12.76	7.05	6.88	9.64	8.72	18.77
D Multivitamínicos y Energía	33.44	21.92	20.49	20.24	25.18	22.27	21.06	24.24
D Proteínas Aisladas	17.76	9.67	30.94	11.55	14.40	16.12	45.65	32.38
D Recuperación	14.81	15.25	15.39	14.74	18.37	26.69	15.33	14.82
D Dieta	53.30	21.57	21.78	27.98	37.33	46.33	31.35	34.62
D Proteínas Líquidas	33.13	29.19	33.09	39.12	41.49	21.13	28.18	26.09
D Superalimentos	14.19	14.43	13.29	13.28	16.09	17.02	14.28	15.20
D Energía, Potencia y Resistencia	20.04	18.67	19.89	20.96	19.89	24.01	18.68	23.69
D Probióticos	5.79	5.45	5.38	5.17	5.63	10.92	5.73	5.50
D Barras y Snack Protéicos	115.33	105.73	123.36	122.68	112.01	112.39	104.86	117.82
D Proteínas de Carne	8.86	9.30	8.81	9.53	11.32	10.11	12.19	9.47
D Cuidado Óseo y Articular	23.30	28.55	23.30	24.46	27.38	25.00	29.09	27.47
D Salud Cardiovascular	9.71	10.07	10.06	11.03	12.22	11.85	11.44	12.49
D Nutricosmética	22.48	25.56	21.99	24.14	23.64	20.73	25.34	19.25
D Defensas	18.23	19.56	18.39	16.62	20.93	19.33	21.28	20.68
D Geles	53.80	42.81	42.76	41.88	44.82	46.78	44.11	44.46
D Ganadores de Peso	4.07	4.06	4.60	4.96	6.65	4.88	4.41	5.56
D Proteínas Vegana	12.51	14.24	10.12	12.05	12.66	13.68	12.83	21.93
D Contenedores	41.41	45.79	52.21	36.72	39.28	54.42	49.96	51.33
D Whey Protein	24.25	23.17	24.45	37.45	24.45	33.54	26.42	25.57
D Bebidas Funcionales	10.71	5.83	5.93	7.69	7.37	11.72	18.28	11.03
E Multivitamínicos y Energía	104.64	68.32	97.25	70.79	286.92	126.99	79.43	233.66
E Proteínas Aisladas	39.69	43.73	27.55	29.33	58.52	52.61	44.84	44.72
E Recuperación	20.44	25.77	20.45	20.99	23.79	27.28	28.93	27.92
E Dieta	46.66	48.95	47.91	44.55	49.36	54.50	47.67	55.72
E Proteínas Líquidas	19.05	19.26	20.65	31.63	24.22	29.17	26.14	41.53
E Superalimentos	86.06	63.03	85.89	78.61	91.14	101.11	88.11	110.96
E Energía, Potencia y Resistencia	38.47	38.58	38.72	37.43	38.71	38.61	38.85	39.41

E Probióticos	35.12	26.96	68.06	62.01	85.30	46.96	60.28	44.46
E Barras y Snack Protéicos	294.92	291.71	554.70	391.97	235.99	234.67	240.87	575.15
E Proteínas de Carne	6.83	10.05	6.51	7.25	9.11	10.72	8.51	11.05
E Cuidado Óseo y Articular	185.02	76.28	164.83	126.02	118.53	148.59	144.59	119.92
E Salud Cardiovascular	83.60	59.73	133.14	149.13	136.54	141.50	141.35	191.35
E Nutricosmética	15.62	14.84	15.60	15.40	16.13	15.44	15.89	18.17
E Defensas	55.12	46.39	55.21	55.55	55.60	42.88	68.02	61.19
E Geles	97.36	137.58	147.85	109.84	93.95	94.52	93.86	127.88
E Ganadores de Peso	6.43	5.08	6.62	7.98	5.08	10.12	5.66	5.66
E Proteínas Vegana	15.45	15.90	17.84	17.37	16.32	17.78	18.42	22.84
E Contenedores	169.69	190.66	232.18	155.70	112.23	79.14	91.65	86.99
E Whey Protein	54.66	47.62	57.52	54.86	49.30	49.78	67.26	65.24
E Bebidas Funcionales	20.25	26.41	32.29	61.15	24.68	24.44	15.07	27.49

Anexo B.5: MAE para las 100 series del Grupo 2

	ARIMA	SEWMA	DEWMA	TEWMA	RNN-12-1	RNN-12-2	RNN-24-1	RNN-24-2
A Multivitamínicos y Energía	0.14	0.13	0.17	0.18	0.23	0.11	0.28	0.11
A Proteínas Aisladas	0.28	0.23	0.25	0.26	0.27	0.26	0.25	0.28
A Recuperación	0.18	0.18	0.20	0.19	0.15	0.26	0.24	0.22
A Dieta	0.25	0.20	0.21	0.23	0.19	0.22	0.19	0.19
A Proteínas Líquidas	0.34	0.30	0.31	0.43	0.54	0.44	0.33	0.31
A Superalimentos	0.34	0.26	0.42	0.42	0.16	0.21	0.29	0.31
A Energía, Potencia y Resistencia	0.22	0.20	0.22	0.20	0.22	0.19	0.20	0.20
A Probióticos	0.28	0.22	0.29	0.28	0.24	0.50	0.67	0.33
A Barras y Snack Protéicos	0.14	0.15	0.14	0.15	0.14	0.19	0.14	0.24
A Proteínas de Carne	0.33	0.55	0.30	0.29	0.37	0.58	0.40	0.68
A Cuidado Óseo y Articular	0.19	0.14	0.17	0.21	0.15	0.15	0.15	0.17
A Salud Cardiovascular	0.32	0.25	0.30	0.42	0.60	0.25	0.25	0.23
A Nutricosmética	0.16	0.18	0.16	0.18	0.20	0.20	0.19	0.18
A Defensas	0.26	0.19	0.24	0.21	0.20	0.21	0.24	0.22
A Geles	0.62	0.69	0.77	0.67	0.55	0.49	0.55	0.47
A Ganadores de Peso	0.19	0.21	0.21	0.23	0.19	0.38	0.18	0.23
A Proteínas Vegana	0.17	0.21	0.17	0.19	0.27	0.23	0.20	0.16
A Contenedores	0.93	0.65	0.79	0.73	0.56	0.81	0.32	0.30
A Whey Protein	0.23	0.18	0.17	0.15	0.18	0.18	0.17	0.18
A Bebidas Funcionales	0.34	0.27	0.29	0.33	0.48	0.91	0.44	0.53
B Multivitamínicos y Energía	0.26	0.23	0.20	0.27	0.42	0.49	0.46	0.46
B Proteínas Aisladas	0.32	0.21	0.31	0.33	0.22	0.21	0.25	0.23
B Recuperación	0.18	0.21	0.18	0.16	0.39	0.25	0.24	0.21
B Dieta	0.21	0.21	0.21	0.15	0.25	0.25	0.28	0.35
B Proteínas Líquidas	0.35	0.70	0.78	0.61	1.41	1.04	0.60	0.80
B Superalimentos	0.33	0.27	0.34	0.32	0.25	0.32	0.44	0.36

B Energía, Potencia y Resistencia	0.15	0.15	0.17	0.18	0.21	0.19	0.26	0.20
B Probióticos	0.32	0.41	1.08	0.48	1.25	0.69	0.40	0.67
B Barras y Snack Protéicos	0.19	0.19	0.21	0.21	0.20	0.21	0.20	0.19
B Proteínas de Carne	0.60	1.19	0.62	0.85	0.97	0.84	0.88	0.70
B Cuidado Óseo y Articular	0.14	0.15	0.18	0.18	0.27	0.20	0.15	0.57
B Salud Cardiovascular	0.25	0.23	0.24	0.24	0.26	0.24	0.36	0.26
B Nutricosmética	0.30	0.35	0.30	0.29	0.29	0.37	0.34	0.48
B Defensas	0.41	0.34	0.40	0.40	0.32	0.35	0.33	0.36
B Geles	0.35	0.35	0.35	0.32	0.37	0.38	0.27	0.44
B Ganadores de Peso	0.41	0.46	0.48	0.46	0.67	0.91	0.43	0.55
B Proteínas Vegana	0.32	0.29	0.44	0.46	0.41	0.42	0.71	0.44
B Contenedores	0.34	0.35	0.36	0.36	0.35	0.30	0.37	0.36
B Whey Protein	0.26	0.24	0.28	0.26	0.26	0.24	0.24	0.26
B Bebidas Funcionales	0.68	0.65	0.77	0.71	0.53	1.65	0.49	0.84
C Multivitamínicos y Energía	0.52	0.21	0.22	0.20	0.26	0.28	0.25	0.26
C Proteínas Aisladas	0.32	0.30	0.71	0.79	0.26	0.42	0.65	0.25
C Recuperación	0.34	0.47	0.29	0.43	0.49	0.46	0.43	0.50
C Dieta	0.62	0.61	0.61	0.59	0.51	0.54	0.51	0.53
C Proteínas Líquidas	1.31	1.54	1.07	1.85	1.20	1.79	1.18	1.02
C Superalimentos	0.34	0.34	0.38	0.45	0.22	0.25	0.22	0.27
C Energía, Potencia y Resistencia	0.25	0.25	0.27	0.31	0.23	0.25	0.25	0.28
C Probióticos	0.70	0.62	1.57	1.87	1.80	0.43	1.15	0.78
C Barras y Snack Protéicos	0.35	0.33	0.41	0.35	0.32	0.31	0.26	0.41
C Proteínas de Carne	0.75	0.94	0.92	1.08	1.46	1.15	0.69	0.77
C Cuidado Óseo y Articular	0.43	0.31	0.61	0.68	0.59	0.69	0.95	0.21
C Salud Cardiovascular	0.62	0.37	0.45	0.52	0.10	0.21	0.13	0.11
C Nutricosmética	0.12	0.13	0.15	0.15	0.24	0.12	0.15	0.11
C Defensas	0.24	0.22	0.23	0.23	0.23	0.29	0.21	0.24
C Geles	0.62	0.56	1.61	0.63	0.71	0.71	0.48	0.45
C Ganadores de Peso	0.35	0.41	0.35	0.50	0.54	0.47	0.35	0.35
C Proteínas Vegana	0.31	0.31	0.29	0.34	0.31	0.35	0.30	0.27
C Contenedores	0.26	0.27	0.31	0.38	0.38	0.43	0.28	0.39
C Whey Protein	0.38	0.49	0.45	0.46	0.46	0.53	0.46	0.47
C Bebidas Funcionales	1.20	1.15	1.95	1.14	1.17	1.39	1.49	2.72
D Multivitamínicos y Energía	0.40	0.22	0.21	0.22	0.20	0.24	0.18	0.27
D Proteínas Aisladas	0.40	0.21	0.65	0.25	0.26	0.32	0.95	0.71
D Recuperación	0.40	0.42	0.42	0.37	0.47	0.64	0.43	0.41
D Dieta	0.41	0.21	0.21	0.25	0.31	0.40	0.21	0.25
D Proteínas Líquidas	0.48	0.42	0.48	0.64	0.63	0.25	0.40	0.35
D Superalimentos	0.27	0.26	0.28	0.24	0.29	0.31	0.26	0.27
D Energía, Potencia y Resistencia	0.20	0.17	0.20	0.20	0.17	0.23	0.16	0.25
D Probióticos	0.37	0.31	0.34	0.33	0.31	0.54	0.32	0.34
D Barras y Snack Protéicos	0.18	0.17	0.19	0.19	0.17	0.18	0.16	0.16
D Proteínas de Carne	1.08	1.13	1.08	1.13	1.34	1.24	1.56	1.23



D Cuidado Óseo y Articular	0.28	0.32	0.28	0.30	0.30	0.33	0.31	0.34
D Salud Cardiovascular	0.30	0.30	0.34	0.36	0.34	0.30	0.31	0.27
D Nutricosmética	0.41	0.48	0.40	0.45	0.43	0.36	0.48	0.33
D Defensas	0.29	0.29	0.29	0.25	0.28	0.34	0.30	0.44
D Geles	1.14	0.79	0.79	0.65	0.87	0.95	0.83	0.86
D Ganadores de Peso	0.28	0.27	0.32	0.35	0.46	0.29	0.30	0.27
D Proteínas Vegana	0.34	0.35	0.28	0.30	0.37	0.36	0.37	0.57
D Contenedores	0.25	0.28	0.33	0.23	0.19	0.23	0.25	0.29
D Whey Protein	0.18	0.16	0.18	0.23	0.15	0.24	0.18	0.21
D Bebidas Funcionales	0.44	0.31	0.36	0.39	0.30	0.56	0.86	0.51
E Multivitamínicos y Energía	0.44	0.29	0.41	0.30	1.15	0.53	0.35	0.94
E Proteínas Aisladas	0.28	0.31	0.19	0.21	0.42	0.37	0.32	0.35
E Recuperación	0.25	0.34	0.25	0.24	0.31	0.35	0.38	0.40
E Dieta	0.25	0.26	0.26	0.25	0.25	0.27	0.25	0.24
E Proteínas Líquidas	0.22	0.24	0.23	0.37	0.24	0.25	0.28	0.41
E Superalimentos	0.74	0.55	0.74	0.68	0.79	0.87	0.77	0.95
E Energía, Potencia y Resistencia	0.25	0.25	0.26	0.24	0.24	0.26	0.24	0.24
E Probióticos	0.68	0.52	1.30	1.18	1.62	0.93	1.16	0.87
E Barras y Snack Protéicos	0.26	0.26	0.45	0.33	0.20	0.21	0.19	0.26
E Proteínas de Carne	0.34	0.68	0.38	0.44	0.61	0.53	0.55	0.72
E Cuidado Óseo y Articular	1.01	0.44	0.91	0.70	0.66	0.81	0.79	0.66
E Salud Cardiovascular	0.76	0.56	1.17	1.30	1.20	1.22	1.26	1.68
E Nutricosmética	0.18	0.16	0.18	0.17	0.18	0.17	0.16	0.20
E Defensas	0.36	0.29	0.36	0.37	0.38	0.26	0.44	0.38
E Geles	1.46	1.95	2.04	1.43	1.36	1.38	1.40	0.81
E Ganadores de Peso	0.21	0.19	0.21	0.27	0.19	0.29	0.20	0.20
E Proteínas Vegana	0.21	0.20	0.27	0.26	0.20	0.22	0.24	0.29
E Contenedores	0.62	0.69	0.83	0.60	0.44	0.33	0.26	0.25
E Whey Protein	0.28	0.24	0.29	0.29	0.25	0.26	0.33	0.32
E Bebidas Funcionales	0.61	0.78	0.94	1.74	0.44	0.51	0.33	0.55

Anexo B.6: MAPE para las 100 series del Grupo 2

	A	B	C	D	E
Multivitamínicos y Energía	(38.14, 0.4)	(29.81, 0.41)	(19.05, 0.39)	(31.41, 0.44)	(81.87, 0.49)
Proteínas Aisladas	(29.79, 0.48)	(22.16, 0.48)	(9.51, 0.47)	(17.79, 0.56)	(31.84, 0.38)
Recuperación	(16.71, 0.17)	(18.89, 0.28)	(12.77, 0.33)	(16.55, 0.26)	(37.55, 0.28)
Dieta	(39.49, 0.21)	(48.98, 0.25)	(20.48, 0.26)	(31.03, 0.22)	(51.8, 0.19)
Proteínas Líquidas	(22.95, 0.39)	(15.26, 0.7)	(8.6, 0.71)	(23.74, 0.41)	(35.63, 0.51)
Superalimentos	(24.83, 0.41)	(13.28, 0.33)	(9.04, 0.24)	(13.34, 0.39)	(53.04, 0.53)
Energía, Potencia y Resistencia	(31.12, 0.29)	(31.52, 0.3)	(9.9, 0.22)	(29.06, 0.32)	(32.96, 0.2)
Probióticos	(15.67, 0.74)	(9.31, 1.0)	(7.52, 0.74)	(8.26, 0.86)	(29.22, 0.89)
Barras y Snack Proteicos	(303.52, 0.25)	(153.63, 0.26)	(147.46, 0.35)	(293.39, 0.37)	(447.0, 0.27)

Proteínas de Carne	(10.7, 0.47)	(13.49, 0.66)	(7.67, 0.77)	(15.01, 0.56)	(21.26, 0.55)
Cuidado Óseo y Articular	(42.27, 0.47)	(24.34, 0.55)	(23.4, 0.49)	(28.68, 0.64)	(82.8, 0.7)
Salud Cardiovascular	(23.36, 0.39)	(12.77, 0.34)	(10.89, 0.32)	(11.61, 0.39)	(53.39, 0.53)
Nutricosmética	(17.14, 0.52)	(8.35, 0.47)	(7.91, 0.42)	(16.73, 0.84)	(34.74, 0.66)
Defensas	(34.21, 0.52)	(18.12, 0.64)	(20.74, 0.5)	(22.29, 0.68)	(67.92, 0.66)
Geles	(79.28, 0.47)	(85.08, 0.48)	(43.83, 0.45)	(39.0, 0.38)	(132.46, 0.47)
Ganadores de Peso	(5.92, 0.29)	(8.45, 0.35)	(3.79, 0.45)	(7.5, 0.37)	(10.77, 0.31)
Proteínas Vegana	(18.3, 0.89)	(11.09, 0.9)	(8.28, 0.95)	(13.46, 0.93)	(32.72, 0.9)
Contenedores	(83.51, 0.51)	(49.34, 0.31)	(12.98, 0.24)	(40.01, 0.27)	(93.01, 0.36)
Whey Protein	(39.12, 0.21)	(34.81, 0.21)	(17.01, 0.23)	(32.93, 0.21)	(45.38, 0.16)
Bebidas Funcionales	(17.73, 0.66)	(14.21, 0.73)	(6.29, 0.61)	(9.35, 0.65)	(19.45, 0.6)

Anexo B.7: Desviación estándar y coeficiente de variación para las 100 series del Grupo 2

	ARIMA	SEWMA	DEWMA	TEWMA	RNN-12-1	RNN-12-2	RNN-24-1	RNN-24-2
ARIMA	-	0.20	0.29	0.15	0.48	0.67	0.48	0.75
SEWMA	0.20	-	0.65	0.62	0.04	0.59	0.75	0.21
DEWMA	0.29	0.65	-	0.43	0.10	0.79	0.97	0.33
TEWMA	0.15	0.62	0.43	-	0.05	0.43	0.56	0.18
RNN-12-1	0.48	0.04	0.10	0.05	-	0.03	0.06	0.64
RNN-12-2	0.67	0.59	0.79	0.43	0.03	-	0.79	0.39
RNN-24-1	0.48	0.75	0.97	0.56	0.06	0.79	-	0.30
RNN-24-2	0.75	0.21	0.33	0.18	0.64	0.39	0.30	-

Anexo B.8: P-valores para t-test Student's en las series del Grupo 2, y mes 1, usando MRSEE

	ARIMA	SEWMA	DEWMA	TEWMA	RNN-12-1	RNN-12-2	RNN-24-1	RNN-24-2
ARIMA	-	0.36	0.49	0.87	0.72	0.03	0.36	0.05
SEWMA	0.36	-	0.05	0.63	0.20	0.00	0.16	0.01
DEWMA	0.49	0.05	-	0.52	0.88	0.06	0.60	0.09
TEWMA	0.87	0.63	0.52	-	0.62	0.02	0.33	0.05
RNN-12-1	0.72	0.20	0.88	0.62	-	0.04	0.51	0.08
RNN-12-2	0.03	0.00	0.06	0.02	0.04	-	0.18	0.99
RNN-24-1	0.36	0.16	0.60	0.33	0.51	0.18	-	0.14
RNN-24-2	0.05	0.01	0.09	0.05	0.08	0.99	0.14	-

Anexo B.9: P-valores para t-test Student's en las series del Grupo 2, y mes 2, usando MRSEE

	ARIMA	SEWMA	DEWMA	TEWMA	RNN-12-1	RNN-12-2	RNN-24-1	RNN-24-2
ARIMA	-	0.40	0.61	0.55	0.73	0.02	0.99	0.26
SEWMA	0.40	-	0.18	0.96	0.34	0.01	0.56	0.17
DEWMA	0.61	0.18	-	0.42	0.95	0.04	0.76	0.44
TEWMA	0.55	0.96	0.42	-	0.45	0.01	0.61	0.14
RNN-12-1	0.73	0.34	0.95	0.45	-	0.03	0.71	0.44
RNN-12-2	0.02	0.01	0.04	0.01	0.03	-	0.01	0.11
RNN-24-1	0.99	0.56	0.76	0.61	0.71	0.01	-	0.22
RNN-24-2	0.26	0.17	0.44	0.14	0.44	0.11	0.22	-

Anexo B.10: P-valores para t-test Student's en las series del Grupo 2, y mes 3, usando MRSEE

	ARIMA	SEWMA	DEWMA	TEWMA	RNN-12-1	RNN-12-2	RNN-24-1	RNN-24-2
ARIMA	-	0.78	0.14	0.54	0.08	0.08	0.43	0.14
SEWMA	0.78	-	0.08	0.73	0.02	0.04	0.37	0.11
DEWMA	0.14	0.08	-	0.08	0.49	0.43	0.83	0.45
TEWMA	0.54	0.73	0.08	-	0.02	0.05	0.28	0.06
RNN-12-1	0.08	0.02	0.49	0.02	-	0.76	0.39	0.72
RNN-12-2	0.08	0.04	0.43	0.05	0.76	-	0.33	0.90
RNN-24-1	0.43	0.37	0.83	0.28	0.39	0.33	-	0.30
RNN-24-2	0.14	0.11	0.45	0.06	0.72	0.90	0.30	-

Anexo B.11: P-valores para t-test Student's en las series del Grupo 2, y mes 4, usando MRSEE

	ARIMA	SEWMA	DEWMA	TEWMA	RNN-12-1	RNN-12-2	RNN-24-1	RNN-24-2
ARIMA	-	0.19	0.38	0.04	0.02	0.12	0.00	0.00
SEWMA	0.19	-	0.05	0.00	0.00	0.01	0.00	0.00
DEWMA	0.38	0.05	-	0.19	0.12	0.49	0.04	0.00
TEWMA	0.04	0.00	0.19	-	0.88	0.62	0.51	0.06
RNN-12-1	0.02	0.00	0.12	0.88	-	0.31	0.45	0.04
RNN-12-2	0.12	0.01	0.49	0.62	0.31	-	0.16	0.00
RNN-24-1	0.00	0.00	0.04	0.51	0.45	0.16	-	0.18

RNN-24-2	0.00	0.00	0.00	0.06	0.04	0.00	0.18	-
----------	------	------	------	------	------	------	------	---

Anexo B.12: P-valores para t-test Student's en las series del Grupo 2, y mes 5, usando MRSEE

	ARIMA	SEWMA	DEWMA	TEWMA	RNN-12-1	RNN-12-2	RNN-24-1	RNN-24-2
ARIMA	-	0.03	0.41	0.08	0.04	0.00	0.12	0.00
SEWMA	0.03	-	0.01	0.00	0.00	0.00	0.00	0.00
DEWMA	0.41	0.01	-	0.18	0.11	0.01	0.28	0.01
TEWMA	0.08	0.00	0.18	-	0.52	0.05	0.93	0.05
RNN-12-1	0.04	0.00	0.11	0.52	-	0.10	0.42	0.07
RNN-12-2	0.00	0.00	0.01	0.05	0.10	-	0.03	0.92
RNN-24-1	0.12	0.00	0.28	0.93	0.42	0.03	-	0.01
RNN-24-2	0.00	0.00	0.01	0.05	0.07	0.92	0.01	-

Anexo B.13: P-valores para t-test Student's en las series del Grupo 2, y mes 6, usando MRSEE

	ARIMA	SEWMA	DEWMA	TEWMA	RNN-12-1	RNN-12-2	RNN-24-1	RNN-24-2
ARIMA	-	0.05	0.03	0.24	0.18	0.03	0.48	0.07
SEWMA	0.05	-	0.00	0.01	0.00	0.00	0.03	0.00
DEWMA	0.03	0.00	-	0.30	0.74	0.84	0.28	1.00
TEWMA	0.24	0.01	0.30	-	0.69	0.43	0.75	0.54
RNN-12-1	0.18	0.00	0.74	0.69	-	0.55	0.43	0.76
RNN-12-2	0.03	0.00	0.84	0.43	0.55	-	0.18	0.82
RNN-24-1	0.48	0.03	0.28	0.75	0.43	0.18	-	0.27
RNN-24-2	0.07	0.00	1.00	0.54	0.76	0.82	0.27	-

Anexo B.14: P-valores para t-test Student's en las series del Grupo 2, y mes 7, usando MRSEE

	ARIMA	SEWMA	DEWMA	TEWMA	RNN-12-1	RNN-12-2	RNN-24-1	RNN-24-2
ARIMA	-	0.12	0.18	0.27	0.17	0.53	0.12	0.64
SEWMA	0.12	-	0.01	0.02	0.01	0.62	0.01	0.18
DEWMA	0.18	0.01	-	0.78	0.72	0.12	0.69	0.67
TEWMA	0.27	0.02	0.78	-	0.64	0.24	0.59	0.81
RNN-12-1	0.17	0.01	0.72	0.64	-	0.02	0.97	0.47

RNN-12-2	0.53	0.62	0.12	0.24	0.02	-	0.03	0.25
RNN-24-1	0.12	0.01	0.69	0.59	0.97	0.03	-	0.40
RNN-24-2	0.64	0.18	0.67	0.81	0.47	0.25	0.40	-

Anexo B.15: P-valores para t-test Student's en las series del Grupo 2, y mes 8, usando MRSEE

	ARIMA	SEWMA	DEWMA	TEWMA	RNN-12-1	RNN-12-2	RNN-24-1	RNN-24-2
ARIMA	-	0.13	0.22	0.42	0.03	0.02	0.01	0.00
SEWMA	0.13	-	0.02	0.06	0.00	0.00	0.00	0.00
DEWMA	0.22	0.02	-	0.74	0.12	0.24	0.11	0.02
TEWMA	0.42	0.06	0.74	-	0.10	0.19	0.08	0.01
RNN-12-1	0.03	0.00	0.12	0.10	-	0.64	0.97	0.30
RNN-12-2	0.02	0.00	0.24	0.19	0.64	-	0.66	0.11
RNN-24-1	0.01	0.00	0.11	0.08	0.97	0.66	-	0.14
RNN-24-2	0.00	0.00	0.02	0.01	0.30	0.11	0.14	-

Anexo B.16: P-valores para t-test Student's en las series del Grupo 2, y mes 9, usando MRSEE

	ARIMA	SEWMA	DEWMA	TEWMA	RNN-12-1	RNN-12-2	RNN-24-1	RNN-24-2
ARIMA	-	0.00	0.18	0.33	0.19	0.32	0.49	0.08
SEWMA	0.00	-	0.00	0.00	0.00	0.00	0.00	0.00
DEWMA	0.18	0.00	-	0.87	0.76	0.83	0.51	0.43
TEWMA	0.33	0.00	0.87	-	0.67	0.96	0.67	0.46
RNN-12-1	0.19	0.00	0.76	0.67	-	0.59	0.34	0.65
RNN-12-2	0.32	0.00	0.83	0.96	0.59	-	0.69	0.28
RNN-24-1	0.49	0.00	0.51	0.67	0.34	0.69	-	0.16
RNN-24-2	0.08	0.00	0.43	0.46	0.65	0.28	0.16	-

Anexo B.17: P-valores para t-test Student's en las series del Grupo 2, y mes 10, usando MRSEE

	ARIMA	SEWMA	DEWMA	TEWMA	RNN-12-1	RNN-12-2	RNN-24-1	RNN-24-2
ARIMA	-	0.16	0.03	0.02	0.10	0.22	0.76	0.97
SEWMA	0.16	-	0.00	0.00	0.01	0.04	0.16	0.28
DEWMA	0.03	0.00	-	0.79	0.99	0.67	0.08	0.08

TEWMA	0.02	0.00	0.79	-	0.85	0.57	0.06	0.10
RNN-12-1	0.10	0.01	0.99	0.85	-	0.66	0.09	0.10
RNN-12-2	0.22	0.04	0.67	0.57	0.66	-	0.31	0.28
RNN-24-1	0.76	0.16	0.08	0.06	0.09	0.31	-	0.79
RNN-24-2	0.97	0.28	0.08	0.10	0.10	0.28	0.79	-

Anexo B.18: P-valores para t-test Student's en las series del Grupo 2, y mes 11, usando MRSEE

	ARIMA	SEWMA	DEWMA	TEWMA	RNN-12-1	RNN-12-2	RNN-24-1	RNN-24-2
ARIMA	-	0.11	0.00	0.00	0.01	0.00	0.06	0.32
SEWMA	0.11	-	0.01	0.00	0.05	0.04	0.45	0.96
DEWMA	0.00	0.01	-	0.19	0.78	0.48	0.06	0.02
TEWMA	0.00	0.00	0.19	-	0.33	0.14	0.02	0.01
RNN-12-1	0.01	0.05	0.78	0.33	-	0.69	0.15	0.05
RNN-12-2	0.00	0.04	0.48	0.14	0.69	-	0.32	0.11
RNN-24-1	0.06	0.45	0.06	0.02	0.15	0.32	-	0.43
RNN-24-2	0.32	0.96	0.02	0.01	0.05	0.11	0.43	-

Anexo B.19: P-valores para t-test Student's en las series del Grupo 2, y mes 12, usando MRSEE