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

ESTIMACIÓN DE LA DEMANDA DE PRODUCTOS PERECIBLES EN UN SUPERMERCADO

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERA CIVIL INDUSTRIAL

LORETO CATALINA VALDIVIA ESPINOSA

PROFESOR GUÍA:
ALEJANDRA PUENTE CHANDIA

MIEMBROS DE LA COMISIÓN:
CAROLINA SEGOVIA RIQUELME
CHARLES THRAVES CORTÉS-MONROY

SANTIAGO DE CHILE
2019

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGENIERA CIVIL INDUSTRIAL
POR: LORETO CATALINA VALDIVIA ESPINOSA
FECHA: 2019
PROF. GUÍA: ALEJANDRA PUENTE CHANDIA

ESTIMACIÓN DE LA DEMANDA DE PRODUCTOS PERECIBLES EN UN SUPERMERCADO

El presente trabajo tiene como objetivo estimar la demanda de frutas y verduras de un supermercado de Santiago, Chile. Estos productos, catalogados como productos perecibles, poseen la característica de tener una corta vida útil, ya que tienen la particularidad de deteriorarse rápidamente. Dadas las características de estos productos, tanto su demanda como sus precios fluctúan en el tiempo dependiendo de la fecha de vencimiento de éstos. Por ende, es de vital importancia estimar su demanda y así evitar tener problemas que puedan significar la pérdida de dinero para un supermercado. En este sentido, al estimar su demanda se pueden disminuir las mermas del supermercado y con ello, aumentar sus ganancias.

Se trabaja con datos históricos de las ventas de un supermercado chileno. Se posee un total de 630 datos, los cuales van desde octubre de 2014 a junio de 2016. De estos, 600 son utilizados como set de entrenamiento, 7 como *Lead Time* y 23 como set de testeo. Dentro de las variables explicativas que se utilizan para estimar la demanda están el precio de otros productos que se compran en conjunto con el producto analizado, la temperatura promedio del día, los días de la semana, el tipo de variación del precio con respecto al día anterior, traducidos en aumentos o descuentos de 10 %, 20 % o más del 20 %, entre otros.

Se realiza la estimación de la demanda para la Palta Hass Extra a Granel y para el Tomate a Granel. Para ello, se usan métodos de regresiones y series de tiempo para identificar los elementos clave que influyen en la demanda de estos productos perecibles y para estimar esta demanda, respectivamente. De esta manera, se utilizan once métodos de *machine learning*, los cuales son Naïve Forecast, Moving Average, ARIMA, SARIMA, Suavización Exponencial Triple, Regresiones Lineales Múltiples, Árboles de Regresiones, Random Forest, Support Vector Machine, Redes Neuronales Artificiales y Redes Neuronales Recurrentes.

Se concluye que el mejor modelo tanto para la Palta Hass Extra a Granel como para el Tomate a Granel es el modelo de Random Forest, el cual posee un error absoluto medio, MAE, en el set de testeo de alrededor de 51 kilos para las paltas y 70 kilos para el tomate. Esto es equivalente a un error porcentual absoluto medio, MAPE, de 25 % de ventas diarias para las paltas y de 22 % de ventas diarias para el tomate. Las ventas en promedio de ambos productos son de 208 kilos y 354 kilos para las paltas y los tomates, respectivamente.

Se aconseja a futuros investigadores realizar los mejores modelos obtenidos en este trabajo de título, como Redes Neuronales Artificiales y Random Forest, pero en bases de datos más grandes para que los modelos puedan aprender de la data. Además, se aconseja agregar otro tipo de variables explicativas como el espacio ocupado por frutas y verduras en góndola.

A mi mamá, mi hermana y mi abuela.

Agradecimientos

Agradezco a mi mamá que lo ha dado todo por mí y porque siempre ha confiado en mis capacidades. Por su apoyo y cariño.

A mis amigos que supieron comprenderme y apoyarme en este período; por sus consejos y amistad. En especial a Cristobal por sus años de amistad y a Constanza por ser tan buena onda y por tener buena disposición a ayudar.

À Jean pour son soutien dans les moments les plus durs de ce mémoire et de la vie en génèrale, pour sa patience et sa bonne humeur. Je t'aime.

Tabla de Contenido

Índice de Tablas

Índice de Ilustraciones

Capítulo 1

Introducción

1.1. Antecedentes generales

Los bienes perecibles como huevos, leche, carne, mariscos, flores, pasteles, vegetales, frutas, etc., son una parte importante de la venta de los supermercados, representando un 36 %[1] del total de las ventas de este mismo. Sin embargo, según un estudio de la Cámara de Comercio de Chile CNC, entre diciembre del 2017 y diciembre del 2018 el índice real de ventas de los productos perecibles, exceptuando las carnes, bajó en un 1,3 % [7].

Los métodos para estimar la demanda de un bien varían según el producto o servicio ofrecido. La estimación más sencilla corresponde a la de productos no perecibles ya que poseen una amplia demanda en el tiempo y no pierden calidad en el corto plazo. Sin embargo, los productos perecibles como frutas y verduras tienen una demanda incierta dado a que estos productos son de corta vida útil y se deterioran fácilmente [12]. La intuición detrás de esto es que las personas consumen en menor cantidad productos que son estéticamente menos atractivos. Este nivel de estética en supermercados se debe, en su mayoría, al nivel de madurez de un producto; un ejemplo claro es el nivel de madurez del plátano, ya que una vez maduro se torna de color oscuro, poco deseable para el consumo. Es por ello que es de gran importancia, tanto para proveedores como para distribuidores, realizar un buen pronóstico de demanda para este tipo de productos, de tal manera que la cantidad de productos perecibles en un supermercado coincida con la demandada, reduciéndose así los quiebres o el sobre stock y los costos operacionales [24].

Según información de la Organización de las Naciones Unidas para la Agricultura y Alimentación [9], el 33 % de los alimentos producidos para el consumo humano se pierde o se desperdicia en el mundo, lo que representa unos 1.300 millones de toneladas de comida al año; 1.300 millones de toneladas en los cuales se invirtieron recursos naturales, espacio, vidas, tiempo, entre otros. Dentro de lo que concierne a un supermercado, las pérdidas provienen principalmente por malas condiciones de almacenamiento y la disminución de la calidad de los productos perecibles (por la estética de éstos o por su avanzado nivel de "madurez").

[1] En Apéndice A.1 se detalla el desgloce de los productos que están incluídos en este resultado.

En vista que los productos perecibles son delicados y tienen un tiempo de vida corto, una vez que se han producido, sus mercados se caracterizan por tener una demanda variable y por las marcadas fluctuaciones en los precios, como se aprecia en la Figura 1.1. Estas condiciones hacen que sea fácil equivocarse en la posible demanda de estos productos, sobreestimándolas en la mayor parte del tiempo.

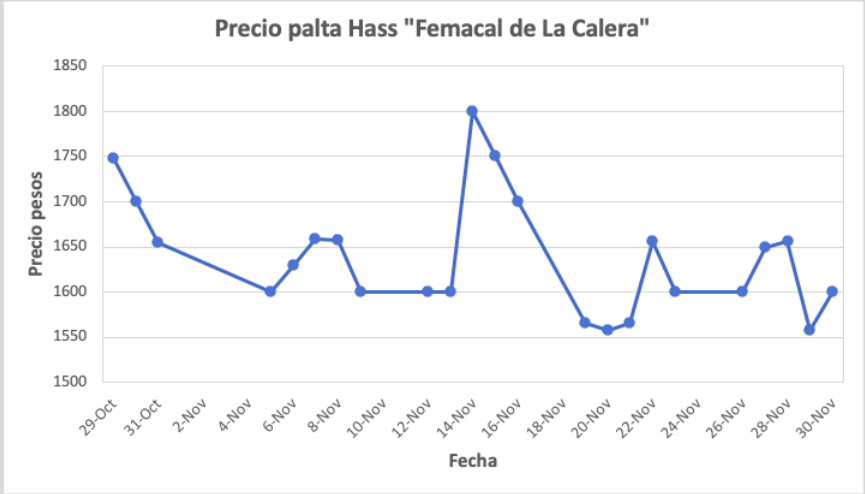


Figura 1.1: Variación del precio de la Palta Hass en Femacal de La Calera

Escasos son los estudios que se han realizado al respecto. Esto debido principalmente a las características particulares que poseen frutas y verduras, mencionadas anteriormente. Finalmente, estas son las razones por las cuales en este trabajo de investigación se intentará identificar cuál o cuáles son los mejores modelos para determinar la demanda de frutas y verduras de un supermercado.

1.2. Sector de supermercados en Chile

El primer supermercado en Chile se remonta al año 1957 con Almac, el cual fue además el primer supermercado a nivel latinoamericano. A partir de ese momento, la proliferación de supermercados en la región continuó de manera exponencial llegando a tener hoy en día 1384 a nivel nacional [8].

Como se aprecia en la Tabla 1.1, las ventas de los supermercados en los últimos 4 años han ido en alza, siendo el 2018 el año donde se tienen la mayor cantidad de ventas. Es entonces esperable que los índices de ventas a precios corrientes y a precios constantes[2] vayan también en aumento.

[2]La diferencia entre los índices de ventas a precios corrientes y constantes es que el primero se calcula con el valor de las ventas de cada día, o sea, incluye la inflación. El segundo se calcula en base a un año, que en este caso es el 2014.

Año	Ventas a precios corrientes [millones de pesos]	Índice de ventas a precios corrientes	Índice de ventas a precios constantes
2014	720.958	100	100
2015	781.436	108,39	101,79
2016	828.909	114,97	103,45
2017	865.964	120,11	105,62
2018	908.461	126,01	108,85

Tabla 1.1: Ventas e índices de supermercados en Chile. Índices a base promedio año 2014=100

1.3. Ventas de productos perecibles en supermercados

Como se observa en la Tabla 1.2, las ventas de productos perecibles como frutas y verduras en supermercados fluctúa año a año. Esto se debe principalmente a que este tipo de productos, como se señala anteriormente, tiene una corta vida útil y se deterioran fácilmente. Es imperativo, entonces, poder estimar la demanda de estos productos para no incurrir en quiebres de stock, o de lo contrario, de sobre stock.

Año	Ventas a precios corrientes [millones de pesos]	Índice de ventas a precios corrientes	Índice de ventas a precios constantes
2014	720.958	96,6	90,1
2015	781.436	94,2	88,9
2016	828.909	94,5	90,2
2017	865.964	95,2	92,8
2018	908.461	93,6	91,5

Tabla 1.2: Ventas e índices de productos perecibles, exceptuando las carnes, en supermercados de Chile. Índices a base promedio 2013=100

1.4. Justificación del tema

Actualmente, no existen estudios sobre la estimación de demanda de frutas y verduras en supermercados chilenos, generando graves problemas de stock y grándes pérdidas de dinero. Las principales causas son que los productos perecibles, a diferencia de los no perecibles, duran menos tiempo por lo que tienen que ser consumidos en un cierto intervalo de tiempo. Fuera de este intervalo de tiempo, los productos en muchas ocasiones son devueltos a sus productores o simplemente desechados.

Como se aprecia en la Tabla 1.2, la demanda anual de productos perecibles fluctúa a través de los años por lo que es difícil identificar cuál será la demanda para los períodos siguientes. Es importante entonces poder calcular esta demanda para poder evitar los problemas ya mencionados y, en este sentido, poder evitar las mermas en el mercado.

Una de las principales hipótesis de porqué ocurre esta fluctuación en la demanda es dada la particularidad que tienen los productos perecibles de poseer una corta vida útil; de esta

manera, la gente no los compra cuando éstos comienzan a deteriorarse o pierden el factor estético por el cual generalmente es comprado. Además, en vista de su rápido deterioro, los supermercados se ven en la obligación de bajar los precios de éstos para no perder dinero.

Una solución para esto sería calcular la demanda para no tener que incurrir en la baja de precios ni en la pérdida de estos mismos. De esta manera, con la demanda calculada, se puede ahorrar en almacenaje, se disminuirían las mermas y además, con ello, aumentarían las ganancias del supermercado.

1.5. Pregunta de investigación y objetivos

En este trabajo de investigación se busca responder a las siguientes preguntas: ¿Cuál será la demanda diaria de frutas y verduras dentro de un mes, sabiendo que éstas se deterioran fácilmente, su precio cambia constantemente y los clientes los compran según su apariencia estética? y, ¿Qué modelo de estimación se debería utilizar para estimar esta demanda con el menor error posible?

1.5.1. Objetivo General

Estimar la demanda de productos perecibles, frutas y verduras, de un supermercado para disminuir las mermas y con ello, aumentar los ingresos de este mismo.

1.5.2. Objetivos Específicos

Los objetivos específicos son los siguientes:

1. Identificar los datos o parámetros relevantes para los modelos. En este sentido, determinar qué variables pueden influir en la demanda.
2. Estimar la demanda con distintos modelos: Naïve Forecasting, Moving Average, Holt-Winter's Exponential Smoothing, ARIMA, Seasonal ARIMA, Regresiones Lineales Múltiples, Redes Neuronales Artificiales, Redes Neuronales Recurrentes, Support Vector Machine, Árboles Binarios y Random Forest.
3. Extrapolar los modelos a otra fruta o verdura de similares características para evaluar el desempeño de las distintas metodologías.
4. Encontrar el mejor modelo a través de indicadores como $MAPE$ y MAE[3].

[3]Ambos indicadores, tanto $MAPE$ como MAE se explican en el Capítuo 2, Marco Teórico.

1.6. Resultados esperados y alcance del proyecto

Se espera poder identificar cuál o qué métodos son los que entregan las mejores predecciones para la demanda de frutas y verduras dadas ciertas variables explicativas; con el fin de poder reducir las mermas que tiene un supermercado y poder optimizar los espacios utilizados en bodega, por ejemplo. Para ello, se realizan los modelos predictivos descritos en el Capítulo II de Marco Teórico y se comparan de tal manera de determinar cual es el más apropiado. De esta manera, se analiza si estos modelos entregan buenas predicciones y, si además, son lo suficientemente robustos como para ser replicados.

Este trabajo no busca implementar los resultados obtenidos, por lo que no se busca dar recomendaciones estratégicas a seguir para mejorar los niveles de stock.

1.7. Metodología de investigación

Como se trata de un proyecto con implementación en minería de datos, se utiliza el modelo CRISP-DM. CRISP-DM, del acrónimo Cross Industry Standard Process for Data Mining, es una metodología que sirve para extraer y analizar información desde grandes repositorios de datos. Es parte de la minería de datos que se basan en el proceso de Knowledge Discovery in Databases KDD.

CRISP-DM está relacionada profundamente con el enfoque de este proyecto pues éste obliga a revisar constantemente que el trabajo realizado vaya alineado con los objetivos del trabajo, y en caso de tener algún problema con los datos, por ejemplo, poder volver al paso anterior para corregir los errores. Actualmente es muy usado en proyectos de minería de datos y desarrollo de softwares, y se caracteriza por ser un conjunto de tareas que van de lo más general a lo más específico [2].

El ciclo de vida de un proyecto realizado con esta metodología consiste en seis etapas. La secuencia de etapas no es rígida, y alguna de ellas son bidireccionales, por lo que permite revisar parcial o totalmente las etapas como se muestra en la Figura 1.2 [6].

A continuación, se enumeran las principales actividades para cada una de las etapas de la metodología anteriormente planteada:

1. **Entendimiento del negocio**
   (a) Se determinan los objetivos del proyecto.
   (b) Se analizan las características de frutas y verduras para ver qué es lo que las hace "especiales".
2. **Entendimiento de los datos**
   (a) Se realiza una recolección inicial de los datos, con el objetivo de establecer un primer contacto con el problema; datos que pueden provenir tanto de la base de datos del supermercado a estudiar como de la cadena de supermercados y de internet.

Figura 1.2: Etapas del modelo CRISP-DM.
Fuente: Chapman & otros, 2000 [6]

(b) Se eligen las frutas y verduras que serán estudiadas en el proyecto; una fruta o verdura para correr los modelos y otra para verificar que estén bien planteados. Para ello, se analizan las frutas y verduras que más se venden en la cadena de supermercados y las que generan mayores ingresos (se realiza un trade-off entre estos dos aspectos).

(c) Se elige el supermercado del cual se extraen los datos y se aplican los modelos. Este supermercado se elige de manera de poder extrapolar los datos a otros supermercados de la misma cadena.

3. **Preparación de los datos**

   (a) Se realiza una exploración de los datos. Se identifican los datos clave para la realización de los modelos; para ello se realiza un análisis de canasta de compra con las frutas y verduras escogidas en el punto 2.(b).

   (b) Se hace una evaluación general de la calidad de los datos. Para ello se realiza una limpieza de estos mismos, identificando problemas de calidad y de completitud. Se eliminan datos correlacionados.

   (c) Se estructuran los datos. Se generan nuevos campos a partir de otros ya existentes para crear nuevos registros.

4. **Modelamiento**

   (a) Se seleccionan las técnicas de modelado más adecuadas para el proyecto. Para ello hay que considerar el objetivo principal que es estimar la demanda de productos perecibles como frutas y verduras. En este sentido, se eligen las técnicas que se adecúan a series de tiempo y regresiones. Además, se tratan los modelos como series de tiempo, o sea, el set de testeo y el de entrenamiento no se eligen de manera aleatoria como se suele hacer, sino que se utiliza un primer periodo como el set de entrenamiento, y un segundo (más corto que el primero) como el de testeo.

   (b) Se construyen todos los modelos.

5. **Evaluación**
   (a) Se evalúan los resultados de los modelos predictivos a través de los indicadores $MAPE$ y MAE, dándole prioridad a la hora de escoger un modelo a la métrica de $MAPE$.
   (b) Se realizan los mismos modelos, pero para la segunda fruta o verdura para evaluar el desempeño de estos.
   (c) Se revisa el proceso, para ello se califica el proceso entero de Data Mining, con el objetivo de identificar elementos que pudieran ser mejorados.
   (d) Se verifica que los mejores modelos para cada fruta o verdura estudiada sean robustos. Para ello, se replican los modelos en distintos sets de entrenamiento y de testeo pero de la misma base de datos[4].
6. Aplicación del mejor modelo
   (a) Esta fase no es implementada ya que para este trabajo de título no se trabaja directamente con el supermercado en cuestión.

[4]Se elige un set de entrenamiento y testeo más pequeño dentro de la misma base de datos de tal manera de demostrar que los mejores modelos son robustos.

Capítulo 2

Marco teórico

La demanda de productos perecibles es un tema que no ha sido muy estudiado en las últimas décadas. La mayoría de los estudios se enfocan principalmente en la demanda de productos no perecibles puesto que éstos cuentan con mercados establecidos, su calidad no sufre deteriorors en el corto ni mediano plazo (por lo que un producto que no se vende hoy puede venderse mañana) y porque las variaciones del precio al interior del mercado por lo general reflejan características ampliamente identificadas en el producto (tamaño, color, calidad, entre otros) y no se hace distinción acerca del origen del producto [9].

El punto de partida de este estudio se basa en que frutas y verduras tienen un corto período de vida, provocando que sus precios varíen considerablemente según su estado de maduración [21].

2.1. Métodos para la estimación de la demanda

En esta sección son descritos los métodos que son utilizados para la estimación de la demanda de productos perecibles. Uno de los principales problemas son el tipo de datos que utilizan estos modelos, los cuales la alumna no posee dentro de las bases de datos del supermercado. Dentro de los datos que se utilizan generalmente en estos modelos se encuentran, por ejemplo, la temperatura de los productos, el precio, su stock, su ubicación dentro del supermercado, entre otros. Sin embargo, como se ve en el Capítulo 3, se realiza un análisis de canasta para agregar variables a los modelos y poder suplir este problema.

Las técnicas de predicción de demanda de frutas y verduras que se utilizan en este trabajo incluyen:

- Naïve Forecasting
- Moving Average
- ARIMA
- Seasonal ARIMA

- Suavización Exponencial Triple de Holt-Winters
- Regresión Lineal Múltiple
- Árboles de Regresión
- Random Forest para Regresiones
- Support Vectos Machine
- Redes Neuronales Artificiales
- Redes Neuronales Recurrentes

En este trabajo se consideran los modelos de *Naïve Forecasting*, *Moving Average*, *ARIMA* y *Regresión Lineal Múltiple* como modelos *tradicionales* y el resto como más *avanzados*. De esta manera, se espera que los modelos más avanzados tengan un mejor rendimiento que los modelos tradicionales; esto porque los métodos avanzados incluyen la no linealidad de los datos y porque se espera tener un nivel significante de demanda no lineal. En lo que resta de esta sección, se explicarán brevemente estos modelos.

La notación principal a utilizar es la que se muestra a continuación[5]:

- $\hat{y}_{t+n}$: demanda a predecir en tiempo $t+n$
- y_t: demanda en el período t

2.1.1. Naïve Forecasting

Este es el método más básico que se utiliza para predecir. La premisa de este método es que el punto esperado es igual al último punto observado:

$$\hat{y}_{t+1} = y_t \tag{2.1}$$

También se puede asumir que los k puntos esperados son iguales a los k puntos anteriores.

Aunque este método luzca simple o ingenuo es útil para crear un punto de partida en el análisis. Numerosos estudios de predicción lo utilizan cuando los datos no poseen una considerable diferencia entre los días, y algunos demuestran que *Naïve Forecasting* es mejor que otros métodos como *Moving Average* o *Trend*, cuando no se ve mucha variación en los datos [5].

2.1.2. Moving Average

Dada una secuencia $\{a_i\}_{i=1}^{N}$, una n-media móvil o *n-moving average* se define como una nueva secuencia $\{s_i\}_{i=1}^{N-n+1}$ la cual proviene de la media aritmética de n elementos de la secuencia a_i. Moving average es una técnica para tener una idea de la tendencia del set

[5]El resto de la notación se encuentra en Apéndice B.1.

de datos. Esta metodología es extremadamente útil para predecir tendencias a lo largo del tiempo y además, sirve para tener un primer acercamiento con los datos.

2.1.3. ARIMA

En estudios relativamente recientes de demanda de alimentos perecibles se utiliza el modelo ARIMA [13]. El modelo autorregresivo integrado de promedio móvil o ARIMA es un modelo que estudia predicciones de series de tiempo, donde estas series de tiempo pueden considerarse como la realización de un proceso estocástico que se observa secuencialmente a lo largo del tiempo. El modelo ARIMA es un caso particular del modelo ARMA en el cual sí existe una raíz unitaria. EL modelo ARMA es a su vez una combinación del proceso autorregresivo $AR(p)$ y el proceso de media móvil MA(q). Ambos procesos son procesos de series de tiempo que intentan explicar la demanda a partir de datos pasados. La diferencia es que el primero tiene memoria a largo plazo por lo que le cuesta reaccionar rápidamente ante "shocks" o perturbaciones y el segundo, tiene corta memoria, reaccionando ágilmente a perturbaciones, pero "olvidando" la información del pasado [4].

Este modelo es un buen modelo utilizado en estadísticas, econometría e ingeniería por varias razones: (i) es considerado como uno de los modelos con mejor desempeño en términos de pronóstico, (ii) se utilizan como referencia para modelos más sofisticados y (iii) porque son de fácil implementación y alta flexibilidad dada su estructura multiplicativa [10].

Los parámetros de un modelo $ARIMA(p, \mathrm{d}, q)$ se definen como sigue:

- p es el número de términos autorregresivos;
- d es el número de diferencias que se aplican a la serie de tiempo para que sea estacionaria; y
- q es el número de medias móviles o *moving average* que realiza el proceso.

De esta forma, se construye un modelo de regresión lineal que incluye el número y el tipo de términos especificados, de tal manera que la serie de tiempo sea estacionaria. Es necesario que la serie de tiempo sea estacionaria para eliminar tendencias y estructuras estacionales que pueden afectar negativamente el modelo de regresión. Finalmente, el modelo de regresión lineal que se busca tiene la siguiente forma:

$$\hat{y}_t = \delta + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \tag{2.2}$$

Con:

- δ una constante;
- y_{t-1}, y_{t-p} las ventas en el período $t-1$ y $t-p$;
- ε_{t-1}, ε_{t-p} los residuos de los períodos $t-1$ y $t-p$, los cuales constituyen el ruido blanco; y
- ϕ, θ los coeficientes de los procesos autorregresivos y de media móvil, respectivamente.

2.1.4. Seasonal ARIMA

Cuando se tienen efectos de temporalidad, es necesario hacer uso del Seasonal ARIMA o ARIMA temporal para incluir el efecto de la temporalidad de los datos. En este caso, el modelo ARIMA se denota como $ARIMA(p, \mathrm{d}, q)(P, D, Q)s$. Aquí, (p, d, q) son los parámetros no-temporales descritos anteriormente, mientras que (P, D, Q) siguen la misma intuición, pero para la componente temporal de la serie de tiempo. El término s es la periodicidad de la serie.

En la literatura, este modelo ha sido utilizado para estudiar productos no perecibles de una cadena de supermercados en Chile [1], sin embargo, escasos son los estudios donde se aplica a productos perecibles.

2.1.5. Suavización Exponencial Triple de Holt-Winters

El método de Suavización Exponencial Triple de Holt-Winters se utiliza para pronosticar series de tiempo y su nombre triple viene ya que se utiliza para pronosticar tendencia, temporalidad y estacionalidad. De esta manera, esta metodología utiliza tres ecuaciones para *suavizar*; una para la atenuación de la serie de tiempo, otra para la tendencia y una última para la estacionalidad. La ecuación de atenuación de la serie de tiempo o de pronóstico para el período t se calcula de la siguiente forma:

$$A_t = \alpha \left(\frac{y_t}{R_{t-L}} \right) + (1 - \alpha)(A_{t-1} + T_{t-1}) \tag{2.3}$$

Donde α corresponde a la constante de atenuación la cual toma valores en el intervalo $0 < \alpha < 1$, T_{t-1} corresponde a la tendencia del período $t - 1$ y R_{t-L} a la estacionalidad del período $t - L$. L se considera como el largo del ciclo de estacionalidad[6]. La tendencia del período t se modela como sigue:

$$T_t = \beta(A_t + A_{t-1}) + (1 - \beta)T_{t-1} \tag{2.4}$$

Donde β es el coeficiente de tendencia el cual toma valores entre el intervalo $0 < \beta < 1$. La estacionalidad del período t se formula a continuación:

$$R_t = \gamma \left(\frac{y_t}{A_t} \right) + (1 - \gamma)R_{t-L} \tag{2.5}$$

Donde el parámetro γ se refiere al coeficiente de estacionalidad el cual, al igual que los coeficientes anteriores, se encuentra en el intervalo $0 < \gamma < 1$. Finalmente, la predicción para k períodos en el futuro dado el período t es:

[6]Si se tiene por ejemplo un ciclo anual, se puede tener una estacionalidad mensual $L = 12$, cuaternaria $L = 4$; si se poseen ciclos diarios se puede tener una estacionalidad horaria $L = 24$, etc.

$$\hat{y}_{t+k} = (A_t + kT_t)R_{t-L+k} \tag{2.6}$$

Se espera que con este modelo se pueda detectar tanto la tendencia como la estacionalidad y la temporalidad de los datos.

2.1.6. Regresión Lineal Múltiple

Regresión Lineal Múltiple es uno de los métodos más utilizados dada su fácil interpretabilidad. En estudios relacionados a la predicción de demanda de productos perecibles como frutas y verduras, la demanda se calcula en función del precio p, la calidad del alimento perecible $q(t)$ y la cantidad de estanterías destinadas para el producto n_g [23]. De esta manera, la función de demanda puede ser descrita como sigue:

$$y(t) = -cp + \mathrm{e}n_g + \mathrm{d}q(t) \tag{2.7}$$

Donde c es la elasticidad del precio de los alimentos perecibles, e es la sensibilidad de demanda según la cantidad de espacio designado al producto (cantidad de espacio en góndolas, por ejemplo) y d, la sensibilidad de demanda con respecto a la calidad del producto.

Si bien estos son los parámetros más utilizados, se pueden utilizar otros (según la base de datos que se posea). Una regresión lineal múltiple cualquiera que relaciona una variable dependiente con variables independientes se escribe de la siguiente manera:

$$y_{\mathrm{i}} = a_1 x_{\mathrm{i}1} + a_2 x_{\mathrm{i}2} + \cdots + a_n x_{\mathrm{i}n} + \varepsilon_{\mathrm{i}}, \qquad \mathrm{i} \in \{1, \ldots, n\} \tag{2.8}$$

El objetivo de este modelo es poder estimar los parámetros a_{i} utilizando los datos de la muestra. Las variables x_{i} son las variables explicativas de y, y los ε_{i} es lo que no se observa del modelo, el cual distribuye típicamente como $N(0, \sigma)$.

2.1.7. Árboles de Decisión

Los árboles binarios o árboles de decisión son un tipo de algoritmo de aprendizaje supervisado en la cual existe una variable objetivo predefinida. Las variables de entrada y salida pueden ser categóricas o continuas y este tipo de algoritmos divide el espacio de los predictores (las variables independientes) en regiones distintas y no sobrepuestas [18]. En la Figura 2.1 se puede ver un ejemplo de árbol de decisión el cual posee seis regiones separadas.

Una de las características de los árboles binarios es que éstos siguen un enfoque de división binaria recursiva, o más conocido como *top-down greedy approach*. En ésta, se analiza la mejor variable para la ramificación en el proceso de división que se esté realizando.

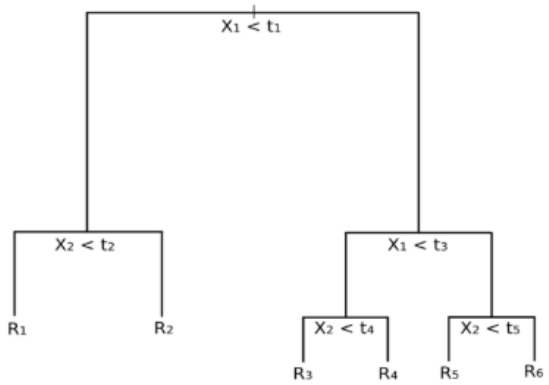


Figura 2.1: Ejemplo de Árbol Binario con seis regiones separadas.
Fuente: Orellana A., 2018 [18]

Este es uno de los métodos menos utilizados en la predicción de demanda de frutas y verduras y se debe principalmente a que se necesitan una gran cantidad de variables para que los resultados sean robustos [22].

Una de las grandes ventajas de los árboles de decisión es que son fáciles de entender y explicar ya que se sabe cuáles de las variables son las que afectan en el resultado. En este sentido, se puede identificar la importancia de las variables del modelo. Sin embargo, este es un método que lleva generalmente al sobreajuste.

2.1.8. Random Forest

Así como todos los modelos descritos anteriormente, un árbol binario también tiene problemas de sesgo y de varianza. Esto se intenta disminuir con la metodología de Random Forest en la cual se utiliza la técnica de *bagging* para reducir la varianza de las predicciones. Esta técnica lo que hace es generar subconjuntos de árboles dentro del set de entrenamiento para que la correlación de las variables, si es que existe, no afecte en los resultados, reduciendo la varianza. En la Figura 2.2 se ilustra el proceso de bagging.

En otras palabras, Random Forest es una técnica que utiliza múltiples árboles de decisión y la metodología de bagging, la cual consiste en entrenar cada árbol de decisión en una muestra de datos distinta donde se realiza el muestreo con reemplazo, para reducir la varianza en la predicción. La idea básica que está detrás de este modelo, es de combinar los resultados que arrojan una gran cantidad de árboles de decisiones en vez de dejarle la responsabilidad a uno solo.

Este modelo no se ha aplicado en la literatura para productos perecibles. Queda como un modelo que la alumna propone para resolver esta problemática.

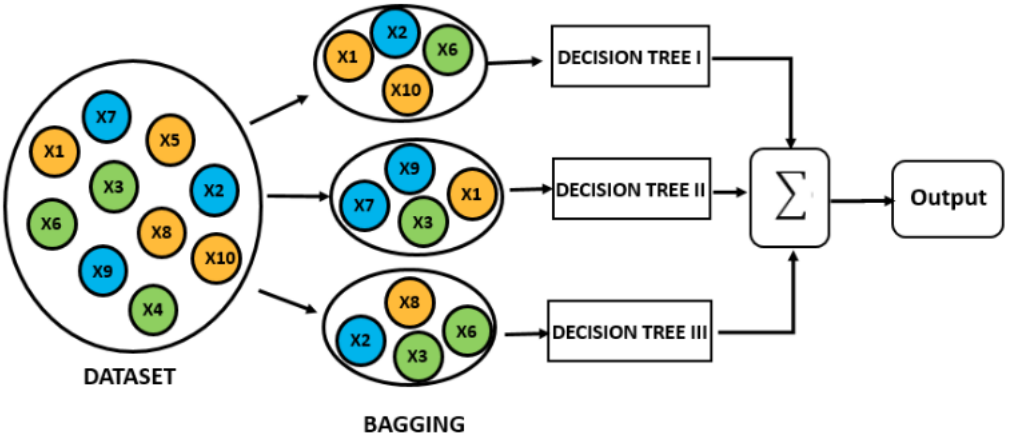


Figura 2.2: Ilustración del proceso de bagging.
Fuente: Orellana A., 2018 [18]

2.1.9. Support Vector Machine

Una máquina de vectores de soporte, SVM por su acrónimo en inglés, es un clasificador definido por un hiperplano separador. En otras palabras, dados los datos de entrenamiento, el algoritmo genera un hiperplano óptimo que categoriza nuevos ejemplos. Como los SVM tienen un mayor rendimiento de generalización y garantizan mínimos globales para datos de entrenamiento, se sabe que una regresión de vectores de soporte funciona bien para pronosticar la demanda de productos agrícolas perecederos [12].

Support Vector Machine puede ser aplicado tanto a problemas de clasificación como problemas de regresión; en este caso se le llama Support Vector Regressor, SVR. Este modelo tiene un parámetro de ajuste llamado *kernel* el cual es una función utilizada para mapear datos desde una baja dimensión a una alta dimensión.

La diferencia de este método con regresiones es que en regresiones se intenta minimizar la tasa de error y en SVR se intenta ajustar el error dentro de un cierto umbral [3].

Si se considera un modelo de regresión de la forma:

$$\hat{y} = \mathbf{w}^T \mathbf{x} + b \tag{2.9}$$

Donde $\mathbf{x}$ son los input o variables independientes, b y $\mathbf{w}$ son los parámetros encontrados con la data y $\hat{y}$ la predicción, lo que se busca es ajustar el error dentro de un umbral, pero sin sobre-ajustarlo. Visualmente, queda como se muestra en la Figura 2.3.

En la Figura 2.3, ξ_i y ξ_i^* representan la distancia que existe entre los puntos que se encuentran fuera del margen con el margen. Finalmente, lo que se busca en SVR [16] es

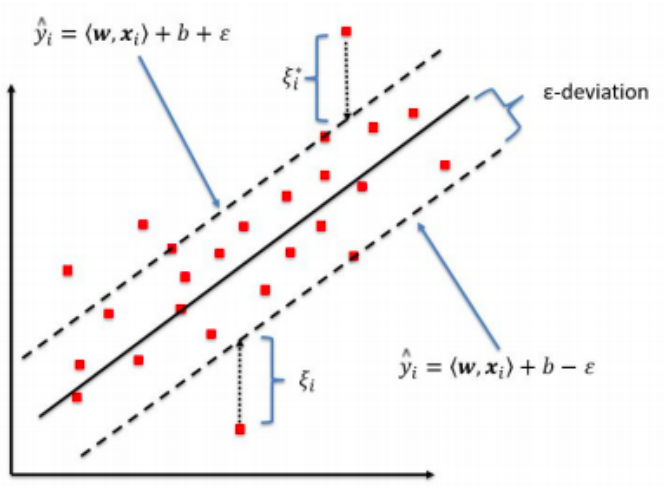


Figura 2.3: Ilustración de Support Vector Regressor en dos dimensiones.
Fuente: Kleynhans & otros [16]

poder resolver la siguiente ecuación:

$$\begin{aligned} &minimizar \ \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{\mathrm{i}=1}^{n}(\xi_{\mathrm{i}}^* + \xi_{\mathrm{i}}) \\ &s.a \begin{cases} \hat{y}_{\mathrm{i}} - \langle \mathbf{w}, \mathbf{x_i} \rangle - b \leq \varepsilon + \xi_{\mathrm{i}}^* \\ \langle \mathbf{w}, \mathbf{x_i} \rangle + b - \hat{y}_{\mathrm{i}} \leq \varepsilon + \xi_{\mathrm{i}} \end{cases} \end{aligned} \tag{2.10}$$

Donde C es un parámetro que penaliza las observaciones fuera de los márgenes y cuya finalidad es evitar el sobre-ajustamiento.

De esta manera, sólo los puntos fuera de los márgenes son utilizados para la realización de predicciones [12].

2.1.10. Redes Neuronales Artificiales

Tanto las redes neuronales como las redes neuronales recurrentes son frecuentemente utilizadas para predecir series de tiempo [11]. Existe una basta cantidad de redes neuronales; en este informe se hace referencia sólo a aquellas que tienen una propagación hacia atrás, las que son conocidas como *feed-forward error back-propagation neural nets*. En estas redes, los elementos individuales, las neuronas de la red, están organizadas en capas de tal manera que las señales de salida de las neuronas de una capa se transmiten a todas las neuronas de la capa siguiente. En este sentido, el flujo de activación de neuronas va en un solo sentido y pasa capa por capa. El número mínimo de capas que se puede tener son dos capas, la de entrada y la de salida, sin embargo, se pueden agregar capas entremedio llamadas capas ocultas las cuales sirven para aumentar el poder computacional de las redes neuronales. En la Figura 2.4 se muestra una representación de este tipo de redes neuronales. De esta manera, a las Redes Neuronales Artificiales hay que entregarle la cantidad de capas de entradas, de

capas ocultas, de neuronas (puntos azules de la Figura 2.4) y de capas de salidas. Además, se le debe entregar el número de *epochs*, que es la cantidad de veces que el aprendizaje ocurre. Así, el proceso de aprendizaje de una red neuronal se repite epoch tras epoch hasta que el rendimiento de la red converge a un valor aceptable[7].

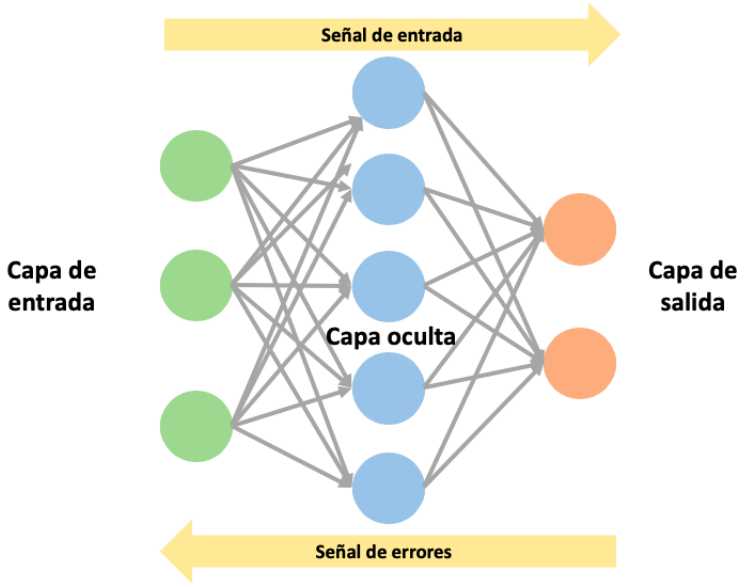


Figura 2.4: Representación de una red neuronal con propagación hacia atrás

Las redes neuronales artificiales están hechas para cumplir con un mapeo requerido utilizando algoritmos de entrenamiento. El algoritmo de entrenamiento común para las redes *feed-forward* se denomina propagación por error [20].

Este tipo de algoritmos es uno de los más utilizados cuando se tiene una gran cantidad de datos ya que tiene un alto poder predictivo y puede ser fácilmente automatizado; además de tener la habilidad de manejar patrones complejos no lineales [15]. Sin embargo, uno de sus principales puntos en contra es que tiene una muy baja interpretabilidad.

2.1.11. Redes Neuronales Recurrentes

Las redes neuronales recurrentes se incluyen en este análisis ya que se considera que la demanda de productos perecibles como frutas y verduras es una serie caótica, la cual no posee, en ocasiones, una gran lógica. Las redes neuronales recurrentes lo que hacen es generar un *back-propagation* o una propagación hacia atrás que permite obtener mejores patrones de aprendizaje a través del tiempo. Esto quiere decir que una red neuronal recurrente puede hacer coincidir cierto patrón (algún patrón de los datos) a través del tiempo el cual se extiende más allá de la ventana de tiempo actual proporcionada [5]. En la Figura 2.5 se muestra una representación de este tipo de redes neuronales.

[7]Un valor muy pequeño de epoch puede llevar a un aprendizaje deficiente por la parte de la red neuronal. Un valor sumamente elevado puede llevar al caso contrario, a un sobreajuste de la red neuronal. Cabe destacar que este comportamiento de bajo aprendizaje o de sobreajuste también puede ocurrir con la cantidad de capas ocultas y la cantidad de neuronas que se le entregan a esta misma.

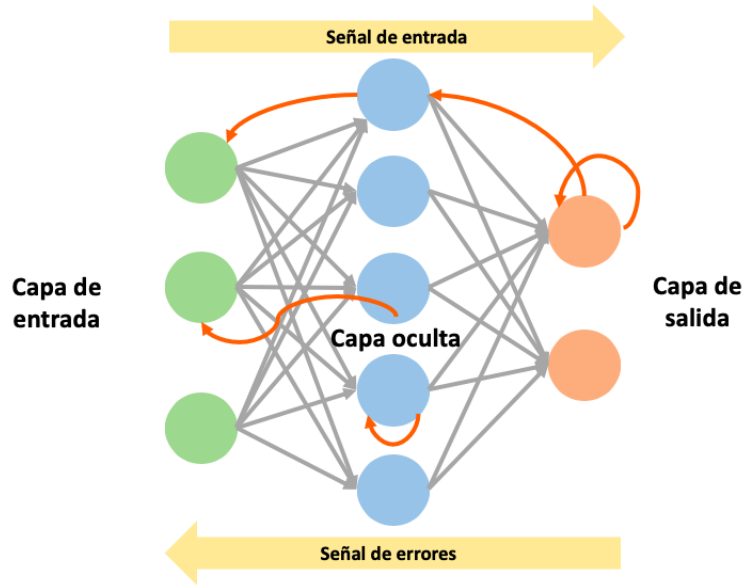


Figura 2.5: Representación de una red neuronal recurrente con propagación hacia atrás

2.2. Indicadores de errores de predicción

Los estadísticos que se utilizan para comparar los modelos anteriormente descritos son MAPE y MAE. Ambos se utilizan para medir distintas formas de errores de predicciones.

2.2.1. MAPE

El Error Porcentual Absoluto Medio o MAPE, por su acrónimo en inglés, es un indicador del desempeño del pronóstico de la demanda el cual mide las variaciones porcentuales que existen entre la demanda real y la demanda pronosticada. Se calcula como sigue:

$$MAPE = \frac{\sum_{t=1}^{n} \frac{|y_t - \hat{y}_t|}{|y_t|}}{n} \tag{2.11}$$

Donde y_t es la demanda en t, $\hat{y}_t$ es la demanda pronosticada en t y n es la cantidad de datos pronosticados, en este caso, la cantidad de días.

2.2.2. MAE

El Error Absoluto Medio o MAE, por su acrónimo en inglés, es una métrica que mide la magnitud de los errores en un set de predicciones, sin considerar la dirección del error (error mayor o menor que el valor real).

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t| \tag{2.12}$$

Se decide utilizar el indicador MAE como segundo indicador en la toma de decisiones. Esto porque el indicador MAPE favorece en el error cuando las predicciones son menores al valor real y castiga cuando las predicciones son mayores. De esta manera, se utiliza MAE cuando la decisión de qué método es mejor se torna complicada.

En este trabajo de título el indicador MAE se interpreta como la diferencia promedio de kilos de palta o tomate, dependiendo del caso, entre las ventas reales y las pronosticadas.

2.2.3. Promedio

Con el fin de poder estudiar la robustez de los modelos, se utiliza el promedio de ambos estadísticos señalados anteriormente, MAPE y MAE. El promedio se calcula como sigue:

$$\bar{x} = \frac{1}{n}\sum_{\mathrm{i}=1}^{n} x_{\mathrm{i}} \tag{2.13}$$

Donde x_{i} representa al estadístico MAPE o MAE, dependiendo del caso.

2.2.4. Desviación estándar

La desviación estándar se utiliza para medir la dispersión de los datos con respecto al promedio y, en este trabajo, se utiliza como ayuda en el análisis de robustez. La desviación estándar se calcula como sigue:

$$s = \sqrt{\frac{1}{n-1}\sum_{\mathrm{i}=1}^{n}(x_{\mathrm{i}} - \bar{x})^2} \tag{2.14}$$

Al igual que en el caso anterior, x_{i} representa al estadístico MAPE o MAE. Por otra parte, $\bar{x}$ representa a la media aritmética y se calcula como se observa en la ecuación 2.13.

En ambas ecuaciones anteriores, tango la ecuación 2.13 como la ecuación 2.14, n representa la cantidad de MAPEs o MAEs que se tengan. En este trabajo se utilizan seis bases de datos por lo que $n = 6$. La conformación de estas bases de datos se explican en el Capítulo 4, Aplicación de la metodología.

Capítulo 3

Descripción de los datos

Los datos que serán utilizados en este trabajo provienen principalmente de las ventas diarias de frutas y verduras de un gran supermercado del retail que, por temas de confidencialidad, quedará en estado de anonimato. Los datos provienen de 21 bases de datos las cuales poseen todas las ventas de la cadena de supermercado que van desde octubre del 2014 a junio del 2016 (630 filas de datos), de manera diaria. Dentro de los datos que se pueden utilizar, en la base de datos se encuentran solamente las cantidades de las transacciones, las ventas diarias y el stock[8] de los productos (para ciertos meses). De esta manera, se obtiene el precio del kilo de las frutas y verduras como la división entre el total de la venta diaria y la cantidad total vendida de la fruta y/o verdura escogida. En la Figura 3.1 se aprecia visualmente cómo son estas bases de datos.

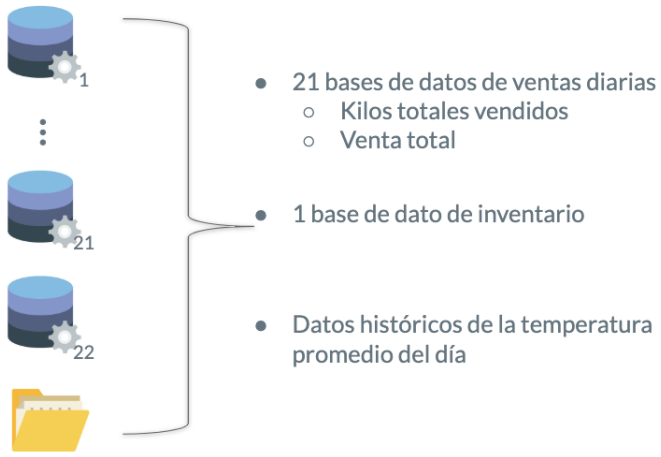


Figura 3.1: Diagrama de las bases de datos

Se realiza un análisis de canasta de compras para para ver qué datos de venta de otros productos se pueden agregar como variables explicativas para la fruta o verdura seleccionada. Se agrega además la variable exógena temperatura promedio del día a la base de datos para ver si esta afecta en las compras. Asimismo, se agregan cinco variables para ver si hubo un

[8]El stock proviene de otra base de datos que proviene, al igual que las otras bases de datos, a la cadena de supermercados analizada.

aumento o un descuento en el precio con respecto al día anterior y ver si esta afecta en el comportamiento de los clientes. Estas variables fueron clasificadas como un aumento o descuento de un 0 a 10 %, 10 % a 20 % o más de un 20 %. Aquí se agregan sólo cinco variables para no incurrir en la "trampa de variables dummy" o más conocido como *Dummy Variable Trap*[9]. De esta manera, la única variable que no se incluye es la variable de aumento de precio por más del 20 %.

Se agregan además variables binarias que describen qué tipo de día es, si se trata de un fin de semana o no o si es el primer fin de semana del mes; esto para ver si hay más compras en ese fin de semana dado el sueldo que reciben los chilenos por su trabajo. Cabe destacar que los días feriados no renunciables fueron considerados como parte del fin de semana. Otras variables que se agregan de manera binaria son las de los días de semana (seis variables, una para cada día de la semana, excepto el día lunes, para evitar la trampa de variables dummy).

3.1. Análisis de la base de datos

A continuación, se muestra un análisis de las ventas de frutas y verduras de manera general con el fin de escoger cuáles serán las frutas o verduras analizadas para el estudio. Se realiza el mismo procedimiento para la elección del supermercado a analizar.

3.1.1. Análisis de las frutas y verduras

Se analizan las ventas de frutas y verduras de toda la cadena de 48 supermercados de la cadena con respecto a los kilos vendidos y los ingresos que generan estos mismos, como se muestra en la Tabla 3.1. A partir de esta tabla, se decide que el mejor producto a analizar son las paltas dado a que es la fruta que más ingresos genera y dado a que es la segunda fruta más vendida, luego del tomate. Es dentro del interés de este trabajo evaluar estos modelos en otra verdura para evaluar el desempeño de estos. Es por ello que la segunda fruta elegida es el tomate. Se elige el tomate ya que, al igual que la palta, es una fruta que representa características similares: se deteriora fácilmente, la gente lo consume regularmente y representa el tercer lugar dentro de las ventas de frutas y verduras.

Dentro de la categoría paltas, existe un total de 22 tipos de paltas que se venden durante todo el año en estos supermercados. Se elige aquella que tiene el mayor impacto en los ingresos de la tienda, la Palta Hass Extra a Granel como se muestra en la Tabla 3.2.

De manera análoga, dentro de la categoría tomates, existe un total de 32 variedades distintas que se venden durante todo el año. Se elige entonces el que genera el mayor ingreso para el supermercado, el Tomate Granel, como muestra la Tabla 3.3.

[9]La Dummy Variable Trap es un escenario en el que las variables independientes son multicolineares, un escenario en el que dos o más variables están altamente correlacionadas. En otras palabras, una variable puede ser predicha a partir de las otras.

Fruta o verdura	Kilos vendidos [millones de Kg]	Ingresos generados [millones de pesos]
Paltas	2,67	7.787
Papas	1,69	3.031
Tomates	2,83	2.922
Limones	2,08	2.162
Champiñones	2,03	2.134
Lechuga Escarola	1,87	1.911
Bananas	2,37	1.856
Zanahorias	1,41	1.521
Frutillas	0,66	1.447
Apio	0,89	1.339
Cebollas	1,51	1.295

Tabla 3.1: Resumen análisis de frutas y verduras de toda la cadena de supermercados (21 meses)

Tipo de palta	Kilos vendidos [millones de Kg]	Ingresos generados [millones de pesos]
Palta Hass Extra a Granel	1,97	6.066
Palta Hass Malla 1 Kg	0,56	1.403
Palta Hass Orgánica Malla 900 grs	0,42	106
Palta Fuerte a Granel	0,33	80
Palta Edranol a Granel	0,28	65
Palta Negra de la Cruz Malla 1 Kg	0,24	58

Tabla 3.2: Resumen de ventas de los primeros 6 tipos de paltas (21 meses)

Tipo de tomate	Kilos vendidos [millones de Kg]	Ingresos generados [millones de pesos]
Tomate a Granel	2,83	2.825
Tomate Beef a Granel	0,40	569
Tomate Romanita Pote 340 grs	0,26	554
Tomate Racimo a Granel	0,23	350
Tomate Salad a Granel	0,32	321
Tomate Malla 1 Kg	0,24	213

Tabla 3.3: Resumen de ventas de los primeros 6 tipos de tomates (21 meses)

3.1.2. Análisis de supermercados

Se analizan las ventas totales por local en un total de 48 locales (48 supermercados de la misma cadena). Dentro de estos, se analiza el total de ventas de cada uno en un período de 18 meses para ver el índice de ventas de cada uno como se muestra en el Apéndice A.2. Se elige entonces el supermercado cuyas ventas son cercanas al promedio de ventas general, en este caso el supermercado número 18. De esta manera, al elegir un supermercado "promedio", se espera que en un futuro se puedan extrapolar los resultados al resto de los supermercados (extrapolar los resultados a otro supermercado queda fuera del alcance de este trabajo de memoria, como mencionado en el Capítulo 1).

3.1.3. Análisis de la Palta Hass Extra a Granel

Se analiza la serie de tiempo de las ventas en kilo de la palta a estudiar como se muestra en la siguiente Figura:

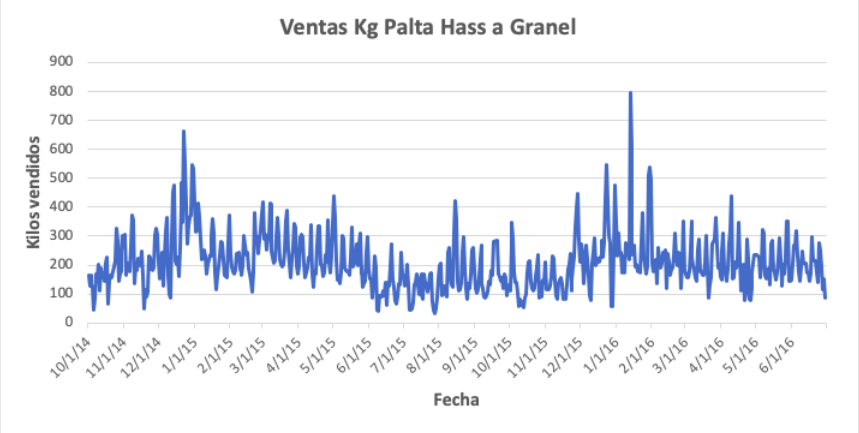


Figura 3.2: Ventas en Kg de Palta Hass Extra a Granel

Esta serie de tiempo posee las siguientes características:

Tendencias	Valor
Mínimo	33,48
Máximo	797,26
Desviación Estándar	96,18
Media	192,80
Promedio	208,02

Tabla 3.4: Características de la serie de tiempo de la Palta Hass

Como se aprecia en la Figura 3.2, a simple vista pareciera tener dos *outliers*, en diciembre de 2014 y enero de 2016. Ahondando más en estos puntos, se puede decir que esta alza en las compras se debe a una baja en el precio como se aprecia en la Figura 3.3. Como se observa en esta Figura, una de las razones del aumento abrupto de las ventas en el día 12 de enero es dado a la baja en el precio. Una de las hipótesis de esta descenso en el precio por más de mil pesos es que hubo una disminución en la calidad de las paltas y el supermercado disminuyó los precios en un esfuerzo de evitar las mermas.

Se realiza además un análisis de ventas por mes (el mismo mes, pero en distintos años) para ver si las ventas se comportan de una manera regular.

Como se aprecia en la Figura 3.4, no existe una tendencia clara en el comportamiento de las ventas de paltas durante los años 2014 y 2015.

Analizando las ventas por día de la semana y por mes, como se ve en la Figura 3.5, se aprecia que existe un aumento en las compras los fines de semana para el mes de octubre[10]

[10]Este mes se muestra como un ejemplo. Ocurre un comportamiento parecido para los otros meses del año.

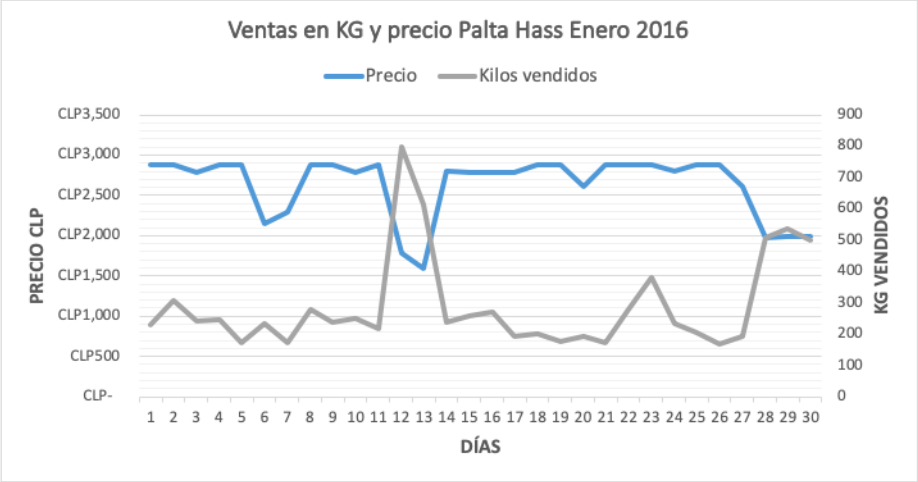


Figura 3.3: Ventas en Kg y precio de la Palta Hass Extra a Granel. Enero 2016

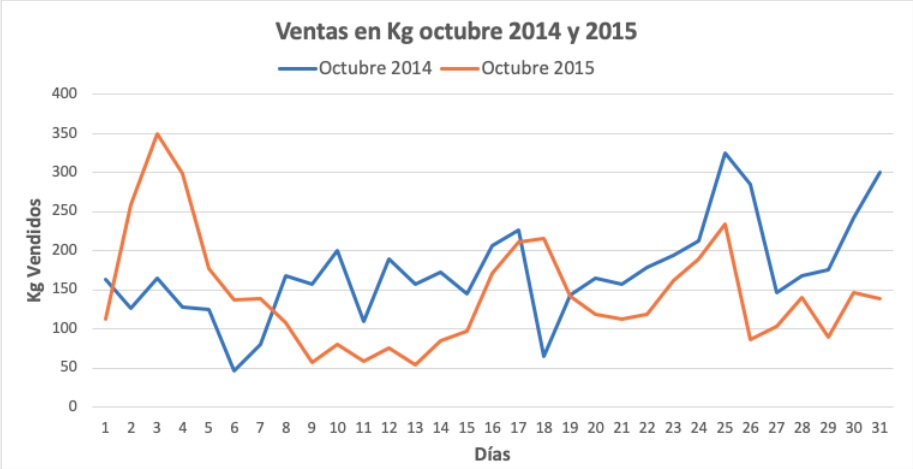


Figura 3.4: Ventas en Kg de octubre 2014 y 2015

para el año 2015, pero no para el año 2014, donde las ventas son mayores los días de semana. Se espera, en ese sentido, que la variable día de semana sea una variable que pueda explicar cierto comportamiento de compra de los clientes.

3.1.4. Análisis del Tomate a Granel

De manera análoga al análisis de la palta, se analiza la serie de tiempo de las ventas en kilo del tomate como se muestra en la Figura 3.6.

Esta serie de tiempo posee las características que se muestran en la Tabla 3.5. En vista de estas características se espera que los resultados de las predicciones den acorde a estos números; es decir, que por ejemplo, la métrica MAE sea más grande en los tomates que en las paltas pues se venden en promedio más tomates que paltas, por lo que un error en un 30 %, por ejemplo, en los tomates significa un error más grande en kilos.

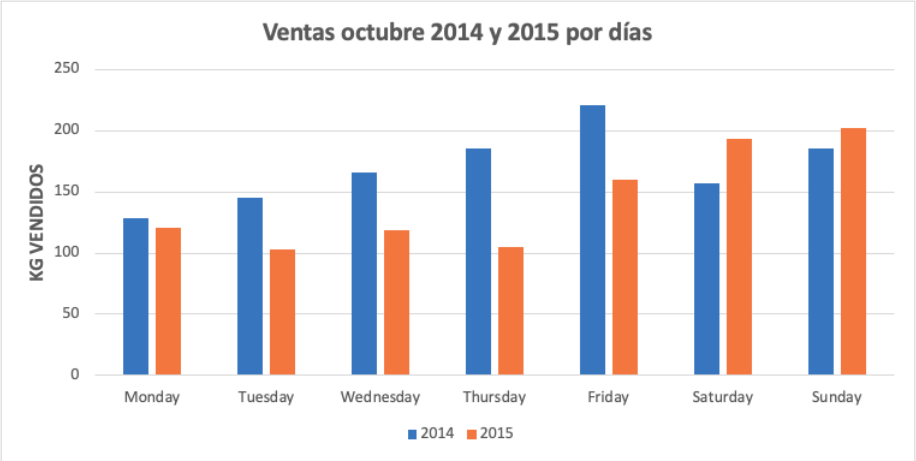


Figura 3.5: Ventas promedio por días de la semana. Octubre 2014 y 2015

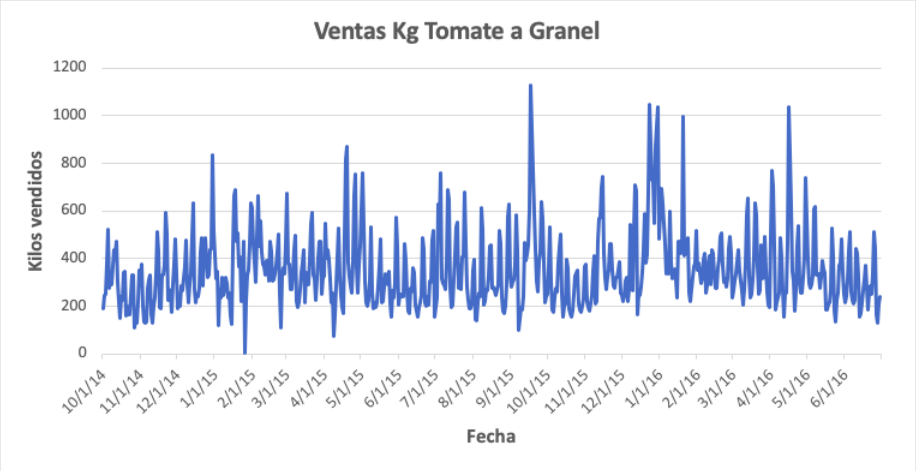


Figura 3.6: Ventas en Kg de Tomate a Granel

Se analiza de forma análoga a la palta las ventas que se tienen de los tomates de manera mensual. Los resultados son los que se muestran en la Figura 3.7.

Al igual que en la Figura 3.4, se puede decir que las ventas de los tomates no poseen una tendencia clara. Esto se ve principalmente en los primeros días de octubre, donde se aprecia de mejor manera la diferencia de las curvas.

Se prosigue a realizar un análisis de las ventas promedio por día como se muestra en la Figura 3.8. Como se muestra en esta Figura, en este caso las ventas se concentran mayoriatiamente en el fin de semana. Se espera, en este sentido, que el tipo de día afecte en las compras; es por ello que se decide agregar variables binarias de los días de la semana como se detalla en el Capítulo 4.

Tendencias	Valor
Mínimo	5,21
Máximo	1126,33
Desviación Estándar	157,15
Media	321,56
Promedio	353,88

Tabla 3.5: Características de la serie de tiempo del Tomate

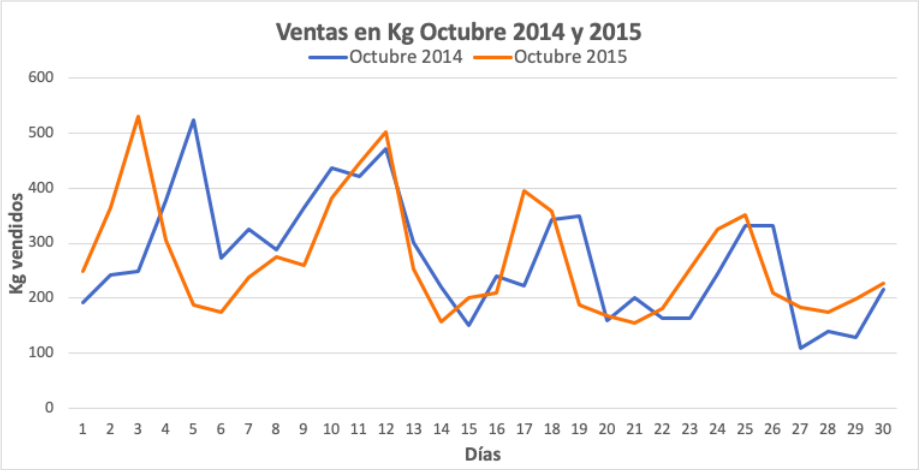


Figura 3.7: Ventas de Tomates en Kg de octubre

3.1.5. Análisis de Canasta

Se realiza un análisis de canasta para ver cuales son los productos que se compran junto con la Palta Hass Extra a Granel y junto al Tomate a Grane para ver cómo afectan la compra de estos productos, los de la canasta de compras, a la compra de la palta o el tomate, respectivamente. Para ello se analizan siete días de compras los cuales fueron elegidos de manera aleatoria tal que hubiese un día de cada uno; un lunes, un martes, y así hasta el domingo. Este análisis se realiza de manera aleatoria para no incurrir en errores y no considerar algún producto que haya estado en oferta en el periodo seleccionado.

Análisis de Canasta Palta Hass

Se analizan un total de 1223 boletas de compras de Palta Hass Extra a Granel y se estudian los productos que se compran en conjunto. Estas 1223 boletas corresponden a siete días de compras, de lunes a domingo, elegidos al azar dentro de la base de datos.

Como se observa en la Tabla 3.6, el producto que más se compra en conjunto con la palta es la marraqueta, apareciendo en un 72,68 % de las boletas analizadas. Le siguen la hallulla, el tomate a granel, los plátanos orgánicos y la malla de 1kg de limón. Se agregan entonces cinco variables explicativas más al análisis de compras de paltas: precio de la marraqueta, hallulla, tomate a granel, plátano orgánico y malla de 1kg de limón.

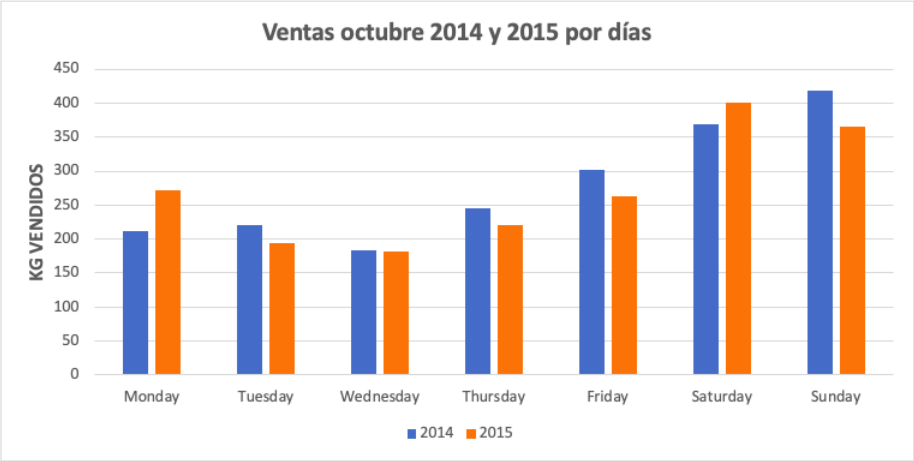


Figura 3.8: Ventas promedio de Tomate por días de la semana. Octubre 2014 y 2015

Ranking	Porcentaje	Nombre
1	72,68 %	Marraqueta
2	65,81 %	Hallulla
3	53,93 %	Tomate a Granel
4	21,60 %	Plátano Orgánico Premium a Granel
5	21,32 %	Limón Malla 1 Kg

Tabla 3.6: Análisis de Canasta de Compras Palta Hass

Análisis de Canasta Tomate

De manera análoga al análisis de la palta, se realiza un estudio de canasta de compras para conocer qué productos se compran en conjunto con el Tomate a Granel.

Ranking	Porcentaje	Nombre
1	28,68 %	Palta Hass Extra a Granel
2	28,41 %	Marraqueta
3	20,98 %	Papa en Malla 2 Kg
4	19,93 %	Hallulla
5	17,62 %	Plátano Extra a Granel

Tabla 3.7: Análisis de Canasta de Compras Tomate

Como se observa en la Tabla 3.7, los primeros cinco productos que más se compran en conjunto con el tomate no poseen un gran porcentaje de ventas como en el caso de la palta. Sin embargo, estos cinco productos se incluyen de todas formas como variables independientes para ver si existe algún efecto de éstos en la compra del Tomate a Granel. Aquí se aprecia que el producto más vendido en conjunto con el Tomate a Granel es la Palta Hass Extra a Granel con un 28,68 %, seguido por la Marraqueta con un 28,41 %, la Malla de 2 Kg de Papa, la Hallulla y finalmente el Plátano Extra a Granel.

3.1.6. Variables independientes

Finalmente, las variables independientes tanto para las paltas como para los tomates son 21. Las variables para la palta se detallan en la Tabla 3.8.

Nombre Variable	Explicación de la variable
P	Precio palta
W	Temperatura promedio del día
Día_()	Días de la semana (6 variables)
S	Stock
Día_()	Feriado, fin de semana o primer fin de semana (3 variables)
D_()	Descuento de hasta 10 %, de 10 % a 20 % y por más de 20 % (3 variables)
A_10	Aumento de hasta 10 % y de 10 % a 20 % (2 variables)
P_()	Precio marraqueta, hallulla, tomate, plátano orgánico y limón (5 variables)[11]

Tabla 3.8: Tabla de las variables independientes que se utilizan en los modelos Paltas

De manera análoga, las variables independientes que se utilizan en el análisis de la demanda de Tomate a Granel quedan detalladas en la Tabla 3.9.

Nombre Variable	Explicación de la variable
P	Precio tomate
W	Temperatura promedio del día
Día_()	Días de la semana (6 variables)
S	Stock
Día_()	Feriado, fin de semana o primer fin de semana (3 variables)
D_()	Descuento de hasta 10 %, de 10 % a 20 % y por más de 20 % (3 variables)
A_10	Aumento de hasta 10 % y de 10 % a 20 % (2 variables)
P_()	Precio palta hass extra a granel, marraqueta, papa en malla 2 Kg, hallulla y plátano orgánico extra a granel (5 variables)[12]

Tabla 3.9: Tabla de las variables independientes que se utilizan en los modelos Tomates

Cabe destacar que la única diferencia entre las tablas 3.8 y 3.9 son las cinco variables de la canasta de compras, respectivamente.

[11] De esta manera, las variables de la canasta de compras para la palta son PM, PH, PT, PP, PL, respectivamente.

[12] De esta manera, las variables de la canasta de compras para el tomate son $Ppalta$, PM, $Ppapa$, PH, $Ppeg$, respectivamente.

Capítulo 4

Aplicación de la metodología

En esta sección se implementan y discuten los modelos descritos en el Capítulo 2. Para lo que sigue, se utiliza un set de entrenamiento de 600 datos, 21 meses, y un set de testeo de 23 datos. Los sets de entrenamiento y de testeo no se obtienen sacando datos de manera aleatoria pues las ventas tienen forma de una serie de tiempo y, por ende, elegir datos aleatoriamente afectaría su temporalidad. Además, se tiene la hipótesis de que existe un *Lead Time* de 7 días entre que se piden los productos a los proveedores hasta que llegan al supermercado. De esta forma, el set de entrenamiento son los primeros 21 meses, luego viene un *Lead Time* de 7 días, y finalmente el set de testeo con los últimos 23 días de la base de datos. En la Figura 4.1 se muestra una representación visual de la división entre set de entrenamiento, *Lead Time* y set de testeo que se realiza en este proyecto.

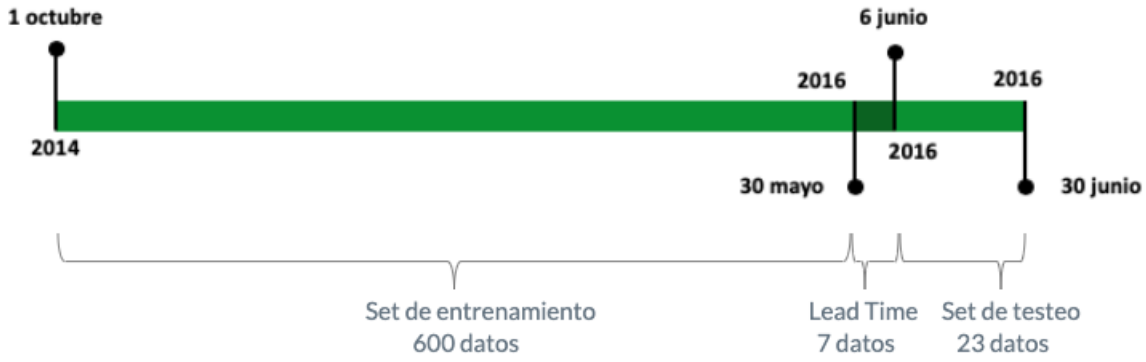


Figura 4.1: Representación visual de la división entre set de testeo y set de entrenamiento

A continuación, se presentan en un comienzo todos los resultados obtenidos de los once modelos para la Palta Hass Extra a Granel. Se prosigue, en una segunda instancia, a presentar los resultados obtenidos para el Tomate a Granel. Se muestra además, en cada modelo, una tabla para ver si los modelos son lo suficientemente robustos, o sea, para ver si son replicables. Para ello, se utilizan cinco bases de datos más pequeñas que la original[13] más la original, y en conjunto se calcula el valor promedio y la desviación estándar de los estadísticos

[13]Cabe destacar que estas cinco bases de datos provienen de igual manera de la base de datos original, solo que poseen menor cantidad de datos.

$MAPE$ y MAE tanto en el set de entrenamiento como en el set de testeo. Estos estadísticos serán llamados $mMAPE$ y $mMAE$ para el valor promedio y, $sMAPE$ y $sMAE$ para la desviación estándar del $MAPE$ y el MAE, respectivamente. En la Figura 4.2 se muestra una representación visual de la base original y las cinco bases más pequeñas para el testeo de robustez.

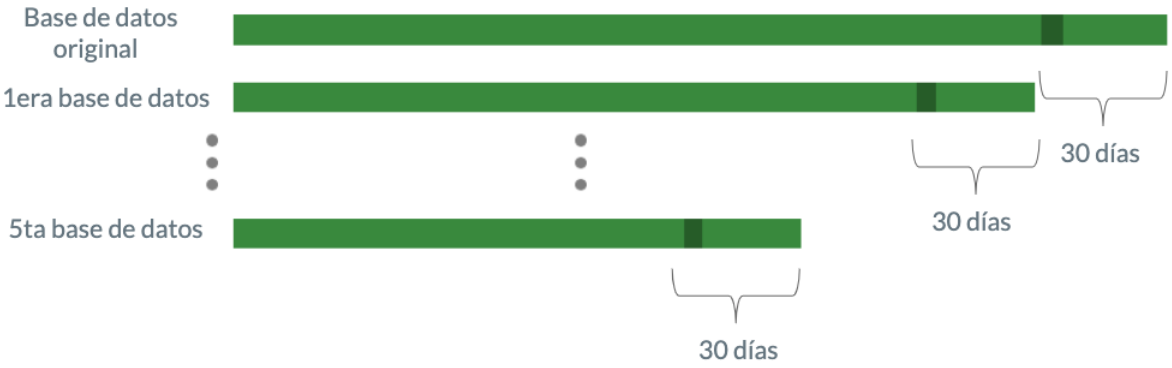


Figura 4.2: Representación visual de la base de datos original y las cinco bases para el testeo de robustez

4.1. Resultados para la Palta Hass Extra a Granel

A continuación, se muestran los resultados de los once modelos explicitados en el Capítulo 2, teniendo como variable dependiente en todos los modelos los kilos vendidos de Palta Hass Extra a Granel.

4.1.1. Naïve Forecast

Este es el método más básico que se emplea en este trabajo de título. Para calcular el pronóstico de la demanda, se utiliza sólo la variable de kilos vendidos de paltas. Como se señala en un principio, se hace el supuesto que existe Lead time de una semana, 7 días, para la llegada de productos, por lo que la predicción se basa en las ventas hechas hace dos semanas. De esta forma se obtiene el pronóstico que se muestra en la Figura 4.3 y en la Tabla 4.1.

Modelo	Set de entrenamiento		Set de testeo	
	MAPE	MAE	MAPE	MAE
Naïve Forecast	41,15 %	70,33	37,25 %	66,93

Tabla 4.1: Resultados predicción Naïve Forecast Paltas

Con respecto a la robustez del modelo, los resultados son los que se muestran en la Tabla 4.2.

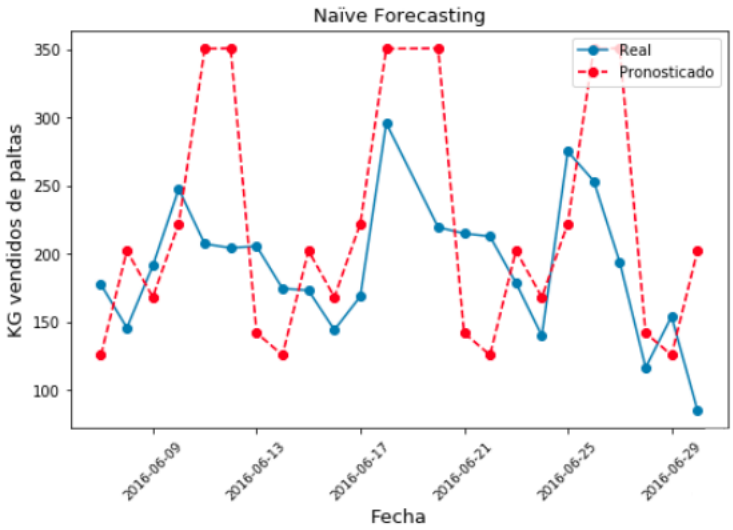


Figura 4.3: Resultados de la técnica de Naïve Forecasting Paltas

Modelo	Set de entrenamiento				Set de testeo			
	mMAPE	sMAPE	mMAE	sMAE	mMAPE	sMAPE	mMAE	sMAE
Naïve F.	42,36 %	0,80 %	71,00	1,73	46,91 %	16,00 %	89,10	33,86

Tabla 4.2: Resultados de robustez Moving Average Paltas

4.1.2. Moving Average

Al igual que el modelo anterior, en este modelo de Moving Average, MA, sólo se utiliza la variable de kilos vendidos por día. Para calcular el pronóstico se calcula la media móvil en el set de entrenamiento variando el tamaño de la muestra. En este sentido, para elegir el tamaño de la media móvil, se varía el tamaño n de $n = 3$ a $n = 301$ (lo cual es la mitad del set de entrenamiento más uno) y se comparan los errores cuadráticos medios. Finalmente, se elige una muestra de tamaño $n = 8$ cuyo error cuadrático medio[14] es el más bajo entre todos los testeados, de 7083 kilos de palta.

Como se muestra en la Figura 4.4, la media móvil en este caso es una curva que no oscila tanto como oscila la curva real. Una media de 8 no representa verdaderamente los datos. Sin embargo, es el que mejor ajusta.

Los resultados de la predicción son los que se muestran en la Tabla 4.3 y los resultados de robustez, los que se muestran en la Tabla 4.4

Modelo	Set de entrenamiento		Set de testeo	
	MAPE	MAE	MAPE	MAE
Moving Average	36,50 %	63,27	23,34 %	37,19

Tabla 4.3: Resultados predicción Moving Average Paltas

[14]El error cuadrático medio se calcula como sigue: $\frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t)^2$.

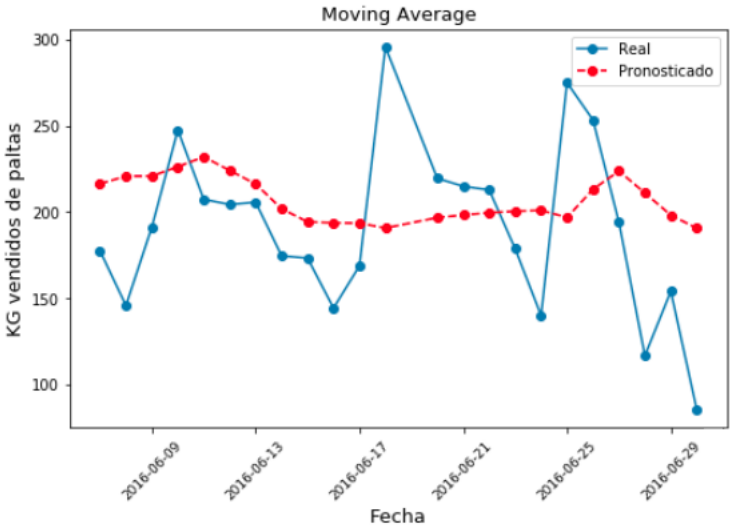


Figura 4.4: Resultados de la técnica de Moving Average Paltas

Modelo	Set de entrenamiento				Set de testeo			
	mMAPE	sMAPE	mMAE	sMAE	mMAPE	sMAPE	mMAE	sMAE
MA	37,04 %	0,44 %	63,01	1,22	28,00 %	8,48 %	58,13	29,34

Tabla 4.4: Resultados de robustez Moving Average Paltas

4.1.3. ARIMA

Este modelo es un modelo más complejo que los dos modelos anteriores, sin embargo, al igual que los modelos anteriormente señalados, se le entrega una sola variable al modelo, la de kilos vendidos por día. La diferencia con los modelos anteriores es que para aplicar este modelo la data se debe tratar previamente.

Se siguen los pasos entonces mencionados en la Figura 4.5. Se genera una serie estacionaria y se comprueba que esta sea estacionaria con el test de Dickey-Fuller[15]. Se comprueba con un 99 % de probabilidad que la serie que se utiliza para los pasos 3 y 4 es estacionaria; resultados en Apéndice C.1.

Una vez generada la serie estacionaria se prosigue a encontrar los parámetros óptimos del modelo. Se estudian entonces los gráficos ACF y $PACF$, los cuales indican los valores de q y p, respectivamente.

Como se aprecia en la Figura 4.6, tanto el gráfico de autocorrelación, ACF, como el de autocorrelación parcial, PACF, cortan por la primera vez el eje superior en 1. Esto quiere decir que $p = q = 1$, o sea, que el modelo posee 1 elemento autorregresivo y 1 elemento de medias móviles. Luego, gracias al test de Dickey-Fuller se sabe que se necesita sólo de

[15]El test de Dickey-Fuller es un test estadístico para chequear estacionalidad. Aquí la hipótesis nula es que la serie de tiempo es no-estacionaria. El resultado del test se compone de un Test Estadístico y Valores Críticos para diferentes niveles de confianza y, si el Test Estadístico es menor al Valor Crítico, se puede rechazar la hipótesis nula y decir que la serie es estacionaria.

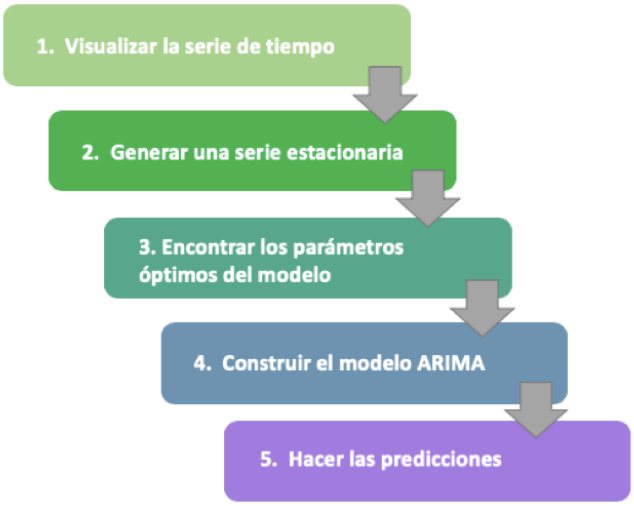


Figura 4.5: Pasos a seguir para poder realizar un modelo ARIMA

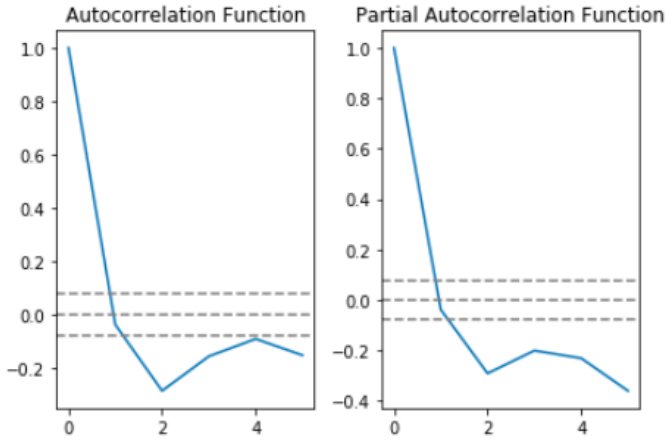


Figura 4.6: Gráficos ACF y PACF, para obtener los mejores parámetros p y q

1 diferencia para que los datos sean estacionarios. De esta manera el modelo que mejor se ajusta es un $ARIMA(1,1,1)$. La predicción se muestra en la Figura 4.7, y los resultados en la Tabla 4.5.

Modelo	Set de entrenamiento		Set de testeo	
	MAPE	MAE	MAPE	MAE
ARIMA	54,34 %	88,35	26,83 %	43,58

Tabla 4.5: Resultados predicción ARIMA Paltas

Finalmente, los resultados de robustez se muestran en la Tabla 4.6.

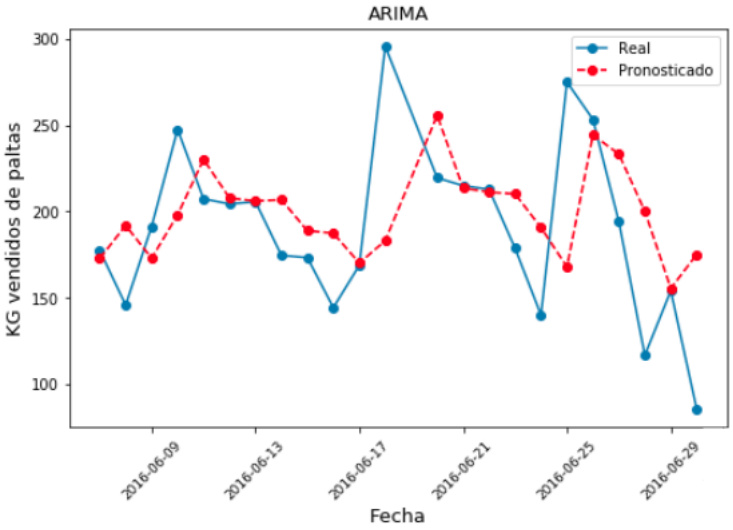


Figura 4.7: Resultados de la técnica de ARIMA Paltas

Modelo	Set de entrenamiento				Set de testeo			
	mMAPE	sMAPE	mMAE	sMAE	mMAPE	sMAPE	mMAE	sMAE
ARIMA	57,08 %	1,98 %	91,26	2,02	35,11 %	9,95 %	70,69	28,39

Tabla 4.6: Resultados de robustez ARIMA Paltas

4.1.4. Seasonal ARIMA

Este modelo lo que busca es poder integrar la temporalidad de los datos en los resultados. Para obtener los mejores parámetros, se realiza una búsqueda de cuadrícula o *greed search* para que, iterativamente, se vayan explorando las distintas combinaciones de parámetros. Para ello, se elige un rango de 0 a 4 para los parámetros $p, \mathrm{d}, q, P, D, Q$.

Dada la escaza cantidad de datos que se posee (solo 600 datos en total), no se puede usar una temporalidad anual (360 días) ya que no existen datos suficientes que reflejen la temporalidad en ese periodo. Se elige una temporalidad semanal de 7 días[16], ya que las compras en general de los supermercados tienen una temporalidad semanal.

Se prosigue a testear $4^6 = 4096$ combinaciones de parámetros para obtener que el mejor modelo es un $ARIMA(0, 1, 3)(1, 2, 3)7$, según la métrica de AIC[17]. En este sentido, la predicción queda como se muestra en la Figura 4.8, y los resultados de la predicción como se muestra en la Tabla 4.7. Finalmente, los resultados de la robustez son los que se muestran en la Tabla 4.8.

[16]Cabe destacar que se toma esta misma temporalidad para los otros modelos que necesitan el parámetro de temporalidad.

[17]Se obtiene un AIC de 6759 para este modelo, el menor AIC entre todos los modelos testeados.

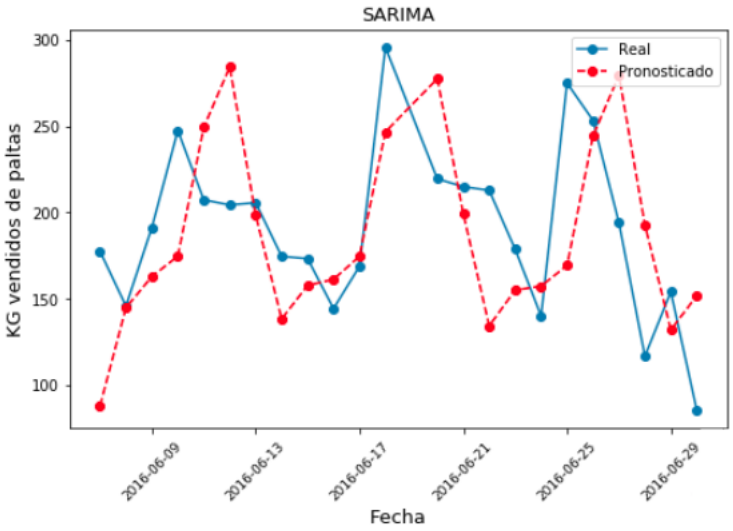


Figura 4.8: Resultados de la técnica de SARIMA Paltas

Modelo	Set de entrenamiento		Set de testeo	
	MAPE	MAE	MAPE	MAE
SARIMA	29,26 %	53,15	24,34 %	43,60

Tabla 4.7: Resultados predicción SARIMA Paltas

Modelo	Set de entrenamiento				Set de testeo			
	mMAPE	sMAPE	mMAE	sMAE	mMAPE	sMAPE	mMAE	sMAE
SARIMA	29,87 %	0,48 %	53,46	1,28	38,52 %	15,80 %	77,03	36,01

Tabla 4.8: Resultados de robustez SARIMA Paltas

4.1.5. Suavización Exponencial Triple de Holt-Winters

En la Suavización Exponencial Triple, SET, se realiza un algoritmo de tal manera que éste entregue los mejores parámetros posibles. Estos parámetros son los que siguen: $\alpha = 0,5$, $\beta = 0$ y $\gamma = 0,15$. De estos resultados se puede decir que los datos no poseen una tendencia ni creciente ni decreciente clara, pero sí poseen una estacionalidad, o sea, que las ventas de palta tienen ciclos. Los resultados de esta predicción se muestran en la Figura 4.9 y en la Tabla 4.9.

Modelo	Set de entrenamiento		Set de testeo	
	MAPE	MAE	MAPE	MAE
Suavización Exponencial Triple	24,62 %	40,78	21,80 %	35,66

Tabla 4.9: Resultados predicción Suavización Exponencial Triple Paltas

Finalmente, los resultados de la robustez son los que se muestran en la Tabla 4.10.

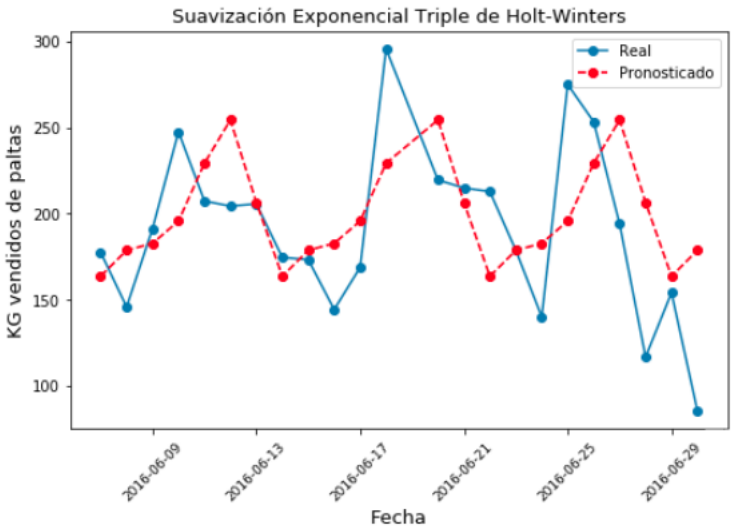


Figura 4.9: Resultados de la técnica de Suavización Exponencial Triple Paltas

Modelo	Set de entrenamiento				Set de testeo			
	mMAPE	sMAPE	mMAE	sMAE	mMAPE	sMAPE	mMAE	sMAE
SET	23,45 %	0,85 %	38,86	1,79	46,83 %	40,95 %	95,24	79,37

Tabla 4.10: Resultados de robustez Suavización Exponencial Triple Paltas

4.1.6. Regresión Lineal Múltiple

Para la regresión lineal múltiple, RLM, se utilizan todas las variables detalladas en la Tabla 3.8 más sus interacciones. Se obtiene entonces un total de 46 variables independientes. Para reducir el número de variables explicativas se utiliza el método Lasso, el cual deja un total de 16 variables[18]. Se continúa a eliminar variables cuya correlación afecta en el resultado y cuyas variables son estadísticamente insignificativas, quedando finalmente con un total de 13 variables[19]. La ecuación de regresión resultante es la siguiente:

$$\begin{aligned}\hat{y}_t = -466,93 - 0,06PT - 0,0013W \cdot PH - 0,0003P \cdot PH + 0,03PL + 0,04S + 0,09PP + \\ 25,26Date\_Viernes + 14,10A\_10 + 17,33Date\_Domingo + \\ 87,70Dia\_FinDeSemana + 166,51D\_mas\_20 + 0,26P + \\ 0,72PM\end{aligned} \tag{4.1}$$

Los resultados de esta regresión lineal múltiple son los que se observan en la Figura 4.10

[18]Para elegir el mejor Lasso se realiza una columna de posibles α con valores muy grandes y muy pequeños, conteniendo todo el rango de escenarios desde el modelo vacío el cual posee solo el intercepto hasta el modelo de OLS. Se prosigue a realizar un *k-fold Crossvalidation* para testear el mejor Lasso; se elige entonces un *10-fold Crossvalidation* obteniendo finalmente $\alpha = 0{,}092$ para el modelo Lasso.

[19]El resumen de este análisis se encuentra en Apéndice C.2.

y los resultados de la predicción los que se muestran en la Tabla 4.11.

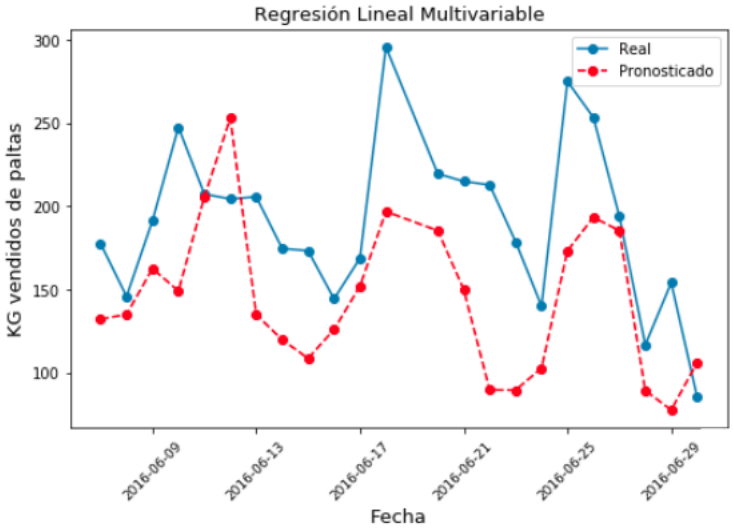


Figura 4.10: Resultados de la técnica de Regresión Lineal Múltiple Paltas

Modelo	Set de entrenamiento		Set de testeo	
	MAPE	MAE	MAPE	MAE
Regresión Lineal Múltiple	30,51 %	47,83	26,73 %	52,27

Tabla 4.11: Resultados predicción Regresión Lineal Múltiple Paltas

Finalmente, los resultados de la robustez de este modelo se observan en la Tabla 4.12.

Modelo	Set de entrenamiento				Set de testeo			
	mMAPE	sMAPE	mMAE	sMAE	mMAPE	sMAPE	mMAE	sMAE
RLM	31,72 %	0,85 %	48,86	0,81	23,67 %	6,08 %	50,13	13,00

Tabla 4.12: Resultados de robustez Regresión Lineal Múltiple Paltas

4.1.7. Árbol de Regresión

Al igual que para la Regresión Lineal Múltiple, en Árboles de Regresiones, AR, se utilizan todas las variables detalladas en la Tabla 3.8. La diferencia con Regresión Lineal Múltiple es que en esta metodología no se incluyen las interacciones entre las variables.

Se testean distintos árboles con distintas profundidades; los detalles del estudio se encuentran en el Apéndice C.3. De esta forma, se elige el mejor árbol, un árbol de profundidad 7 (o de 7 ramas) cuyas predicciones y detalles son los que se observan en la Figura 4.11 y en la Tabla 4.13. Finalmente, los resultados de robustez son los que se muestran en la Tabla 4.14.

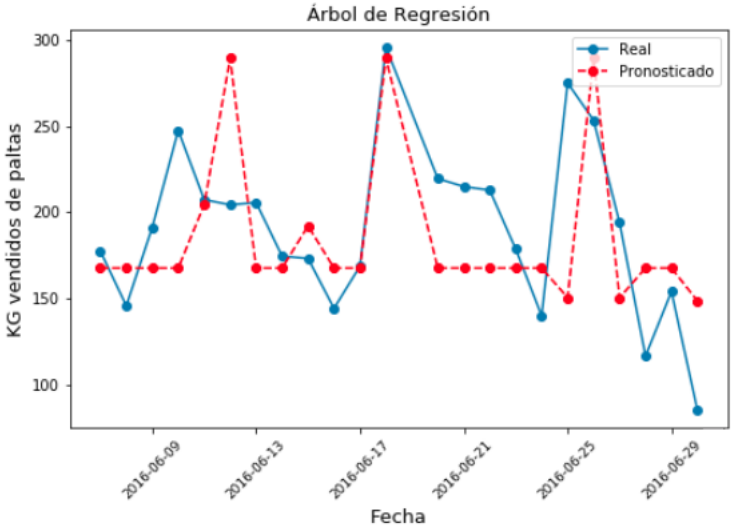


Figura 4.11: Resultados de la técnica de Árbol de Regresión con 7 ramas Paltas

Modelo	Set de entrenamiento		Set de testeo	
	MAPE	MAE	MAPE	MAE
Árbol de Regresión	19,71 %	31,57	20,26 %	39,27

Tabla 4.13: Resultados predicción Árbol de Regresión Paltas

Modelo	Set de entrenamiento				Set de testeo			
	mMAPE	sMAPE	mMAE	sMAE	mMAPE	sMAPE	mMAE	sMAE
AR	19,90 %	1,11 %	30,76	0,79	28,63 %	10,82 %	62,57	35,36

Tabla 4.14: Resultados de robustez Árbol de Regresión Paltas

4.1.8. Random Forest para Regresiones

Para Random Forest, RF, se debe especificar tanto la profundidad del árbol o cantidad de ramas como la cantidad de árboles que se utilizan. Se realiza un algoritmo que testea distintos modelos cuyas profundidades se contienen en un rango de 3 a 8 ramas y la cantidad de árboles se contienen en un rango de 5 a 1000 árboles. En Apéndice C.4 se muestra un extracto a modo de ejemplo de distintos Random Forest testeados[20]. Finalmente se encuentra que el mejor Random Forest es aquel que tiene 6 ramas y 50 árboles. Los resultados de este Random Forest se observan en la Figura 4.12 y en la Tabla 4.15. Los resultados de la robustez se observan en la Tabla 4.16.

Finalmente, la importancia de las variables de esta metodología se encuentran en Apéndice C.5.

[20]Se realiza esta Tabla para demostrar que el mejor árbol es aquel que escoge el algorítmo.

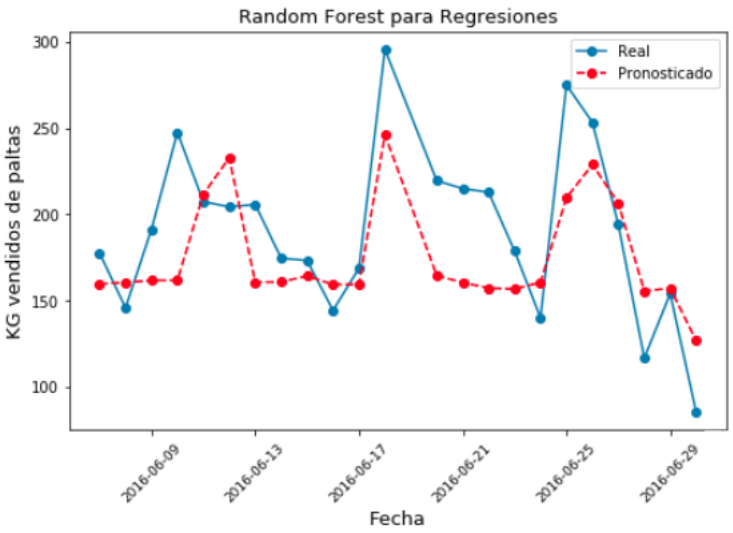


Figura 4.12: Resultados de la técnica de Random Forest Paltas

Modelo	Set de entrenamiento		Set de testeo	
	MAPE	MAE	MAPE	MAE
Random Forest	22,84 %	34,93	16,58 %	31,13

Tabla 4.15: Resultados predicción Random Forest Paltas

Modelo	Set de entrenamiento				Set de testeo			
	mMAPE	sMAPE	mMAE	sMAE	mMAPE	sMAPE	mMAE	sMAE
RF	23,19 %	0,28 %	34,67	0,79	24,78 %	9,05 %	51,47	22,46

Tabla 4.16: Resultados de robustez Random Forest Paltas

4.1.9. Support Vector Regressor

Para Support Vector Regressor, SVR, se testean distintos tipos de *kernel* en el set de entrenamiento como se observan en la Tabla 4.17.

Kernel	R^2
Lineal	46,77 %
Polinomial	56,27 %
Sigmoide	2,61 %
Radial	52,50 %

Tabla 4.17: Valores de R^2 para elección de kernel

Como se muestra en la Tabla 4.17, el mejor kernel según R^2 es un kernel polinomial, kernel que es utilizado para testear el modelo. Se prosigue entonces a realizar la metodología de Support Vector Regressor obteniendo los resultados detallados en la Figura 4.13 y la Tabla 4.18.

Figura 4.13: Resultados de la técnica de Support Vector Regressor Paltas

Modelo	Set de entrenamiento		Set de testeo	
	MAPE	MAE	MAPE	MAE
Support Vector Regressor	29,58 %	47,25	16,07 %	28,74

Tabla 4.18: Resultados predicción Support Vector Regressor Paltas

Finalmente, los resultados de robustez son los que se muestran en la Tabla 4.19.

Modelo	Set de entrenamiento				Set de testeo			
	mMAPE	sMAPE	mMAE	sMAE	mMAPE	sMAPE	mMAE	sMAE
SVR	35,35 %	12,34 %	52,26	18,18	25,38 %	7,03 %	53,84	25,86

Tabla 4.19: Resultados de robustez Support Vector Regressor Paltas

4.1.10. Redes Neuronales Artificiales

Para esta metodología, RNA, se testean varias RNA con distinta cantidad de capas ocultas, de neuronas y de epochs como se muestra en Apéndice C.6. Finalmente, se tiene que la mejor RNA es aquella que posee 21 variables en la capa de entrada (las 21 variables independientes del problema), 3 capas ocultas con 10 neuronas cada una, 700 epochs de aprendizaje y *una* capa de salida la cual representa la predicción de demanda. Los resultados de este método se muestran en la Figura 4.14 y en la Tabla 4.20.

Modelo	Set de entrenamiento		Set de testeo	
	MAPE	MAE	MAPE	MAE
Redes Neuronales Artificiales	24,95 %	39,39	24,42 %	40,95

Tabla 4.20: Resultados predicción Redes Neuronales Artificiales Paltas

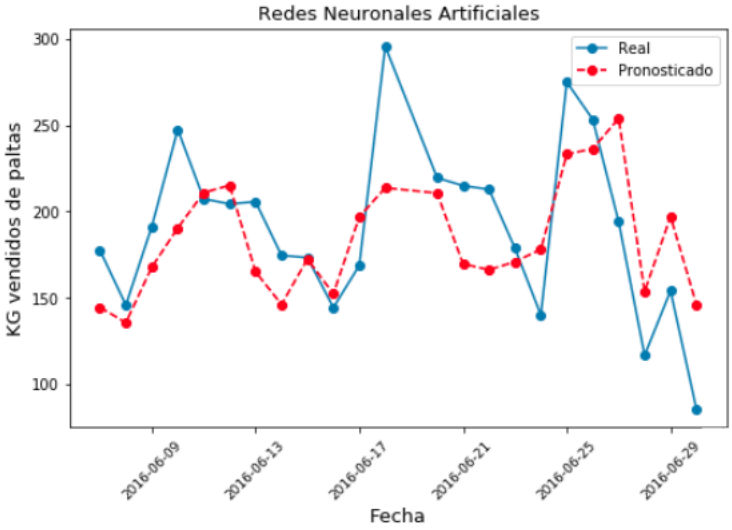


Figura 4.14: Resultados de la técnica de Redes Neuronales Artificiales Paltas

Finalmente, la robustez de esta RNA se muestra en la Tabla 4.21.

Modelo	Set de entrenamiento				Set de testeo			
	mMAPE	sMAPE	mMAE	sMAE	mMAPE	sMAPE	mMAE	sMAE
RNA	26,37 %	1,14 %	40,28	0,92	28,90 %	11,07 %	55,22	20,56

Tabla 4.21: Resultados de robustez Redes Neuronales Artificiales Paltas

4.1.11. Redes Neuronales Recurrentes

Para este modelo se testean distintas combinaciones de parámetros de Redes Neuronales Recurrentes, RNR, como se observa en Apéndice C.7[21]. Finalmente se escoge aquella que posee *una* capa de entrada (la variable a predecir), 2 capas ocultas con 50 neuronas cada una, 200 epochs y una capa de salida, ya que ésta posee el menor error $MAPE$ en el set de entrenamiento con respecto a las otras.

Los resultados de esta metodología se muestran en la Figura 4.15 y en la Tabla 4.22.

Modelo	Set de entrenamiento		Set de testeo	
	MAPE	MAE	MAPE	MAE
Redes Neuronales Recurrentes	26,83 %	48,34	19,81 %	34,23

Tabla 4.22: Resultados predicción Redes Neuronales Recurrentes Paltas

Finalmente, la robustez de esta RNR se muestra en la Tabla 4.23.

[21] Cabe destacar que todas estas Redes Neuronales Recurrentes testeadas poseían una variable *Dropout* de un 20 % la cual se encarga de eliminar el 20 % de las neuronas más correlacionadas en cada epoch.

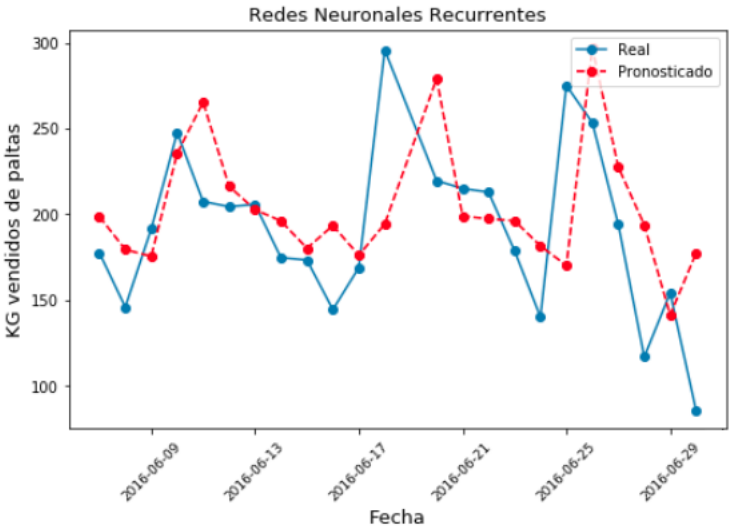


Figura 4.15: Resultados de la técnica Redes Neuronales Recurrentes Paltas

Modelo	Set de entrenamiento				Set de testeo			
	mMAPE	sMAPE	mMAE	sMAE	mMAPE	sMAPE	mMAE	sMAE
RNR	29,27 %	1,47 %	48,57	1,82	27,20 %	7,91 %	54,22	21,32

Tabla 4.23: Resultados de robustez Redes Neuronales Recurrentes Paltas

A continuación, se muestran los resultados obtenidos para el Tomate a Granel. En el capítulo siguiente, Capítulo 5 Análisis de resultados, se realiza el análisis de los resultados obtenidos anteriormente y los resultados del Tomate a Granel.

4.2. Resultados para el Tomate a Granel

Se prosigue a realizar los mismos métodos testeados anteriormente en la Palta Hass Extra a Granel pero ahora para el Tomate a Granel. Los resultados y la robustez se muestran a continuación.

4.2.1. Naïve Forecast

Para calcular el pronóstico de la demanda en Naïve Forecast, NF, se utiliza sólo la variable de kilos vendidos de tomate. Se hace el supuesto que la reposición tiene un Lead time de una semana, 7 días, al igual que en el modelo para la palta. De esta manera se obtiene el pronóstico como se muestra en la Figura 4.16, y la predicción como se detalla en la Tabla 4.24. Los resultados de la robustez se muestran en la Tabla 4.25.

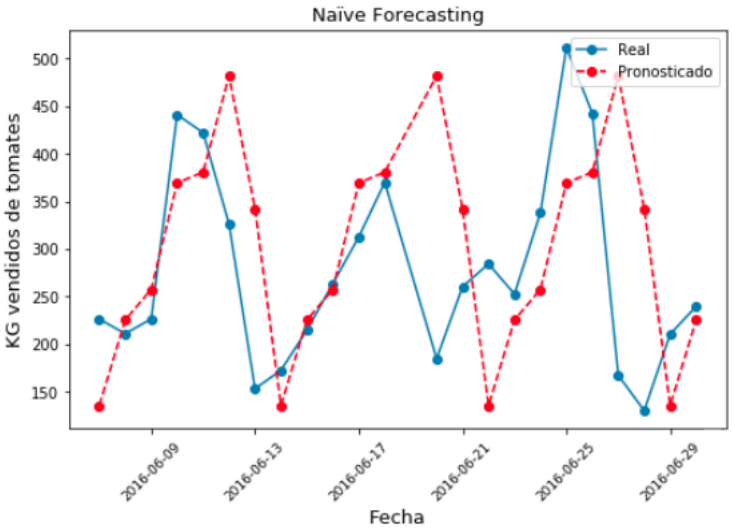


Figura 4.16: Resultados de la técnica de Naïve Forecasting Tomates

Modelo	Set de entrenamiento		Set de testeo	
	MAPE	MAE	MAPE	MAE
Naïve Forecasting	49,93 %	120,49	44,54 %	94,50

Tabla 4.24: Resultados predicción Naïve Forecast Tomates

Modelo	Set de entrenamiento				Set de testeo			
	mMAPE	sMAPE	mMAE	sMAE	mMAPE	sMAPE	mMAE	sMAE
Naive F.	52,64 %	2,35 %	120,23	2,26	40,42 %	22,95 %	127,47	76,97

Tabla 4.25: Resultados de robustez Naïve Forecast Tomates

4.2.2. Moving Average

Al igual que el modelo anterior, en este modelo sólo se utiliza la variable de kilos vendidos por día. Para calcular el pronóstico se calcula la media móvil en el set de entrenamiento variando el tamaño de la muestra. Para ello, para elegir el tamaño de la media móvil, se varía el tamaño n de $n = 3$ a $n = 301$ (lo cual es la mitad del set de entrenamiento más uno) y se comparan los errores cuadráticos medios. Finalmente, se elige una muestra de tamaño $n = 8$ cuyo error cuadrático medio es el más bajo, de 22075.

Como se muestra en la Figura 4.17, la media móvil en este caso es una curva que no oscila tanto como oscila la curva real; ocurre lo mismo que en el caso de la palta. En este sentido, una media de 8 no representa verdaderamente los datos, sin embargo, es el que mejor ajusta. Finalmente, los resultados de la predicción son los que se muestran en la Tabla 4.26 y, los resultados de robustez, los que se muestran en la Tabla 4.27.

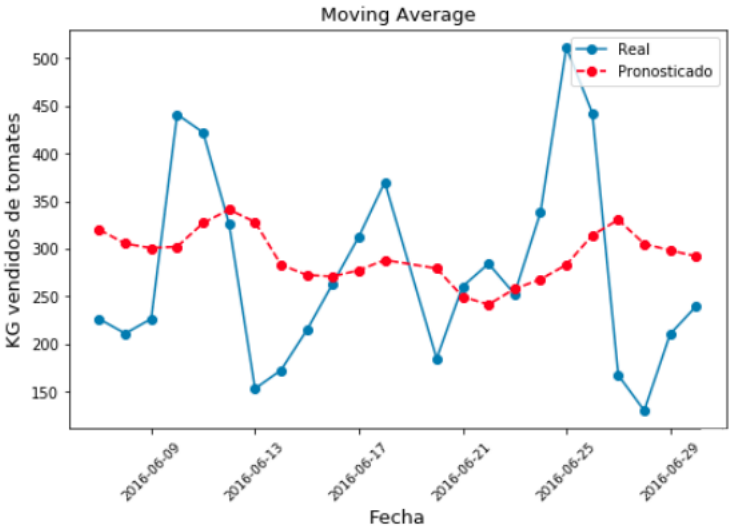


Figura 4.17: Resultados de la técnica de Moving Average Tomates

Modelo	Set de entrenamiento		Set de testeo	
	MAPE	MAE	MAPE	MAE
Moving Average	35,13 %	111,85	33,62 %	77,54

Tabla 4.26: Resultados predicción Moving Average Tomates

Modelo	Set de entrenamiento				Set de testeo			
	mMAPE	sMAPE	mMAE	sMAE	mMAPE	sMAPE	mMAE	sMAE
MA	35,26 %	0,51 %	111,09	1,51	30,11 %	7,46 %	96,49	27,66

Tabla 4.27: Resultados de robustez Moving Average Tomates

4.2.3. ARIMA

Al igual que los modelos anteriores, a este modelo se le entrega una sola variable al modelo, la de kilos vendidos por día de tomates. La diferencia con los modelos anteriores es que para aplicar este modelo se debe tratar la data previamente.

Al igual que para el caso de las paltas, se siguen los pasos mencionados en la Figura 4.5. Se genera una serie estacionaria y se comprueba que esta sea estacionaria con el test de Dickey-Fuller. Se comprueba con un 99 % de probabilidad que la serie que se utiliza para los pasos 3 y 4 de la Figura 4.5 es estacionaria; resultados en Apéndice D.1.

Una vez generada la serie estacionaria se prosigue a encontrar los parámetros óptimos del modelo. Se estudian entonces los gráficos ACF y $PACF$. De esta manera, a partir de la Figura 4.18, se obtiene que $p = 3$ y $q = 2$, o sea, que el modelo posee 3 términos autorregresivos y 2 términos de medias móviles. Luego, gracias al test de Dickey-Fuller se sabe que se necesita sólo de 1 diferencia, d = 1, para que los datos sean estacionarios. De esta manera el modelo que mejor se ajusta es un $ARIMA(3, 1, 2)$. La predicción se muestra en la Figura 4.19.

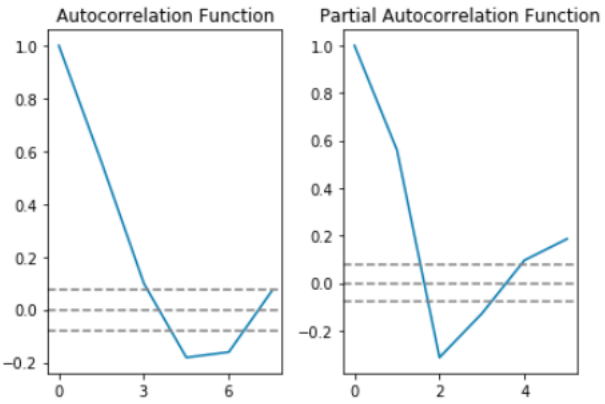


Figura 4.18: Gráficos ACF y PACF, para obtener los mejores parámetros p y q

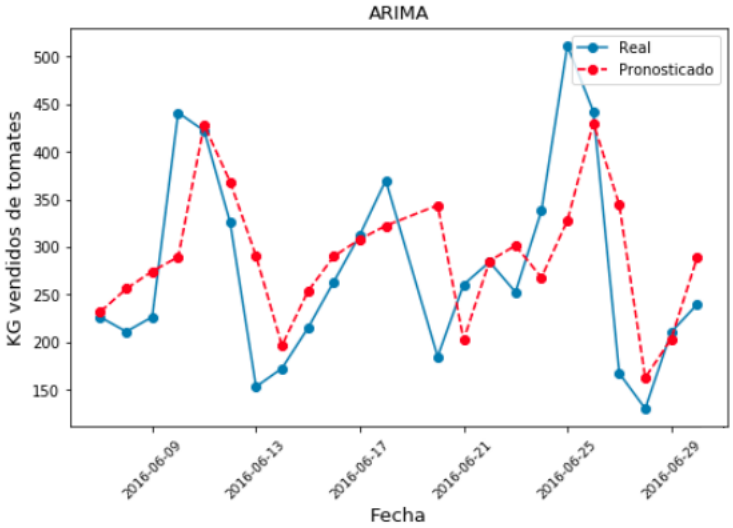


Figura 4.19: Resultados de la técnica de ARIMA Tomates

En este caso, los resultados de predicción se muestran en la Tabla 4.28 y, los resultados de la robustez en la Tabla 4.29.

Modelo	Set de entrenamiento		Set de testeo	
	MAPE	MAE	MAPE	MAE
ARIMA	45,77 %	140,73	40,85 %	99,10

Tabla 4.28: Resultados predicción ARIMA Tomates

Modelo	Set de entrenamiento				Set de testeo			
	mMAPE	sMAPE	mMAE	sMAE	mMAPE	sMAPE	mMAE	sMAE
ARIMA	38,83 %	9,41 %	127,11	30,70	45,90 %	2,66 %	135,02	17,70

Tabla 4.29: Resultados de robustez ARIMA Tomates

4.2.4. Seasonal ARIMA

Este modelo lo que busca es poder integrar la estacionalidad de los datos en los resultados, al igual que en el caso de la palta. Para obtener los mejores parámetros, se realiza una búsqueda de cuadrícula o *greed search* para que, iterativamente, se vayan explorando las distintas combinaciones de parámetros. Para ello, se elige un rango de 0 a 4 para los parámetros $p, \mathrm{d}, q, P, D, Q$.

Al igual que para el caso de la palta, se elige una estacionalidad semanal de 7 días, ya que las compras en general de los supermercados tienen una temporalidad semanal.

Se prosigue a testear $4^6 = 4096$ combinaciones de parámetros para obtener que el mejor modelo es un $ARIMA(0,1,3)(1,2,3)7$[22], según la métrica de AIC[23]. En este sentido, la predicción queda como se muestra en la Figura 4.20.

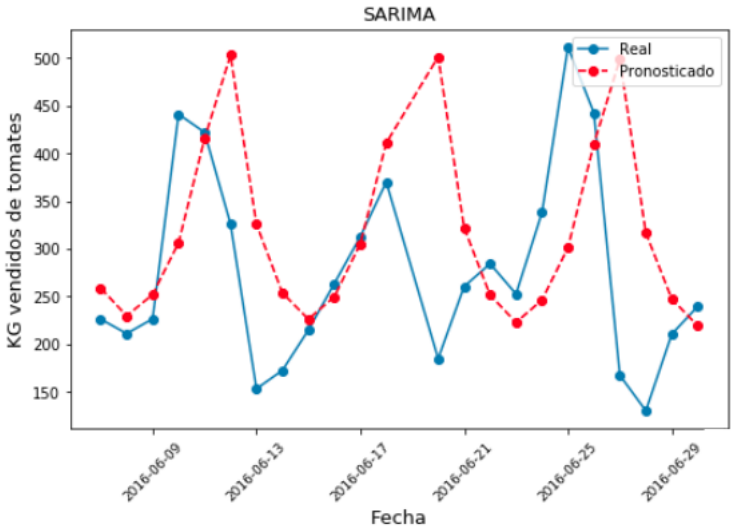


Figura 4.20: Resultados de la técnica de SARIMA Tomates

Los resultados de la predicción anterior son los que se muestran en la Tabla 4.30 y, los resultados de robustez, los que se muestran en la Tabla 4.31.

Modelo	Set de entrenamiento		Set de testeo	
	MAPE	MAE	MAPE	MAE
SARIMA	26,90 %	87,81	42,13 %	90,26

Tabla 4.30: Resultados predicción SARIMA Tomates

[22]Cabe destacar que estos resultados son los mismos que para la palta. Esto no quiere decir que el MAPE y el MAE vayan a ser los mismos, sino que el modelo que mejor ajusta es el que posee la configuración de $ARIMA(0,1,3)(1,2,3)7$.

[23]Se obtiene un AIC de 7017 para este modelo, el menor AIC entre todos los modelos testeados.

Modelo	Set de entrenamiento				Set de testeo			
	mMAPE	sMAPE	mMAE	sMAE	mMAPE	sMAPE	mMAE	sMAE
SARIMA	27,48 %	0,62 %	89,05	1,64	36,65 %	11,19 %	99,33	20,84

Tabla 4.31: Resultados de robustez SARIMA Tomates

4.2.5. Suavización Exponencial Triple de Holt-Winters

Para la Suavización Exponencial Triple, SET, los mejores parámetros que nos entrega el algoritmo son los que siguen: $\alpha = 0,55$, $\beta = 0$ y $\gamma = 0,15$. De estos resultados se puede decir que los datos no poseen una tendencia ni creciente ni decreciente clara, pero sí poseen una estacionalidad, al igual que para el caso de la palta. La predicción se muestran en la Figura 4.21.

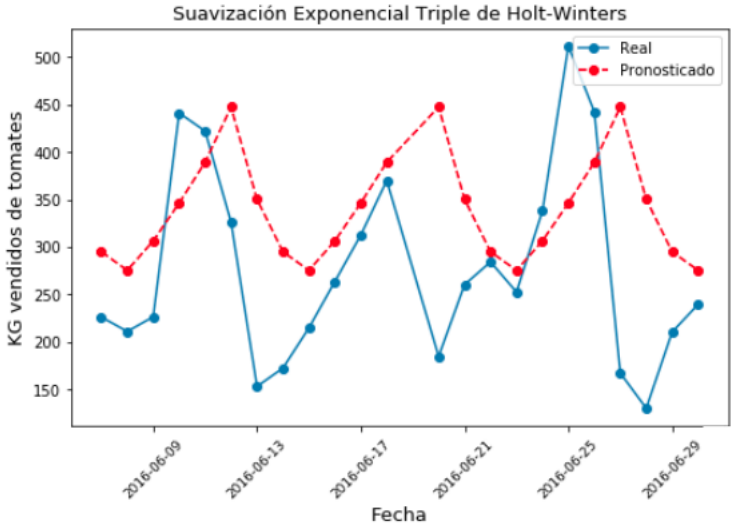


Figura 4.21: Resultados de la técnica de Suavización Exponencial Triple Tomates

Finalmente, los resultados de la predicción se muestran en la Tabla 4.32 y, los resultados de la robustez, en la Tabla 4.33.

Modelo	Set de entrenamiento		Set de testeo	
	MAPE	MAE	MAPE	MAE
Suavización Exponencial Triple	28,92 %	70,09	46,12 %	95,53

Tabla 4.32: Resultados predicción Suavización Exponencial Triple Tomates

Modelo	Set de entrenamiento				Set de testeo			
	mMAPE	sMAPE	mMAE	sMAE	mMAPE	sMAPE	mMAE	sMAE
SET	28,98 %	0,39 %	67,60	1,88	54,39 %	48,44 %	172,14	167,37

Tabla 4.33: Resultados de robustez Suavización Exponencial Triple Tomates

4.2.6. Regresión Lineal Múltiple

Para la Regresión Lineal Múltiple, RLM, se utilizan todas las variables detalladas en la Tabla 3.9 más sus interacciones. Se obtiene entonces un total de 45 variables independientes, al igual que para el tomate. Para reducir el número de variables explicativas se utiliza el método Lasso, el cual deja un total de 15 variables[24]. Se continúa a eliminar variables cuya correlación afecta en el resultado y cuyas variables son estadísticamente insignificativas, quedando finalmente con un total de 11 variables[25]. La ecuación de regresión resultante es la siguiente:

$$\begin{aligned}\hat{y}_t = 551 - 35,89D\_20\_10 - 5,84W \cdot Date\_Domingo - 4,93 \cdot Date\_Miercoles - \\ 3,95W \cdot Date\_Jueves - 3,51W \cdot Date\_Lunes - 0,18P - \\ 0,03Ppalta + 0,02Ppapa + 0,03S + 117,46D\_mas\_20 + \\ 159,15Dia\_FinDeSemana\end{aligned} \tag{4.2}$$

Los resultados de esta Regresión Lineal Múltiple son los que se observan en la Figura 4.22 y los resultados de la predicción los que se muestran en la Tabla 4.34.

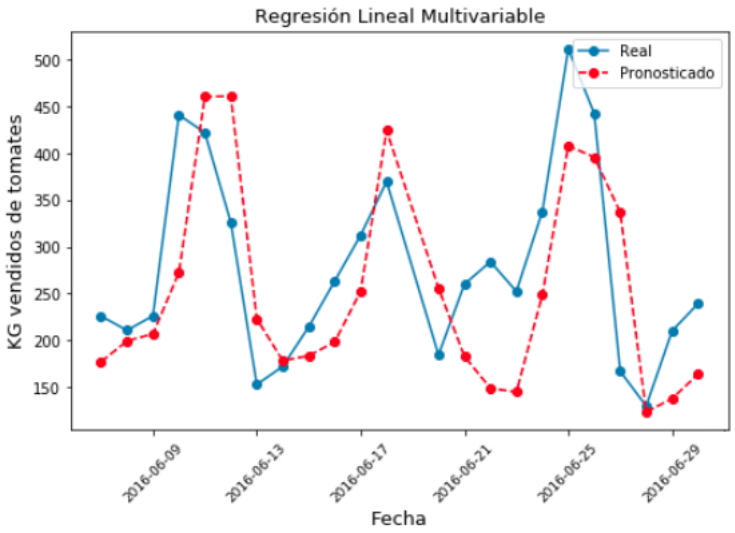


Figura 4.22: Resultados de la técnica de Regresión Lineal Múltiple Tomates

Finalmente, los resultados de la robustez de la Regresión Lineal Múltiple son los que se muestran en la Tabla 4.12.

[24]Para elegir el mejor Lasso se prosigue de la misma manera que para las paltas. Se realiza una columna de posibles α con valores muy grandes y muy pequeños, conteniendo todo el rango de escenarios desde el modelo vacío el cual posee solo el intercepto hasta el modelo de OLS. Se prosigue a realizar un *k-fold Crossvalidation* para testear el mejor Lasso; se elige entonces un *10-fold Crossvalidation* obteniendo finalmente $\alpha = 0{,}139$ para el modelo Lasso.

[25]El resumen detallado de los resultados se encuentran en Apéndice D.2

Modelo	Set de entrenamiento		Set de testeo	
	MAPE	MAE	MAPE	MAE
Regresión Lineal Múltiple	31,31 %	73,11	27,65 %	72,38

Tabla 4.34: Resultados predicción Regresión Lineal Múltiple Tomates

Modelo	Set de entrenamiento				Set de testeo			
	mMAPE	sMAPE	mMAE	sMAE	mMAPE	sMAPE	mMAE	sMAE
RLM	32,88 %	1,41 %	73,06	0,79	21,71 %	5,54 %	73,93	17,17

Tabla 4.35: Resultados de robustez Regresión Lineal Múltiple Tomates

4.2.7. Árbol de Regresión

En esta metodología se utilizan todas las variables detalladas en la Tabla 3.9, sin embargo, contrario a la Regresión Lineal Múltiple, en este modelo no se consideran las interacciones entre las variables.

Se prosigue a testear distintos árboles con distintas profundidades, los detalles del estudio se encuentran en el Apéndice D.3. De esta forma, se elige el mejor árbol, un árbol de profundidad 6 (o de 6 ramas) cuyas predicciones y detalles son los que se observan en la Figura 4.23 y en la Tabla 4.36.

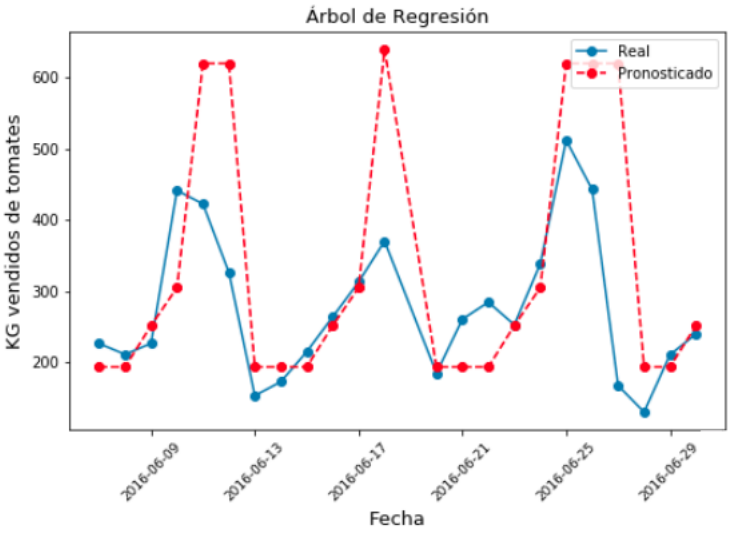


Figura 4.23: Resultados de la técnica de Árbol de Regresión con 6 ramas Tomates

Finalmente, los resultados de la robustez de este Árbol de Regresión se muestran en la Tabla 4.37.

Modelo	Set de entrenamiento		Set de testeo	
	MAPE	MAE	MAPE	MAE
Árbol de Regresión	28,57 %	52,47	34,68 %	91,64

Tabla 4.36: Resultados predicción Árbol de Regresión Tomates

Modelo	Set de entrenamiento				Set de testeo			
	mMAPE	sMAPE	mMAE	sMAE	mMAPE	sMAPE	mMAE	sMAE
AR	28,58 %	0,93 %	50,76	2,27	25,88 %	8,62 %	90,16	36,65

Tabla 4.37: Resultados de robustez Árbol de Regresión Tomates

4.2.8. Random Forest para Regresiones

Para esta metodología se debe especificar tanto la profundidad del árbol o cantidad de ramas como la cantidad de árboles que se utilizan. Se realiza un algoritmo que testea distintos modelos cuyas profundidades se contienen en un rango de 3 a 8 ramas y la cantidad de árboles se contienen en un rango de 5 a 100 árboles como se observa en el Apéndice D.4.

De esta manera, gracias al algoritmo generado se encuentra que el mejor Random Forest es aquel que tiene 7 ramas y 60 árboles.

Los resultados de este Random Forest se observan en la Figura 4.24 y en la Tabla 4.38. La importancia de las variables de esta metodología se encuentran en Apéndice D.5.

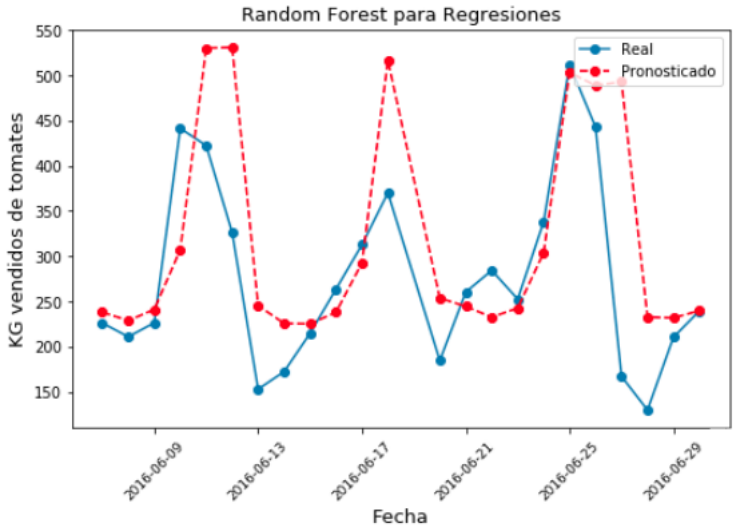


Figura 4.24: Resultados de la técnica de Random Forest Tomates

Finalmente, los resultados de la robustez de este Árbol de Regresión se muestran en la Tabla 4.39.

Modelo	Set de entrenamiento		Set de testeo	
	MAPE	MAE	MAPE	MAE
Random Forest	22,38 %	47,92	28,86 %	66,33

Tabla 4.38: Resultados predicción Random Forest Tomates

Modelo	Set de entrenamiento				Set de testeo			
	mMAPE	sMAPE	mMAE	sMAE	mMAPE	sMAPE	mMAE	sMAE
RF	23,53 %	1,17 %	47,24	0,38	21,66 %	5,62 %	70,32	18,66

Tabla 4.39: Resultados de robustez Random Forest Tomates

4.2.9. Support Vector Regressor

Para esta metodología, Support Vector Regressor, SVR, se testean distintos tipos de *kernel* como se observan en la Tabla 4.40.

Kernel	R^2
Lineal	24,12 %
Polinomial	35,47 %
Sigmoide	2,88 %
Radial	40,50 %

Tabla 4.40: Valores de R^2 para elección de kernel Tomates

Como se muestra en la Tabla 4.40, el mejor kernel según R^2 es un kernel radial. Se prosigue entonces a realizar la metodología de Support Vector Regressor obteniendo los resultados detallados en la Figura 4.25 y la Tabla 4.41.

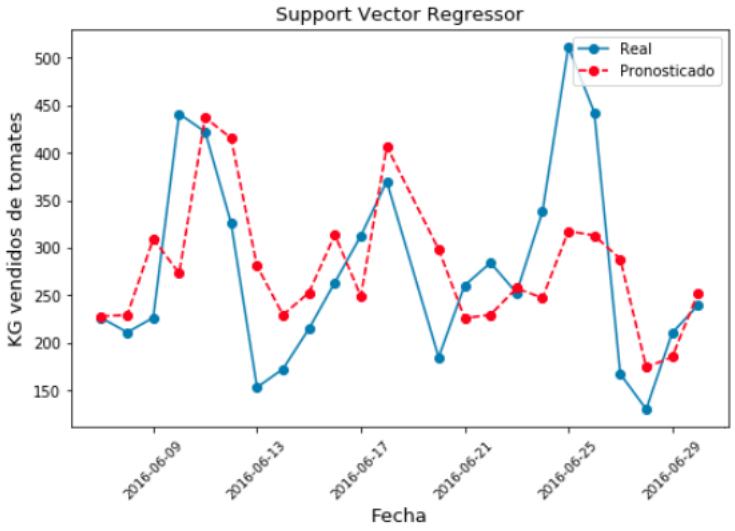


Figura 4.25: Resultados de la técnica de Support Vector Regressor Tomates

Finalmente, los resultados de robustez de este Support Vector Regressor se muestran en la Tabla 4.42.

Modelo	Set de entrenamiento		Set de testeo	
	MAPE	MAE	MAPE	MAE
Support Vector Regressor	26,69 %	68,51	29,49 %	87,77

Tabla 4.41: Resultados predicción Support Vector Regressor Tomates

Modelo	Set de entrenamiento				Set de testeo			
	mMAPE	sMAPE	mMAE	sMAE	mMAPE	sMAPE	mMAE	sMAE
SVR	20,27 %	5,39 %	66,37	11,91	36,08 %	10,09 %	77,35	15,85

Tabla 4.42: Resultados de robustez Support Vector Regressor Tomates

4.2.10. Redes Neuronales Artificiales

Para esta metodología se testean varias Redes Neuronales Artificiales con distinta cantidad de redes neuronales, distinta cantidad de capas ocultas y distintos *epochs* como se muestra en Apéndice D.6.

En este caso, la mejor Red Neuronal Artificial es aquella que posee 21 variables en la capa de entrada (las 21 variables independientes del problema), 4 capas ocultas con 12 neuronas cada una, 400 epochs para el aprendizaje y *una* capa de salida la cual representa la predicción de demanda, a diferencia del caso de la palta la cual poseía 3 capas ocultas con 10 neuronas cada una, 700 epochs y *una* capa de salida.

Los resultados de este método se muestran en la Figura 4.26 y en la Tabla 4.43.

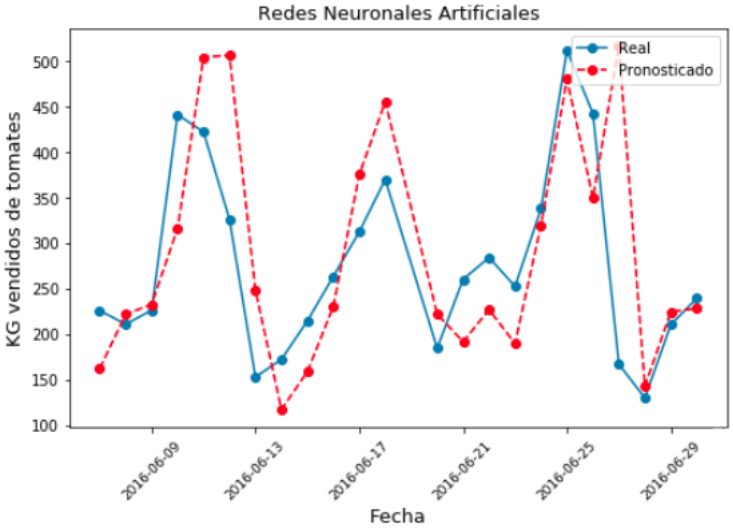


Figura 4.26: Resultados de la técnica de Redes Neuronales Artificiales Tomates

Finalmente, los resultados de robustez de esta Red Neuronal Artificial se muestran en la Tabla 4.44.

Modelo	Set de entrenamiento		Set de testeo	
	MAPE	MAE	MAPE	MAE
Redes Neuronales Artificiales	28,09 %	58,81	25,54 %	65,47

Tabla 4.43: Resultados predicción Redes Neuronales Artificiales Tomates

Modelo	Set de entrenamiento				Set de testeo			
	mMAPE	sMAPE	mMAE	sMAE	mMAPE	sMAPE	mMAE	sMAE
RNA	28,71 %	1,58 %	57,77	1,25	26,24 %	4,21 %	85,94	14,24

Tabla 4.44: Resultados de robustez Redes Neuronales Artificiales Tomates

4.2.11. Redes Neuronales Recurrentes

Para esta metodología, Redes Neuronales Recurrentes, RNR, se testean varias Redes Neuronales Recurrentes como se aprecia en Apéndice D.7. Finalmente, se utiliza aquella red neuronal recurrente que posee *una* capa de entrada (la variable a predecir), 3 capas ocultas con 50 neuronas cada una, 200 epochs y *una* capa de salida, a diferencia del caso de la palta, el cual posee solo 2 capas ocultas. Cabe destacar que además se utiliza un *Dropout* de 20 % para eliminar el 20 % más alto de neuronas correlacionadas en cada epoch.

Los resultados de esta metodología se muestran en la Figura 4.27 y en la Tabla 4.45.

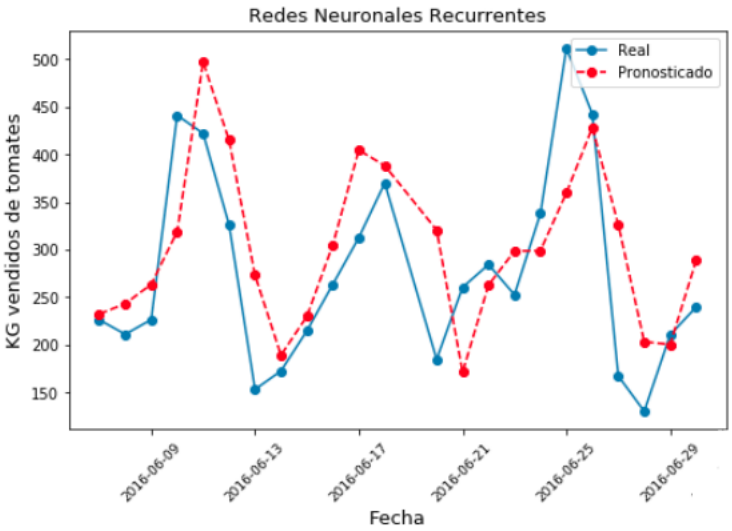


Figura 4.27: Resultados de la técnica Redes Neuronales Recurrentes Tomates

Finalmente, los resultados de robustez de esta Red Neuronal Recurrente se muestran en la Tabla 4.46.

Modelo	Set de entrenamiento		Set de testeo	
	MAPE	MAE	MAPE	MAE
Redes Neuronales Recurrentes	40,01 %	75,83	25,51 %	61,25

Tabla 4.45: Resultados predicción Redes Neuronales Recurrentes Tomates

Modelo	Set de entrenamiento				Set de testeo			
	mMAPE	sMAPE	mMAE	sMAE	mMAPE	sMAPE	mMAE	sMAE
RNR	44,85 %	2,68 %	81,03	2,59	22,68 %	5,85 %	78,04	27,62

Tabla 4.46: Resultados de robustez Redes Neuronales Recurrentes Tomates

Capítulo 5

Análisis de resultados

A continuación, se muestra una recapitulación de todos los resultados obtenidos de los modelos anteriormente testeados, de manera separada para paltas y tomates.

Se utiliza la métrica de MAPE para elegir el mejor modelo y se utiliza la métrica MAE cuando la decisión es difícil de tomar. Aquí MAPE representa el porcentaje de error en la predicción de la demanda de manera diaria y MAE representa el error absoluto de la venta de kilos diarios. Se prosigue a observar los valores obtenidos en el set de entrenamiento y se eliminan los modelos cuyos errores de pronóstico de demanda son los más altos. Luego, con los modelos restantes se analiza su robustez para ver si son efectivamente replicables; en caso de no serlo, los modelos son eliminados. Para el análisis de robustez se utilizan como indicadores el promedio y la desviación estándar del MAPE y del MAE en el set de entrenamiento y el set de testeo denotados como m y s, respectivamente. Es preciso señalar que se espera que el mejor modelo posea una desviación estándar baja[26], de esta manera, se puede decir que el modelo es robusto.

5.1. Análisis de resultados Palta Hass Extra a Granel

El total de los resultados obtenidos para la Palta Hass Extra a Granel son los que se muestran en la Tabla 5.1.

Lo que se busca a partir de la Tabla 5.1 es poder encontrar el o los mejores modelos que predicen la demanda de Palta Hass Extra a Granel. Para ello, se analiza en una primera instancia el valor que posee cada modelo en la métrica de promedio de MAPE m, en el set de entrenamiento; lo que se espera es que esta métrica sea baja. Se busca además un modelo cuyo error MAPE tanto en el set de entrenamiento como en el set de testeo no sean

[26]Que la desviación estándar sea baja significa que los resultados de predicción, en cualquiera de las base de datos testeadas, permanecen siempre cercanos a la media aritmética. Por el contrario, que la desviación estándar sea alta significa que estos resultados se extienden por un rango de valores más amplios y, por ende, las predicciones difieren notoriamente unas de otras. en este sentido, si la desviación estándar es alta, se dice que el modelo no es robusto ya que no es replicable.

Modelo	Set de entrenamiento						Set de testeo					
	MAPE	m[%]	s[%]	MAE	m	s	MAPE	m[%]	s[%]	MAE	m	s
NF	41,1 %	42,4	0.8	70,3	71,0	1,7	37,2 %	46,9	16,0	66,9	89,1	33,9
MA	36,5 %	37,0	0,4	63,3	63,0	1,2	23,3 %	28,0	8,5	37,2	58,1	19,3
ARI	54,3 %	57,1	2,0	88,3	91,3	2,0	26,8 %	35,1	10,0	43,6	70,7	28,4
SAR	29,3 %	29,9	0,5	53,2	53,5	1,3	24,3 %	38,5	15,8	43,6	77,0	36,0
SET	24,6 %	23,5	0,9	40,8	38,9	1,8	21,8 %	46,8	40,9	35,7	95,2	79,4
RLM	30,5 %	31,7	0,8	47,8	48,9	0,8	26,7 %	23,7	6,1	52,3	50,1	13,0
AR	19,7 %	19,9	1,1	31,6	30,8	0,8	20,3 %	28,6	10,8	39,3	62,6	35,4
RF	22,8 %	23,2	0,3	34,9	34,7	0,8	16,6 %	24,8	9,1	31,1	51,5	22,5
SVR	29,6 %	35,3	12,3	47,3	52,3	18,2	16,1 %	25,4	7,0	28,7	53,8	25,9
RNA	25,0 %	26,4	1,1	39,4	40,3	0,9	24,4 %	28,9	11,1	41,0	55,2	20,6
RNR	26,8 %	29,3	1,5	48,3	48,6	1,8	19,8 %	27,2	7,9	34,2	54,2	21,3

Tabla 5.1: Primera tabla comparativa de los modelos testeados Paltas

tan diferentes, sino que sean relativamente parecidos, de esta manera se puede decir que los modelos son consistentes.

A partir de la Tabla 5.1, se puede ver que los modelos que poseen, en promedio, el mayor error porcentual MAPE en el set de entrenamiento son el modelo de ARIMA con un error de $57,1\,\%$ en promedio, seguido por el modelo de Naïve Forecast con $42,4\,\%$ de error en promedio y Moving Average con $37,0\,\%$ de error en promedio. En consecuencia, se eliminan estos modelos del análisis de los mejores modelos ya que su predicción no es tan acertada como se le desea.

Bajo la misma lógica, se prosigue entonces a eliminar aquellos modelos cuyo error en el set de entrenamiento sigan siendo altos. Se elimina entonces el modelo de Support Vector Regressor ya que posee en promedio un error MAPE de $35,3\,\%$. Ahora, en vista de que muchos de los promedio MAPE en el set de entrenamiento no difieren en grandes cantidades entre los modelos que quedan, se decide utilizar el estadístico de desviación estándar tanto en MAPE como en MAE del set de entrenamiento y de testeo para eliminar a los modelos. En consecuencia, se elimina la metodología de SARIMA cuyo error porcentual promedio en el set de entrenamiento es de $29,9\,\%$ y cuya desviación estándar en el MAPE del set de testeo es de $15,8\,\%$. Siguiendo esta misma idea, se elimina el modelo de Suavización Exponencial Triple cuyo error porcentual promedio en el set de entrenamiento es de $23,5\,\%$ y cuya desviación estándar del MAPE en el set de testeo es de $40,9\,\%$, la desviación estándar más alta de todos los modelos testeados.

Modelo	Set de entrenamiento						Set de testeo					
	MAPE	m[%]	s[%]	MAE	m	s	MAPE	m[%]	s[%]	MAE	m	s
RLM	30,5 %	31,7	0,8	47,8	48,9	0,8	26,7 %	23,7	6,1	52,3	50,1	13,0
AR	19,7 %	19,9	1,1	31,6	30,8	0,8	20,3 %	28,6	10,8	39,3	62,6	35,4
RF	22,8 %	23,2	0,3	34,9	34,7	0,8	16,6 %	24,8	9,1	31,1	51,5	22,5
RNA	25,0 %	26,4	1,1	39,4	40,3	0,9	24,4 %	28,9	11,1	41,0	55,2	20,6
RNR	26,8 %	29,3	1,5	48,3	48,6	1,8	19,8 %	27,2	7,9	34,2	54,2	21,3

Tabla 5.2: Segunda tabla comparativa de los modelos testeados Paltas

Se realiza la Tabla 5.2 con el fin de que el análisis sea más sencillo al lector. Como uno de los objetivos principales es poder encontrar el o los mejores modelos que predigan la demanda de Palta Hass, se elimina el modelo de Regresión Lineal Múltiple ya que posee uno de los errores porcentuales en promedio más altos en el set de entrenamiento con $31,7\,\%$ de error, como se muestra en la Tabla 5.2. Si bien, este modelo es uno de los más robustos con $6,1\,\%$ de desviación estándar en el MAPE del set de testeo, no cumple con el objetivo de predecir con bajo porcentaje de error, por lo tanto, es eliminado. Se prosigue a eliminar el método de Redes Neuronales Recurrentes ya que posee el error porcentual en promedio más alto entre los cuatro modelos restantes con $29,3\,\%$ de error en el set de entrenamiento; otra razón para eliminarlo es que su error MAE también es el más alto con $48,3$ kilos de error en la predicción de palta. Otra razón para eliminarlo es que es uno de los modelos menos robustos en el set de entrenamiento, tanto en el MAPE como en el MAE.

De esta manera, en la Tabla 5.3 se muestran los modelos que van quedando.

Modelo	Set de entrenamiento						Set de testeo					
	MAPE	m[%]	s[%]	MAE	m	s	MAPE	m[%]	s[%]	MAE	m	s
AR	19,7 %	19,9	1,1	31,6	30,8	0,8	20,3 %	28,6	10,8	39,3	62,6	35,4
RF	22,8 %	23,2	0,3	34,9	34,7	0,8	16,6 %	24,8	9,1	31,1	51,5	22,5
RNA	25,0 %	26,4	1,1	39,4	40,3	0,9	24,4 %	28,9	11,1	41,0	55,2	20,6

Tabla 5.3: Tercera tabla comparativa de los mejores cuatro modelos para la Paltas Hass

A partir de este momento, se hace cada vez más difícil poder elegir cuál es el mejor modelo, por lo que a partir de ahora se provilegia por sobre todo la robustez de los modelos. En este sentido, se puede decir que el mejor modelo es el modelo de Random Forest, cuya desviación estándar del MAPE en el set de entrenamiento es de $0,3\,\%$. Si bien, este modelo no es el modelo cuya predicción es la más baja, es el modelo más robusto que se tiene. Además, la diferencia de errores es de aproximadamente 3 kilos con respecto a Árboles para Regresiones, modelo cuyo error es el más bajo en el set de entrenamiento; este error de 3 kilos es considerado despreciable comparado a las ventas promedio que se tiene de la palta, de $208,02$ kilos. Para explicar un poco más detalladamente esta parte, se decide no considerar el modelo de Redes Neuronales Ariticiales como el mejor modelo porque posee el error porcentual más grande en el set de entrenamiento con un $26,4\,\%$ de error porcentual en promedio y porque su desviación estándar en el set de entrenamiento, tanto en MAPE como MAE es la mayor, con un $1,1\,\%$ y 0,9 kilos, respectivamente. Por otra parte, si bien Árboles de Regresiones posee el menor MAPE y MAE en el set de entrenamiento, éste posee una alta desviación estándar en el MAPE tanto del set de entrenamiento como en el set de testeo y, como se dijo anteriormente, la diferencia de 3 kilos con respecto al modelo de Random Forest es casi despreciable (cercana al 1,5 % de las ventas promedio de paltas). Es por ello que los modelos de Redes Neuronales Artificiales y de Árboles de Regresiones son descartados, dejando al modelo de Random Forest como el mejor modelo para la predicción de Palta Hass Extra a Granel.

5.2. Análisis de resultados Tomate a Granel

Ahora, se analizan los resultados obtenidos para el Tomate a Granel. El total de estos resultados son los que se observan en la Tabla 5.4.

Modelo	Set de entrenamiento						Set de testeo					
	MAPE	m[%]	s[%]	MAE	m	s	MAPE	m[%]	s[%]	MAE	m	s
NF	49,9 %	52,6	2,3	120,5	120,2	2,3	44,5 %	40,4	22,9	94,5	127,5	77,0
MA	35,1 %	35,3	0,4	111,8	111,1	1,5	33,6 %	30,1	7,5	77,5	96,5	27,7
ARI	45,8 %	38,8	9,4	140,7	127,1	30,7	40,9 %	45,9	2,7	99,1	135,0	17,7
SAR	26,9 %	27,5	0,6	87,8	89,1	1,6	42,1 %	36,6	11,2	90,3	99,3	20,8
SET	28,9 %	29,0	0,4	70,1	67,6	1,9	46,1 %	54,4	48,4	95,5	172,1	167,4
RLM	31,3 %	32,9	1,4	73,1	73,1	0,8	27,6 %	21,7	5,5	72,4	73,2	17,2
AR	28,6 %	28,6	0,9	52,5	50,8	2,3	34,7 %	25,9	8,6	91,6	90,2	36,6
RF	22,4 %	23,5	1,2	47,9	47,2	0,4	28,9 %	21,7	5,6	66,3	70,3	18,7
SVR	26,7 %	20,3	5,4	68,5	66,4	11,9	29,5 %	36,1	10,1	87,8	77,3	15,8
RNA	28,1 %	28,7	1,6	58,8	57,8	1,3	25,5 %	26,2	4,2	65,5	85,9	14,2
RNR	40,0 %	44,8	2,7	75,8	81,0	2,6	25,5 %	22,7	49,5	49,5	78,0	27,6

Tabla 5.4: Primera tabla comparativa de los modelos testeados Tomates

De manera análoga al análisis de la Palta Hass, lo que se busca en este análisis es poder encontrar el o los mejores modelos que predicen la demanda del Tomate a Granel. Para ello, se analiza la métrica del promedio del MAPE en el set de entrenamiento de cada modelo y se eliminan aquellos modelos cuya métrica sea muy elevada. Además, se compara esta misma métrica en el set de testeo de manera de verificar que haya consistencia en las predicciones, o sea, que los resultados sean relativamente parecidos en el set de entrenamiento con el set de testeo. Se utiliza la métrica MAE en el caso que la decisión sea difícil de tomar. Se utiliza además la desviación estándar como apoyo en la decisión ya que se desea un modelo que pediga de mejor manera la demanda de Tomate a Granel pero, que además, sea robusto, o sea, que se pueda replicar.

A partir de la Tabla 5.4 se observa que los modelos con mayor porcentaje de error en promedio en el set de entrenamiento son Naïve Forecast con $52,6\,\%$ de error en promedio. Se prosigue a eliminar el modelo de Redes Neuronales Recurrentes ya que posee el segundo error porcentual en promedio más alto, con $44,8\,\%$ de error en promedio. Se elimina además la metodología de ARIMA, la cual posee un error porcentual promedio de $38,8\,\%$. Cabe destacar que este modelo posee una desviación estándar de $45,9\,\%$ en el set de testeo, siendo entonces uno de los modelos menos robustos.

Bajo el mismo razonamiento, se prosigue a eliminar aquellas metodologías cuyo error porcentual en promedio en el set de entrenamiento sigan siendo altos. En este sentido se elimina el modelo de Moving Average cuyo error porcentual en promedio es de $35,3\,\%$. Se prosigue además a eliminar el modelo de Regresión Lineal Múltiple el cual posee un error porcentual en promedio de $32,9\,\%$ en el set de entrenamiento.

Ahora, se continúa a eliminar el o los modelos que sean los menos robustos del set de modelos que van quedando. Para ello, se crea la Tabla 5.5, para hacer el análisis más sencillo al lector.

Modelo	Set de entrenamiento						Set de testeo					
	MAPE	m[%]	s[%]	MAE	m	s	MAPE	m[%]	s[%]	MAE	m	s
SAR	26,9 %	27,5	0,6	87,8	89,1	1,6	42,1 %	36,6	11,2	90,3	99,3	20,8
SET	28,9 %	29,0	0,4	70,1	67,6	1,9	46,1 %	54,4	48,4	95,5	172,1	167,4
AR	28,6 %	28,6	0,9	52,5	50,8	2,3	34,7 %	25,9	8,6	91,6	90,2	36,6
RF	22,4 %	23,5	1,2	47,9	47,2	0,4	28,9 %	21,7	5,6	66,3	70,3	18,7
SVR	26,7 %	20,3	5,4	68,5	66,4	11,9	29,5 %	36,1	10,1	87,8	77,3	15,8
RNA	28,1 %	28,7	1,6	58,8	57,8	1,3	25,5 %	26,2	4,2	65,5	85,9	14,2

Tabla 5.5: Segunda tabla comparativa de los modelos testeados Tomates

En la Tabla 5.5 se puede apreciar que uno de los modelos menos robusto es el modelo de Suavización Exponencial Triple. Si bien esta metodología posee una baja desviación estándar en el MAPE del set de entrenamiento, con $0,4\,\%$, posee una alta desviación estándar en el set de testeo, de $48,4\,\%$, lo que lo hace ser el modelo menos replicable de los once modelos; es por esto que esta metodología es eliminada del análisis. Bajo la misma lógica se prosigue a eliminar el modelo de Support Vector Regressor el cual posee una alta desviación estándar en el MAPE tanto en el set de entrenamiento como en el set de testeo, con $5,4\,\%$ y $10,1\,\%$, respectivamente. Finalmente, siguiendo la lógica utilizada en un principio, la de eliminar los modelos por sus errores en el set de entrenamiento, se elimina el modelo de SARIMA por poseer un elevado error en promedio en el set de entrenamiento, de $87,8$ kilos de tomate, y por poseer una elevada desviación estándar en el set de testeo de $11,2\,\%$.

En la Tabla 5.6 se muestran los modelos que quedan para hacer el análisis visualmente más fácil.

Modelo	Set de entrenamiento						Set de testeo					
	MAPE	m[%]	s[%]	MAE	m	s	MAPE	m[%]	s[%]	MAE	m	s
AR	28,6 %	28,6	0,9	52,5	50,8	2,3	34,7 %	25,9	8,6	91,6	90,2	36,6
RF	22,4 %	23,5	1,2	47,9	47,2	0,4	28,9 %	21,7	5,6	66,3	70,3	18,7
RNA	28,1 %	28,7	1,6	58,8	57,8	1,3	25,5 %	26,2	4,2	65,5	85,9	14,2

Tabla 5.6: Tercera tabla comparativa de los modelos testeados Tomates

A partir de la Tabla 5.6 se puede apreciar que, si bien la metodología de Árboles de Regresiones es la más robusta en el set de entrenamiento en términos de MAPE, ésta no lo es en el mismo set de entrenamiento pero en el MAE y no lo es en el set de testeo, tanto en MAPE como en MAE; aquí se aprecia que posee la mayor desviación estándar entre los tres modelos, con $8,6\,\%$ en MAPE y $36,6$ en MAPE. Es por esto que este modelo puede ser eliminado como uno de los mejores modelos.

Se prosigue a analizar los dos últimos modelos que quedan, Random Forest y Redes Neuronales Artificiales. A simple vista se puede apreciar que el modelo de Random Forest predice en promedio de mejor manera en el set de entrenamiento, tanto en MAPE como en MAE, sin embargo, en el set de testeo no siempre es así. No obstante, si se habla del error promedio, Random Forest predice en promedio de mejor manera en el set de testeo, tanto en MAPE como en MAE. A pesar de predecir de mejor manera, Random Forest no es el modelo más robusto en el set de testeo, sino que es Redes Neuronales Artificiales, el cual posee una desviación estándar de $4,2\,\%$ en el MAPE y de $14,2$ tomates en el MAE, siendo ambos valores

menores que en Random Forest. Si bien Redes Neuronales Artificiales es más robusto que Random Forest, se considera que, de todas formas, Random Forest predice de mejor manera; esto ya que en promedio Random Forest posee un error de $70,3$ kilos de tomates comparado a $85,9$ kilos en promedio en Redes Neuronales Artificiales.

A partir del análisis anteriormente descrito, se puede decir que ambos modelos son buenas metodologías para predecir la demanda de Tomate a Granel. Sin embargo, al momento de elegir uno, se elige el modelo de Random Forest ya que con este se posee un error promedio de predicción más bajo; existe una diferencia de alrededor de 15 kilos de tomates de predicción entre ambos modelos, o sea, una diferencia porcentual de alrededor de $4,5\,\%$ de error en promedio.

Capítulo 6

Conclusiones

A continuación, se muestran las conclusiones generales de los modelos de predicción de la demanda para la Palta Hass Extra a Granel y el Tomate a Granel seguido por recomendaciones a futuro.

6.1. Conclusiones generales

Como se muestra en un comienzo, es importante poder estimar la demanda de frutas y verduras de un supermercado para poder reducir las mermas, aumentar los ingresos y para, lo más importante, poder evitar el desperdicio de agua, tierra, energía, capital y trabajo necesario para producirlas. Dados estos problemas, se plantean las siguientes preguntas de investigación: ¿Cuál será la demanda de frutas y verduras dentro de un mes, sabiendo que éstas se deterioran fácilmente, su precio cambia constantemente y los clientes los compran según su apariencia estética? ¿Qué modelo de estimación se debería utilizar para estimar esta demanda con el menor error posible?

A lo largo del presente trabajo se aborda el tema de cuál es el mejor modelo para estimar la demanda de frutas y verduras. Para ello se escoge la fruta que más ingresos generaba, la Palta Hass Extra a Granel, en un cierto supermercado de Santiago, y se le aplican once modelos predictivos. Se realiza el mismo procedimiento, pero para la fruta más vendida en este mismo supermercado, el Tomate a Granel, para poder validar los once modelos.

Los datos con los que se cuentan pertenecen a una gran cadena de supermercados chilena la cual consta con 630 datos, de octubre 2014 a mayo 2016. En ella se encuentran las ventas de los productos, el precio y el stock. Dada la baja cantidad de variables que se poseen para estimar la demanda, se agregan otras como el aumento o descuento en el precio con respecto al día anterior, la temperatura promedio del día, el día de la semana, entre otras.

Se realizan entonces once modelos para la palta y el tomate y se comparan sus desempeños con las métricas MAPE y MAE, dándole prioridad a la métrica MAPE, que representa al error porcentual medio absoluto entre las ventas reales y el valor predicho para paltas y

tomates, respectivamente. Además, se analiza la robustez de los modelos para verificar que los modelos sean replicables. Para ello, se testean en cinco bases de datos más pequeñas los mismos once modelos y se analiza la desviación estándar y el promedio de los estadísticos MAPE y MAE. De esta manera, con los desempeños de cada metodología y con su robustez, se escogen los mejores modelos para la Palta Hass Extra a Granel y para el Tomate a Granel.

Se demuestra entonces que los modelos *más avanzados*, como se les llama en el Capítulo 2 de Marco teórico, predicen de mejor manera que aquellos modelos *tradicionales*. Esto se ve principalmente en que, en su mayoría, los modelos de *Naïve Forecasting*, *Moving Average*, *ARIMA* y *Regresión Lineal Múltiple* tuvieron un rendimiento más bajo comparado al resto de las metodologías.

Finalmente, se concluye que el mejor modelo para la Palta Hass Extra a Granel es el modelo de Random Forest cuyo error porcentual en promedio en el set de entrenamiento y set de testeo es de alrededor de 23 % y 25 %, respectivamente, teniendo un error absoluto medio de alrededor de 35 kilos de palta en el set de entrenamiento y de alrededor de 51 kilos de palta en el set de testeo. Como se expuso anteriormente, estos 51 kilos equivalen a un error de alrededor de 25 % de las ventas promedio de Palta Hass Extra a Granel.

Para el caso del Tomate a Granel, el mejor modelo también es Random Forest con un error porcentual de alrededor de 24 % y 22 % en el set de entrenamiento y en el set de testeo, respectivamente. Esto se traduce a un error absoluto medio de 47 kilos de tomate en promedio en el set de entrenamiento y de 70 kilos de tomate en promedio el set de testeo.

Si bien esto es una coincidencia, que el mejor modelo en ambas frutas sea Random Forest para Regresiones, esto puede no ser cierto para otras frutas o verduras. Esto se concluye principalmente por la cercanía que tienen los resultados a otros modelos como Árboles de Regresiones o Redes Neuronales Artificiales para el caso de la palta y Redes Neuronales Artificiales para el caso del tomate. Además, es esperable que no sean los mismos modelos ya que las ventas en kilo de frutas y verduras se comportan de manera distinta (algunos se venden más, otros menos), poseen distinta temporada de cosecha [19] y poseen distintos procesos de deterioro [14].

Por último, como se expuso anteriormente y, en el Capítulo 5, Análisis de resultados, Random Forest supera a los modelos de Suavización Exponencial Triple y Support Vector Regressor y se parece mucho a los resultados de Redes Neuronales Artificiales. El modelo de Random Forest es un modelo más simple de entrenar e implementar que los modelos anteriormente mencionados y, además, no se sobreajusta. Otra cualidad de esta metodología es que reduce la varianza de los resultados ya que es un modelo que calcula el promedio de muchos árboles de regresión simples y, por ende, se reduce la varianza. Uno de los mayores puntos a favor de este modelo es que se puede conocer la importancia de las variables que influyen en el comportamiento de compra de los consumidores y, por ende, se pueden tomar decisiones concretas por parte del supermercado para poder disminuir el error porcentual absoluto medio. En este sentido, las primeras cinco variables más importantes para el caso de la Palta Hass Extra a Granel son *Dia_FinDeSemana* con un $24,83\,\%$ seguido por el precio de la palta P con $18,24\,\%$, stock S con $16,54\,\%$, el precio del limón PL con $16,08\,\%$ y la temperatura promedio del día W con $9,31\,\%$; para el caso del Tomate a Granel, las primeras cinco variables son *Dia_FinDeSemana* con un $38,65\,\%$ seguido por el precio del tomate P

con $16,27\,\%$, stock S con $11,96\,\%$, la temperatura promedio del día W con $6,86\,\%$ y el precio de la papa en malla de 2 kg con $5,82\,\%$.

6.2. Recomendaciones a futuro

El objetivo de esta sección es servir de guía a futuros investigadores para que puedan contribuir a este trabajo, dándoles una instancia para conocer los puntos en los que se puede aportar o mejorar.

En primer lugar, como este trabajo se basa en el análisis de venta de frutas y verduras, hay que tener en cuenta que, en su mayoría, estos productos poseen temporalidad. En este sentido conviene tener una base de datos más grande para poder captar los ciclos que tienen las frutas y verduras, estos pueden ser ciclos de precio o de venta, por ejemplo. Al tener una mayor cantidad de datos, puede que modelos como el de Suavización Exponencial Triple hagan más sentido en el análisis de temporalidad. Otro punto a favor de poseer más datos es que modelos como Support Vector Regressor y Redes Neuronales Artificiales pueden mejorar. Ambos modelos son modelos que necesitan de una gran cantidad de datos para poder aprender de estos mismos y para poder hacer mejores predicciones. Si bien en este trabajo no había una gran cantidad de datos, un poco menos de 2 años de data, ambos modelos pudieron de todas formas predecir la demanda, dejando al segundo modelo, Redes Neuronales Artificiales, como el segundo mejor modelo en el caso de los tomates y tercero en el caso de las paltas.

Por otra parte, el poseer otras variables puede ser de gran importancia. En este trabajo sólo se poseían variables de kilos vendidos, stock y precio; otras variables fueron agregadas para suplir la falta de variables. Variables como la cantidad de espacio que utilizan las frutas y verduras en góndolas serían interesantes a analizar. Puede, en este sentido, que alimentos que son más visibles para los clientes sean más llamativos y, por ende, más comprados.

Finalmente, se pueden testear otros modelos como, por ejemplo, SARIMAX en donde se agregan las variables exógenas o variables independientes al modelo de SARIMA. También se pueden realizar modelos combinados, en donde se combinan los resultados de distintos modelos en distintas proporciones quedando, por ejemplo, $20\,\%$ del modelo Random Forest, $43\,\%$ de Redes Neuronales Articiales, etc.

Bibliografía

[1] L. Aburto and R. Weber. Demand Forecast in a Supermarket using a Hybrid Intelligent System. 2003.

[2] A. Azevedo and M.F. Santos. KDD, SEMMA and CRISP-DM: a parallel overview. 2008.

[3] I. Bhattacharyya. Support vector regression or svr, 2018.

[4] C. Brooks. *Introductory Econometrics for Finance*. Cambridge University Press, The Edinburgh Building, Cambridge, UK, 2008.

[5] R. Carbonneau, K. Laframboise, and R. Vahidov. Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research*, 184(3):1140–1154, 2008.

[6] P. Chapman, J.M. Clinton, R Kerber, T. Khabaza, T. Reinartz, C.R. Shearer, and R. Wirth. *CRISP-DM 1.0: Step-by-step data mining guide*. 2000.

[7] Cámara Nacional de Comercio. Informe comercio región metropolitana diciembre 2018. Índice de ventas reales supermercado tradicional. [En línea] `https://www.cnc.cl/wp-content/uploads/2019/01/Informe-Comercio-Región-Metropolitana-Diciembre-2018.pdf` [Consulta: 26/01/2019].

[8] Instituto Nacional de Estadísticas INE. Índice de ventas de supermercados - isup. [En línea] `https://www.ine.cl/estadisticas/economicas/comercio` [Consulta: 5/03/2019], 2018.

[9] Jefe de la Subdirección de Políticas y Apoyo en Materia de Publicación Electrónica. *Formulación y análisis detallado de proyectos*. Dirección del Centro de Inversiones Organización de las Naciones Unidas para la Agrucultura y la Alimentación, Roma, Italia, 2007.

[10] G Dellino, T. Laudadio, N. Mastronardi, and C. Meloni. Sales Forecasting Models in the Fresh Food Supply Chain. *International Conference on Operations Research and Enterprise Systems*, pages 419–426, 2015.

[11] G. Dorffner. Neural networks for time series processing. 1996.

[12] X. Fang Du, S.C.H. Leung, J. Long Zhang, and K.K. Lai. Demand forecasting of pe-

rishable farm products using support vector machine. *International Journal of System Sciences*, 44:556–567, 2011.

[13] J. Fattah, L. Ezzine, and Z. Aman. Forecasting of demand using ARIMA model. *International Journal of Engineering Business Management*, 10, 2018.

[14] Departamento de Protección Vegetal Fundación de investigacioón agrícola. Deterioro de las frutas y hortalizas frescas en el periodo de poscosecha. [En línea] `http://www.infoagro.com/frutas/deterioro_poscosecha_frutas_hortalizas.htm` [Consulta: 10/07/2019].

[15] R.J. Hyndman and G. Athanasopoulos. *Forecasting: Principles and Practice*. University of Western Australia, Australia, 2014.

[16] T. Kleynhans, A. Montaro, M. ans Gerace, and C. Kanan. Predicting Top-of Atmosphere Thermal Radiance Using MERRA-2 Atmospheric Data with Deep Learning. *Remote Sensing*, 9(11), 2017.

[17] ODEPA. *Los supermercados en la distribución alimentaria y su impacto sobre el sistema agroalimentario nacional*. Facultad de Ciencias y Pecuarías. Universidad de Chile, 2002.

[18] A. Orellana. *Árboles de decisión y Random Forest*. Universidad de Cuenca, Cuenca, 2018.

[19] R. Pizzarro and INDAP Ministerio de Agricultura. Prefiera frutas y verduras de la estación: Le hace bien a su bolsillo y a la tierra. [En línea] `http://www.indap.gob.cl/noticias/detalle/2016/10/19/prefiera-frutas-y-verduras-de-la-estación-le-hace-bien-a-su-bolsillo-y-a-la-tierra` [Consulta: 4/07/2019].

[20] D.E. Rumelhart, G.E. Hinton, and R.J. Williams. Learning internal representations by error propagation. *MIT Press*, 1:318–361, 1985.

[21] M. Shukla and S. Jharkharia. Applicability of ARIMA Models in Wholesale Vegetable Market: An Investigation. *Proceedings of the 2011 Internacional Conference on Industrial Engineering and Operations Management*, 2011.

[22] Q. Wen, L. Sun, S. Hua, and Z. Zhou. Daily Sales Forecasting for Grapes by Support Vector Machine. *International Federation for Information Processing*, 420:351–360, 2014.

[23] Y. Xiao and S. Yang. The Retail Chain Design for Perishable Food: The Case of Price Strategy and Shelf Space Allocation. *Sustainability*, 12(9), 2007.

[24] Tetsuo Iida; Paul Zipkin. Competition and Cooperation in a Two-Stage Supply Chain with Demand Forecasts. *Operations Research*, 58(5):1350–1363, 2010.

Apéndice A

Análisis de supermercados

A.1. Desgloce de la participación de productos perecibles en las ventas totales

Rubro Perecederos	Participación en ventas
Fiambrería	3.42 %
Carnicería	8.15 %
Vegetales	7.49 %
Lácteos	7.55 %
Panadería	5.23 %
Pescadería	1.69 %
Congelados	2.70 %

Tabla A.1: Desgloce de participación de productos perecibles en ventas de supermercados

La tabla obtenida anteriormente se obtiene a partir de los datos históricos que la alumna posee. Comparando estos resultados con los publicados por la Oficina de Estudios y Políticas Agrarias, ODEPA [17], se aprecia que los porcentajes de participación de venta de los productos perecibles se mantienen durante el tiempo.

A.2. Ingresos de supermercados de la cadena

Tabla A.2: Ingresos de los supermercados de la cadena

Supermercados	Ingresos [miles de millones de pesos]
Supermercado 1	184,6
Supermercado 2	166,1
Supermercado 3	132,7
Supermercado 4	115,2
Supermercado 5	115,2
Supermercado 6	100,1
Supermercado 7	97,2
Supermercado 8	95,8
Supermercado 9	88,1
Supermercado 10	88,0
Supermercado 11	83,4
Supermercado 12	78,6
Supermercado 13	78,5
Supermercado 14	70,8
Supermercado 15	70,3
Supermercado 16	61,3
Supermercado 17	56,4
Supermercado 18	55,4
Supermercado 19	54,8
Supermercado 20	53,7
Supermercado 21	52,2
Supermercado 22	51,1
Supermercado 23	50,8
Supermercado 24	50,6
Supermercado 25	48,2
Supermercado 26	48,2
Supermercado 27	43,2
Supermercado 28	43,0
Supermercado 29	42,9
Supermercado 30	41,8
Supermercado 31	40,6
Supermercado 32	39,9
Supermercado 33	36,9
Supermercado 34	36,8
Supermercado 35	32,4
Supermercado 36	32,3
Supermercado 37	30,8
Supermercado 38	30,7
Supermercado 39	12,7
Supermercado 40	10,1

Continúa en la página siguiente

Tabla A.2 – continúa de la página anterior

Supermercados	Ingresos [miles de millones de pesos]
Supermercado 41	7,4
Supermercado 42	6,2
Supermercado 43	6,1
Supermercado 44	5,7
Supermercado 45	5,4
Supermercado 46	5,1
Supermercado 47	4,7
Supermercado 48	3,2

Apéndice B

Notación

B.1. Notación a utilizar

La notación que se utiliza en este trabajo de título se muestra a continuación:

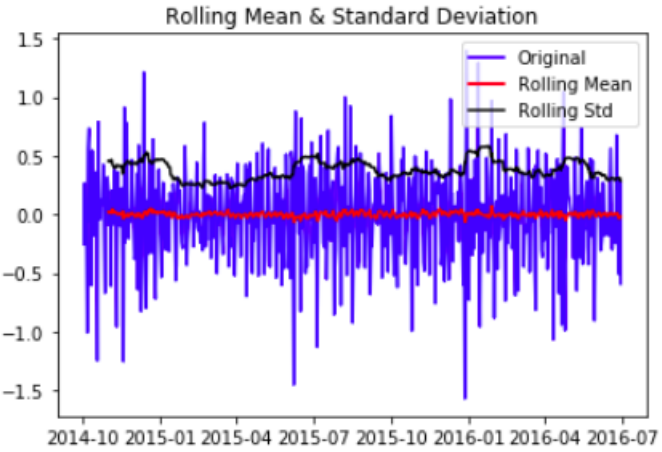
- $\hat{y}_{t+n}$: demanda a predecir en el período $t + n$
- y_t: demanda en el período t
- $\{x_\mathrm{i}\}_{\mathrm{i}=1}^{N}$: secuencia tal que i $\in \{1, ..., N\}$
- p: número de términos autorregresivos
- d: número de diferencias que se aplican a la serie de tiempo para que sea estacionaria
- q: número de medias móviles
- P: número de términos autorregresivos de la componente temporal del modelo SARIMA
- D: número de diferencias que se aplican a la serie de tiempo para que sea estacionaria de la componente temporal del modelo SARIMA
- Q: número de medias móviles de la componente temporal del modelo SARIMA
- s: periodicidad de la serie de tiempo
- δ: una constante
- y_{t-p}: las ventas en el período $t - p$
- ε_{t-p}: residuo del período $t - p$, el cual constituye el ruido blanco
- ϕ, θ: los coeficientes de los procesos autorregresivos y de media móvil, respectivamente
- A_t: atenuación de la serie de tiempo en Suavización Exponencial Triple
- L: largo de un ciclo de estacionalidad
- α: coeficiente de atenuación
- T_t: tendencia en el período t
- R_t: estacionalidad del período t
- β: coeficiente de tendencia
- γ: coeficiente de estacionalidad
- c: elasticidad del precio

- p: precio de alguún producto
- c: sensibilidad de demanda
- n_g: espacio ocupado por un producto en góndola
- d: sensibilidad de la demanda con respecto a la calidad de un producto
- $q(t)$: calidad del alimento perecible
- a_i: coeficientes de la regresión
- ε_i: lo que no se observa del modelo
- $\mathbf{x}$: vector de variables independientes
- $\mathbf{w}^T$: vector de los coeficientes encontrados en Support Vector Regressor
- ξ_i, ξ_i^*: representan la distancia que existe entre los puntos que se encuentran fuera del margen con el margen en Support Vector Regressor
- b: parámetro encontrado con la data en Support Vector Regressor
- C: parámetro que penaliza las observaciones fuera de los márgenes en Support Vector Regressor
- NF: para denotar a Naïve Forecast
- MA: para denotar a Moving Average
- ARI: para denotar a ARIMA
- SAR: para denotar a SARIMA
- SET: para denotar a Suavización Exponencial Triple
- RLM: para denotar a Regresión Lineal Múltiple
- AR: para denotar a Árboles de Regresiones
- RF: para denotar a Random Forest
- SVR: para denotar a Support Vector Regressor
- RNA: para denotar a Redes Neuronales Artificiales
- RNR: para denotar a Redes Neuronales Recurrentes

Apéndice C

Justificación de los modelos Paltas

C.1. ARIMA - Test de Dickey-Fuller




```
Results of Dickey-Fuller Test
Test Static                    -9.641337e+00
p-value                         1.518033e-16
#Lags Used                      1.800000e+01
Number of Observations Used     6.100000e+02
Critical value(1%)             -3.441116e+00
Critical value(5%)             -2.866290e+00
Critical value(10%)            -2.569300e+00
dtype: float64
```

Figura C.1: Test de Dickey-Fuller Paltas

C.2. Regresión Lineal Múltiple - Detalles

Cantidad de variables	Nombre variables	MAE		MAPE		Estadísticos de regresión	
		Train	Test	Train	Test	R^2	R^2 adj
46	Todas las variables	46,59	48,74	29,84 %	23,64 %	0,552	0,522
16	Lasso	47,70	51,77	30,24 %	25,65 %	0,524	0,511
15	PH · día_sábado	47,73	51,80	30,25 %	25,72 %	0,524	0,511
14	W · día_miércoles	47,64	52,39	30,24 %	26,03 %	0,523	0,512
13	P · día_miércoles	47,83	52,19	30,51 %	25,66 %	0,551	0,511

Tabla C.1: Procedimiento llevado a cabo para la elección de variables en la Regresión Lineal Múltiple

En la Tabla C.1 la columna *Nombre variables* se refiere a las variables que se eliminan excepto para la primera fila donde se muestran todos los resultados al testear todas las variables.

	coef	std err	t	P>|t|	[0.025	0.975]
const	-466.9255	298.027	-1.567	0.118	-1052.257	118.406
PT	-0.0598	0.012	-4.960	0.000	-0.083	-0.036
d	-0.0013	0.001	-2.516	0.012	-0.002	-0.000
b	-0.0003	8.47e-05	-3.967	0.000	-0.001	-0.000
PL	0.0342	0.009	3.623	0.000	0.016	0.053
Stock	0.0394	0.006	6.695	0.000	0.028	0.051
PP	0.0904	0.035	2.618	0.009	0.023	0.158
date_Friday	25.2596	8.437	2.994	0.003	8.689	41.830
Aumento_10	14.0964	6.267	2.249	0.025	1.787	26.406
date_Sunday	17.3333	10.043	1.726	0.085	-2.391	37.058
Dia_FinDeSemana	87.7032	8.043	10.904	0.000	71.906	103.500
Desc_mas_20	166.5081	28.646	5.813	0.000	110.247	222.769
PRECIO KG	0.2643	0.084	3.145	0.002	0.099	0.429
PM	0.7194	0.279	2.575	0.010	0.171	1.268

Omnibus:	50.811	Durbin-Watson:	1.003
Prob(Omnibus):	0.000	Jarque-Bera (JB):	172.020
Skew:	0.329	Prob(JB):	4.43e-38
Kurtosis:	5.539	Cond. No.	3.28e+08

Figura C.2: Resumen coeficientes de las variables en el set de entrenamiento. Regresión Lineal Múltiple

En la Figura C.2 las variables d y b representan interacciones entre las variables donde d $= W \cdot PH$ y $b = PrecioKg \cdot PH$. En esta figura se muestran los resultados obtenidos con 13 variables.

C.3. Árbol de Regresión - Elección del mejor árbol

A continuación se muestra una tabla comparativa de distintos árboles con distintas profundidades. Se sigue por mostrar gráficamente el árbol escogido y las variables que lo componen (las variables importantes del árbol escogido).

Profundidad	MAE		MAPE	
	Train	Test	Train	Test
3	49,13	34,19	31,21 %	19,18 %
4	45,81	35,31	29,30 %	19,39 %
5	40,87	38,97	26,06 %	19,34 %
6	36,62	37,45	23,36 %	18,60 %
7	31,57	39,27	19,71 %	20,26 %
8	26,19	39,09	15,73 %	10,79 %
28	0,00	42,18	0,00 %	24,46 %

Tabla C.2: Tabla comparativa para árboles de regresión con distinta cantidad de ramas

De la Tabla C.2 se elige que el mejor modelo es el que tiene una profundidad de 7 ramas ya que posee un MAPE estable entre el set de entrenamiento y el set de testeo de $19,71\,\%$ y $20,26\,\%$, respectivamente.

Figura C.3: Gráfico del árbol de 7 ramas

Este árbol, al poseer 7 ramas de profundidad es de difícil interpretación. Es por ello que se hace un desgloce de la Figura C.3 en las Figuras C.4, C.5 y C.6, las cuales corresponden al lado derecho, el centro y el lado izquierdo del árbol, respectivamente.

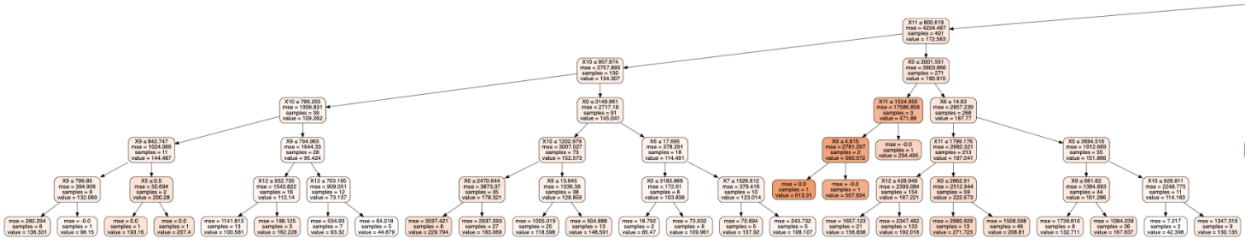


Figura C.4: Gráfico izquierda del árbol de 7 ramas

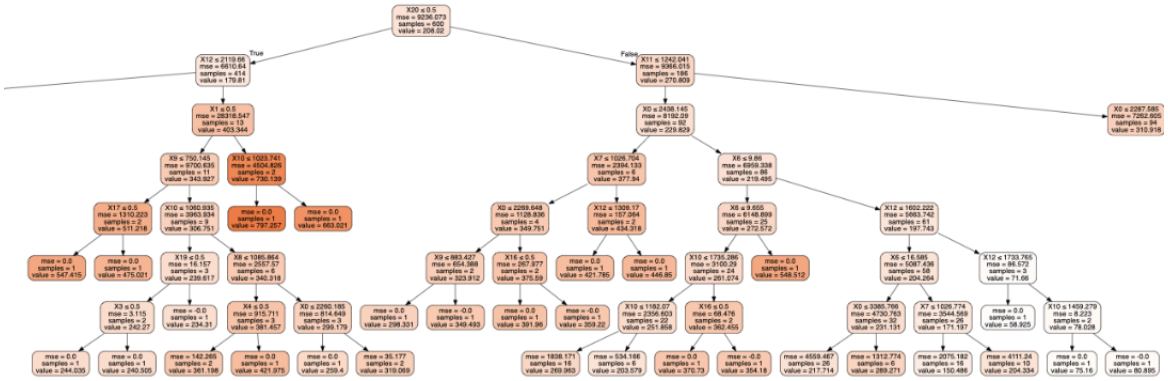


Figura C.5: Gráfico centro del árbol de 7 ramas

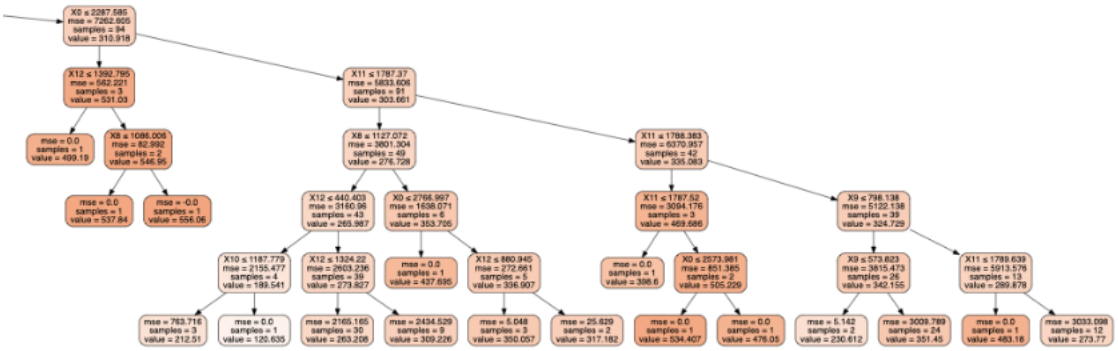


Figura C.6: Gráfico derecha del árbol de 7 ramas

Si bien en las figuras no se aprecia con claridad, el detalle de las variables importantes se muestran a continuación:

- $X0 = PrecioKg$
- $X1 = Desc\_mas\_20$
- $X5 = A\_10\_20$
- $X6 = W$
- $X7 = PH$
- $X8 = PM$
- $X9 = PP$
- $X10 = PT$
- $X11 = PL$
- $X17 = Dia\_Jueves$
- $X20 = Dia\_FinDeSemana$

C.4. Random Forest - Elección del mejor Random Forest

A coninuación, se muestra la tabla comparativa para distintos Random Forest.

Profundidad	Número de árboles	MAE		MAPE	
		Train	Test	Train	Test
5	10	40,10	30,46	25,98 %	16,56 %
5	90	39,01	32,12	25,71 %	16,83 %
5	1000	38,84	32,14	25,59 %	16,86 %
6	50	34,93	31,13	22,84 %	16,58 %
6	60	35,12	34,19	23,04 %	17,02 %
6	75	34,83	31,25	23,02 %	16,80 %
6	85	34,81	31,26	22,95 %	16,67 %
6	90	34,92	31,53	23,03 %	16,75 %
6	500	34,90	31,54	22,97 %	16,65 %

Tabla C.3: Tabla comparativa para Random Forest con distintas profundidades y cantidades de árboles

De la Tabla C.3 se elige el Random Forest cuya profundidad es de 6 ramas y cuya cantidad de árboles es de 50 como se muestra en la cuarta fila de resultados.

C.5. Random Forest - Importancia de las variables

Nombre Variable	Importancia
Dia_FinDeSemana	24,83 %
P	18,24 %
S	16,54 %
PL	16,08 %
W	9,31 %
PT	3,65 %
PP	3,44 %

Tabla C.4: Importancia de las variables modelo escogido Random Forest Paltas

En la Tabla C.4 se muestra un extracto de la importancia de las primeras siete variables del modelo escogido de Random Forest.

C.6. Redes Neuronales Artificiales - Resultados de distintas RNA

Cantidad de capas ocultas	Cantidad de neuronas	Epochs	MAE		MAPE	
			Train	Test	Train	Test
2	10	200	46,43	48,21	29,29 %	26,65 %
2	12	300	44,23	46,08	30,36 %	27,40 %
3	12	300	42,12	43,47	26,81 %	21,96 %
3	10	300	40,86	41,55	29,46 %	27,00 %
3	10	400	45,53	46,27	27,44 %	27,15 %
3	12	400	47,01	45,47	29,34 %	25,24 %
3	10	700	40,37	39,42	25,91 %	21,59 %
4	10	400	46,05	37,85	28,01 %	24,40 %
4	12	400	45,86	39,72	27,19 %	26,62 %

Tabla C.5: Tabla comparativa para Redes Neuronales Artificiales con distintas cantidades de capas, neuronas y epochs

En la Tabla C.5 se elige el mejor modelo según la métrica MAPE. En este caso, la mejor Red Neuronal Artificial es aquella que posee 3 capas ocultas con 10 neuronas cada una y 700 epochs de aprendizaje.

C.7. Redes Neuronales Recurrentes - Resultados de distintas RNR

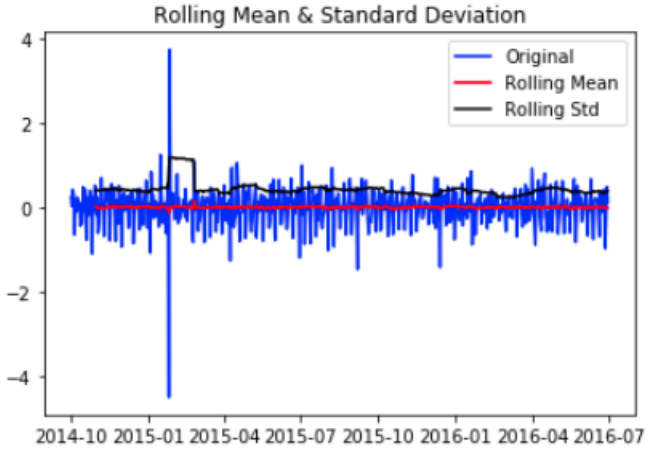
Cantidad de capas ocultas	Cantidad de neuronas	Epochs	Set de entrenamiento		Set de testeo	
			MAPE	MAE	MAPE	MAE
2	40	200	29,96 %	49,37	24,00 %	39,57
2	45	200	28,53 %	48,70	21,97 %	36,62
2	50	200	26,83 %	48,34	19,81 %	34,23
3	45	100	30,27 %	50,29	21,82 %	35,50
3	50	100	27,94 %	49,15	19,48 %	33,06
3	50	200	30,01 %	49,26	24,30 %	40,25
3	50	100	28,64 %	49,33	20,53 %	34,18
4	45	200	33,59 %	55,37	21,89 %	34,98
4	50	100	34,23 %	53,63	25,83 %	41,64

Tabla C.6: Tabla comparativa para Redes Neuronales Recurrentes con distintas cantidades de capas, neuronas y epochs

Apéndice D

Justificación de los modelos Tomates

D.1. ARIMA - Test de Dickey-Fuller




```
Results of Dickey-Fuller Test
Test Static                    -9.504983e+00
p-value                         3.366099e-16
#Lags Used                      1.900000e+01
Number of Observations Used     6.090000e+02
Critical value(1%)             -3.441133e+00
Critical value(5%)             -2.866298e+00
Critical value(10%)            -2.569304e+00
dtype: float64
```

Figura D.1: Test de Dickey-Fuller Tomates

D.2. Regresión Lineal Múltiple - Detalles Tomates

Cantidad de variables	Nombre variables	MAE		MAPE		Estadísticos de regresión	
		Train	Test	Train	Test	R^2	R^2 adj
45	Todas las variables	70,56	59,99	30,99 %	23,35 %	0,575	0,543
15	Lasso	72,95	66,21	31,31 %	24,81 %	0,545	0,534
14	W · día_sábado	73,05	65,20	31,27 %	24,58 %	0,545	0,534
13	W · día_domingo	73,10	65,47	31,28 %	24,63 %	0,545	0,534
12	W	73,10	64,66	31,19 %	24,39 %	0,545	0,534
11	A_10_20	73,11	65,44	31,31 %	24,59 %	0,544	0,534

Tabla D.1: Procedimiento llevado a cabo para la elección de variables en la Regresión Lineal Múltiple

En la Tabla D.1, al igual que en la Tabla C.1, la columna *Nombre variables* se refiere a las variables que se eliminan excepto para la primera fila donde se muestran todos los resultados al testear todas las variables.

	coef	std err	t	P>|t|	[0.025	0.975]
const	551.0025	60.452	9.115	0.000	432.274	669.731
Desc_20_10	-35.8885	18.658	-1.923	0.055	-72.534	0.757
m2	-5.8399	1.088	-5.368	0.000	-7.976	-3.703
mi2	-4.9349	1.089	-4.533	0.000	-7.073	-2.797
j2	-3.9527	1.333	-2.965	0.003	-6.571	-1.335
l2	-3.5134	1.067	-3.293	0.001	-5.609	-1.418
PRECIO KG	-0.1880	0.019	-9.850	0.000	-0.226	-0.151
Ppalta	-0.0314	0.013	-2.396	0.017	-0.057	-0.006
Ppapa	0.0178	0.009	1.949	0.052	-0.000	0.036
Stock	0.0285	0.004	6.695	0.000	0.020	0.037
Desc_mas_20	117.1646	30.130	3.889	0.000	57.990	176.339
Dia_FinDeSemana	159.1513	11.563	13.764	0.000	136.441	181.861

Omnibus:	253.864	Durbin-Watson:	1.081
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1971.620
Skew:	1.683	Prob(JB):	0.00
Kurtosis:	11.218	Cond. No.	5.84e+04

Figura D.2: Resumen coeficientes de las variables en el set de entrenamiento. Regresión Lineal Múltiple

En la Figura D.2 las variables $m2$, $mi2$, $j2$ y $l2$ representan interacciones entre las variables donde $m2 = W \cdot \text{dia\_}miercoles$, $mi2 = W \cdot \text{dia\_}miercoles$, $j2 = W \cdot \text{dia\_}jueves$ y $l2 = W \cdot \text{dia\_}lunes$. En esta figura se muestran los resultados obtenidos con 11 variables.

D.3. Árbol de Regresión - Elección del mejor árbol

A continuación se muestra una tabla comparativa de distintos árboles con distintas profundidades testeados para el tomate. Se sigue por mostrar gráficamente el árbol escogido y las variables que lo componen (las variables importantes del árbol escogido).

Profundidad	MAE		MAPE	
	Train	Test	Train	Test
3	77,50	67,47	35,47 %	31,06 %
4	72,31	62,66	33,63 %	27,33 %
5	62,73	97,34	30,93 %	38,55 %
6	52,47	90,78	28,57 %	34,45 %
7	44,67	91,42	25,82 %	34,91 %
8	38,79	103,34	19,14 %	39,95 %
28	0,00	92,78	0,00 %	35,88 %

Tabla D.2: Tabla comparativa para árboles de regresión con distinta cantidad de ramas Tomates

De la Tabla D.2 se obtiene que el mejor modelo es aquel que tiene una profundidad de 6 ramas ya que posee la menor diferencia entre MAPEs entre el set de entrenamiento y el set de testeo.

Figura D.3: Gráfico del árbol de 6 ramas

Este árbol, al poseer 6 ramas de profundidad es de difícil interpretación. Es por ello, al igual que para el caso de la palta, que se hace un desgloce de la Figura D.3 en las Figuras D.4 y D.5.

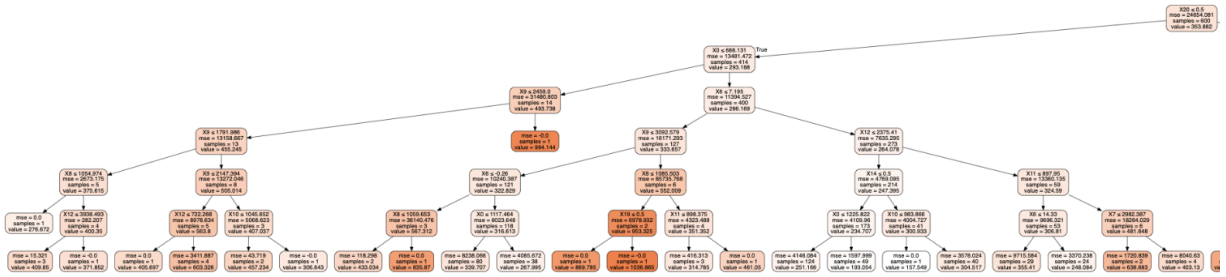


Figura D.4: Gráfico izquierda del árbol de 6 ramas

Si bien en las figuras no se aprecia con claridad, el detalle de las variables importantes se muestran en la lista punteada.

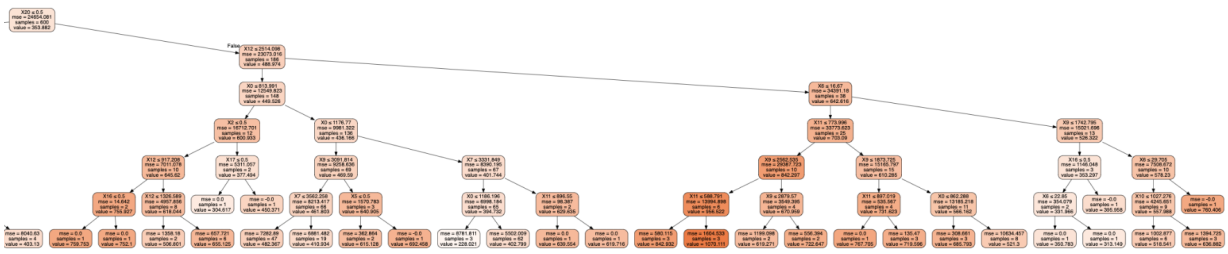


Figura D.5: Gráfico derecha del árbol de 6 ramas

- $X0 = PrecioKg$
- $X5 = A\_10\_20$
- $X6 = W$
- $X7 = Ppalta$
- $X8 = PM$
- $X9 = Ppapa$
- $X10 = PH$
- $X11 = Ppeg$
- $X12 = S$
- $X14 = Dia\_Viernes$
- $X16 = Dia\_Domingo$
- $X17 = Dia\_Jueves$
- $X19 = Dia\_Miercoles$
- $X20 = Dia\_FinDeSemana$

D.4. Random Forest - Elección del mejor Random Forest

Profundidad	Número de árboles	MAE		MAPE	
		Train	Test	Train	Test
6	50	53,01	67,16	25,66 %	29,39 %
6	85	52,70	67,88	25,88 %	29,64 %
7	20	48,72	59,99	22,74 %	27,07 %
7	40	48,04	65,48	22,65 %	28,54 %
7	60	47,92	66,33	22,38 %	28,86 %
7	75	47,74	66,52	22,71 %	29,03 %
7	90	47,78	68,89	22,64 %	29,79 %
7	100	47,65	68,42	22,75 %	29,63 %
7	1000	47,80	66,83	23,03 %	29,38 %

Tabla D.3: Tabla comparativa para Random Forest con distintas profundidades y cantidades de árboles

De la Tabla C.3 se elige el Random Forest cuya profundidad es de 7 ramas y cuya cantidad de árboles es de 60 como se muestra en la cuarta fila de resultados.

D.5. Random Forest - Importancia de las variables

A continuación se muestra la importancia del Random Forest ecogido, aquel que posee 7 ramas y 60 árboles.

Nombre Variable	Importancia
Dia_FinDeSemana	38,65 %
P	16,17 %
S	11,96 %
W	6,86 %
Ppapa	5,82 %
Pplatano	5,31 %
Ppalta	3,87 %

Tabla D.4: Importancia de las variables modelo escogido Random Forest Tomates

D.6. Redes Neuronales Artificiales - Resultados de distintas RNA

A conitnuación se muestra una tabla comparativa de Redes Neuronales Artificiales para el caso del tomate.

Cantidad de capas ocultas	Cantidad de neuronas	Epochs	MAE		MAPE	
			Train	Test	Train	Test
2	10	200	68,57	62,86	30,50 %	21,48 %
2	12	300	66,26	70,41	30,32 %	26,12 %
3	12	300	62,87	64.71	28,61 %	21,96 %
3	10	300	62,23	72,98	29,07 %	28,92 %
3	10	400	64,59	61,71	28,28 %	24,65 %
3	12	400	61,50	74,86	28,74 %	29,85 %
3	10	700	60,98	67,94	28,96 %	34,63 %
4	10	400	61,09	66,33	28,64 %	28,00 %
4	12	400	58,81	65,47	28,09 %	25,54 %

Tabla D.5: Tabla comparativa para Redes Neuronales Artificiales con distintas cantidades de capas, neuronas y epochs

A partir de la Tabla D.5 se elige la Red Neuronal Artificial que posee 4 capas ocultas con 12 neuronas cada una y 400 epochs de aprendizaje. Se elige esta red neuronal ya que es aquella que posee menor MAPE en el set de entrenamiento.

D.7. Redes Neuronales Recurrentes - Resultados de distintas RNR

Cantidad de capas ocultas	Cantidad de neuronas	Epochs	Set de entrenamiento		Set de testeo	
			MAPE	MAE	MAPE	MAE
2	40	200	40,66 %	81,74	25,71 %	61,07
2	45	200	40,29 %	81,56	26,24 %	61,92
2	50	200	40,48 %	81,91	26,39 %	61,73
3	40	100	48,90 %	96,71	39,03 %	84,46
3	48	100	42,61 %	87,74	32,46 %	72,61
3	48	200	40,18 %	80,79	25,59 %	61,96
3	50	100	43,74 %	92,51	36,23 %	79,76
3	50	200	40,01 %	75,83	25,51 %	61,25
3	55	100	40,83 %	81,47	25,42 %	60,76
4	50	100	49,12 %	105,15	45,09 %	96,06

Tabla D.6: Tabla comparativa para Redes Neuronales Recurrentes con distintas cantidades de capas, neuronas y epochs