



**Escuela Superior  
de Ingeniería y Tecnología**  
Universidad de La Laguna

# Trabajo de Fin de Grado

Grado en Ingeniería Informática

## Estimación de las ventas de frutos secos. Un caso de estudio: La Gaviota Alimentación

*Estimating sales of dried fruits. A case of study: La Gaviota  
Alimentación*

Ángel Emilio Capote Pérez

La Laguna, 14 de junio de 2022

D. **J. Marcos Moreno Vega**, con N.I.F. 42.841.047-M, Catedrático de Universidad adscrito al Departamento de Ingeniería Informática y de Sistemas de la Universidad de La Laguna, como tutor

D. **Francisco Javier Pérez Alonso**, con N.I.F. 78.619.759-D, Adjunto a la dirección de La Gaviota Alimentación S.L., como cotutor

## **C E R T I F I C A ( N )**

Que la presente memoria titulada:

*“Estimación de las ventas de frutos secos. Un caso de estudio: La Gaviota Alimentación”*

ha sido realizada bajo su dirección por D. **Ángel Emilio Capote Pérez**, con N.I.F. 45.353.412-B.

Y para que así conste, en cumplimiento de la legislación vigente y a los efectos oportunos firman la presente en La Laguna a 27 de mayo de 2022

78619759D  
FRANCISCO  
JAVIER PÉREZ  
(R: B38426664)

Firmado digitalmente  
por 78619759D  
FRANCISCO JAVIER  
PÉREZ (R: B38426664)  
Fecha: 2022.06.13  
22:35:33 +01'00'

## **Agradecimientos**

En primer lugar, quiero agradecer a todas aquellas personas, compañeros, amigos y familiares que me han acompañado a lo largo de esta etapa, sin su apoyo no habría sido posible.

También agradecer a Francisco Pérez de La Gaviota Alimentación por hacer posible este proyecto.

Hacer especial mención a Marcos Moreno, por todos los conocimientos, tiempo y dedicación que ha puesto en este trabajo.

# Licencia



© Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial 4.0 Internacional.

## Resumen

En este proyecto se lleva a cabo el estudio y la estimación de las ventas de cuatro productos facilitados por la empresa La Gaviota Alimentación. Para ello se han abordado dos paradigmas de predicción, el primero que llamaremos clásico, en el que los datos son presentados como series temporales. El segundo, recoge las técnicas de machine learning. El objetivo es seleccionar el modelo que realice una mejor predicción.

El primer paso ha sido estudiar los datos, en busca de estacionalidad, tendencia y valores anómalos. Una vez identificadas todas las características, se ha estudiado la literatura disponible en busca de modelos aplicables. Seleccionados los modelos, se han aplicado y evaluado los resultados para responder a la pregunta: ¿Cuál de los modelos realiza mejores predicciones?

Finalmente, a partir de la información recopilada en el estudio, se ha construido un prototipo de estimación de ventas, que cuenta con una función de validación y otra de predicción usando los modelos evaluados.

**Palabras clave:** Series temporales, estimaciones de ventas, aprendizaje automático.

## **Abstract**

In this project, the study and estimation of the sales of four products provided by the company La Gaviota Alimentación is carried out. For this, two prediction paradigms have been addressed, the first one that we will call classic, in which the data is presented as time series, the second one, includes machine learning techniques. The goal is to select the model that performs the best prediction.

The first step has been to study the data, looking for seasonality, trend and anomalous values, once all the characteristics have been identified, the available literature has been studied in search of applicable models. Once the models have been selected, the results have been applied and evaluated to answer the question: Which of the models makes better predictions?

Finally, based on the information collected in the study, a sales estimation prototype has been built, which has a validation function and a prediction function using the evaluated models.

**Keywords:** Sales estimations, time series, machine learning.

# Índice general

<b>Introducción</b>	<b>1</b>
Antecedentes y estado del arte	1
La Gaviota Alimentación S.L.	1
El Modelo de Predicción	1
Estado del arte	2
Objetivos	2
Fases	3
<b>Los datos y su tratamiento</b>	<b>4</b>
Los artículos	4
<b>Los modelos de predicción</b>	<b>10</b>
Autoregresión	10
SARIMA	10
Prophet	11
Random Forest	11
<b>Resultados obtenidos</b>	<b>13</b>
Artículo 1	13
Artículo 2	15
Artículo 3	17
Artículo 4	19
<b>Prototipo sistema de predicción de ventas</b>	<b>22</b>
<b>Conclusiones y líneas futuras</b>	<b>25</b>
<b>Summary and Conclusions</b>	<b>26</b>
<b>Tecnologías</b>	<b>27</b>
<b>Presupuesto</b>	<b>28</b>
<b>Apéndice 1: Datos de las ventas</b>	<b>29</b>

<b>Apéndice 2: Tablas con la estimación realizada</b>	<b>31</b>
Artículo 1	31
Artículo 2	32
Artículo 3	34
Artículo 4	35

# Índice de figuras

Figura 1.1: Artículo 1, ventas con covid	5
Figura 1.2: Artículo 1, ventas sin covid	5
Figura 1.3: Artículo 2, ventas con covid	6
Figura 1.4: Artículo 2, ventas sin covid	6
Figura 1.5: Artículo 3, ventas con covid	7
Figura 1.6: Artículo 3, ventas sin covid	7
Figura 1.7: Artículo 4, ventas con covid	8
Figura 1.8: Artículo 4, ventas sin covid	8
Figura 2.1: Funcionamiento Random Forest	12
Figura 3.1: Diagrama de clases del prototipo	22
Figura 3.2: Formato fichero validación random forest	23
Figura 3.3: Demo iniciación programa	23
Figura 3.4: Menú de opciones	23
Figura 3.5: Resultado validaciones	24
Figura 3.6: Resultado predicción SARIMA	24

# Índice de tablas

Tabla 2.4: Artículo 1, Estimación mensual	15
Tabla 2.8: Artículo 2, Estimación mensual	17
Tabla 2.12: Artículo 3, Estimación mensual	19
Tabla 2.16: Artículo 4, Estimación mensual	21
Tabla 1.1: Artículo 1, ventas	29
Tabla 1.2: Artículo 2, ventas	29
Tabla 1.3: Artículo 3, ventas	29
Tabla 1.4: Artículo 4, ventas	30
Tabla 2.1: Artículo 1, Estimación anual	31
Tabla 2.2: Artículo 1, Estimación semestral	31
Tabla 2.3: Artículo 1, Estimación trimestral	32
Tabla 2.5: Artículo 2, Estimación anual	32
Tabla 2.6: Artículo 2, Estimación semestral	33
Tabla 2.7: Artículo 2, Estimación trimestral	33
Tabla 2.9: Artículo 3, Estimación anual	34
Tabla 2.10: Artículo 3, Estimación semestral	34
Tabla 2.11: Artículo 3, Estimación trimestral	35
Tabla 2.13: Artículo 4, Estimación anual	35
Tabla 2.14: Artículo 4, Estimación semestral	36
Tabla 2.15: Artículo 4, Estimación trimestral	36

# 1. Introducción

De la estimación de ventas dependen diversos factores que pueden ser cruciales para la empresa, tales como el aprovisionamiento de materias primas, la planificación de tareas, el logro de objetivos y la capacidad de respuesta. Su correcta predicción supone un reto para las empresas, ya que se ve influenciada por factores como las preferencias de los clientes, el precio de venta, la estación del año o la ocurrencia de eventos impredecibles. Los factores más influyentes a la hora de hacer la estimación no suelen ser controlables por las empresas, lo que dificulta de manera sustancial dar una estimación adecuada.

La presente memoria se estructura en 6 secciones. En la primera sección se enumeran los antecedentes, incluyendo lo relativo a la empresa y a su modelo de predicción actual. También se hace una breve revisión bibliográfica de los trabajos científicos directamente relacionados con los objetivos de este proyecto para establecer el estado del arte. Por último, en esta sección se enumeran los objetivos generales y específicos del trabajo. En la segunda sección, se describen los datos facilitados por la empresa La Gaviota Alimentación, analizando su naturaleza y describiendo el preprocesado que se ha hecho de ellos. En la tercera sección, se tratan los modelos de predicción empleados en el trabajo, brindando una descripción de cada uno de ellos. En la cuarta sección, se exponen, analizan y comparan los resultados obtenidos. En la quinta sección, se aborda la estructura y funcionamiento del prototipo de estimación de ventas. Por último, en la sexta sección, se exponen las conclusiones del trabajo, así como una línea futura.

## 1.1 Antecedentes y estado del arte

### 1.1.1 La Gaviota Alimentación S.L.

La Gaviota Alimentación es una empresa familiar, fundada en 1982, referente principal en la comercialización de frutos secos en Canarias. Cuenta con un catálogo formado por más de 400 frutos secos y diversos cereales, semillas y especias. Tiene presencia en todas las islas y entre sus principales clientes se encuentran los principales operadores de: hostelería, distribución, pastelería e impulso de Canarias.

Para mantener esta posición de referente es fundamental para la empresa realizar una estimación adecuada de las ventas. Esto podría evitar pérdidas por situaciones de desabastecimiento, que ocasiona ventas perdidas, y también por sobreabastecimiento, que conlleva costes derivados de almacenamiento, entre otros.

### 1.1.2 El Modelo de Predicción

Actualmente, La Gaviota Alimentación realiza una previsión simple de las ventas del siguiente mes, basada en la media móvil de los últimos tres meses que, evidentemente, no tiene en cuenta la estacionalidad de las ventas. La observación de los datos históricos de ventas de la compañía sugiere que existe cierta estacionalidad en las ventas de algunos de los datos facilitados. Por ello, se establece como hipótesis de partida del presente trabajo que la inclusión de este comportamiento en los modelos debería mejorar las previsiones que, en estos momentos, hace la compañía.

### 1.1.3 Estado del arte

Tradicionalmente se han utilizado modelos de macroeconomía, basados en series temporales como los autorregresivos, media móvil, ARIMA, SARIMA (Seasonal autoregressive integrated moving average). Actualmente, además de los mencionados anteriormente, se están utilizando modelos cuya base nace en el aprendizaje automático (Machine learning). Deben destacarse, entre otros, los árboles de clasificación, el Random Forest y las Redes Neuronales, que han mostrado un buen desempeño en la predicción de ventas.

A continuación se enumeran y resumen varios de los trabajos prácticos consultados para la realización de este proyecto:

#### **Forecasting Wars: Classical Forecasting Methods vs Machine Learning [5]**

En este artículo se aborda una comparativa entre los métodos más clásicos, aquellos basados en series temporales, y algunos algoritmos de machine learning, usando los datos de ventas de algunos productos de Walmart. El objetivo es hacer la predicción de las ventas de los próximos 28 días. Para ello se construye un modelo preprocesando todos los datos para entrenar y predecir con los algoritmos de ML. Tras llevar a cabo la predicción con diversos métodos se llega a la siguiente conclusión:

Han tenido un mejor desempeño los algoritmos de machine learning, pero esto no significa que se pueda aplicar como norma general, ya que depende mucho del problema. Además, destaca la importancia de tener características con las que alimentar el modelo de ML.

#### **Time-series forecasting using ordinary Machine Learning algorithms [23]**

En él se plantea la eficiencia de un algoritmo de machine learning para la estimación de series temporales, añadiendo un tema muy importante, las empresas no solo se fijan en su nivel de acierto, si no también en los recursos necesarios y la complejidad. El objetivo de este estudio es predecir la temperatura, usando regresión lineal y random forest. Como conclusión, resalta el buen desempeño de random forest y debido a su bajo coste, abre la ventana al uso de algoritmos de machine learning en series temporales.

## 1.2 Objetivos

El objetivo general de este trabajo de fin de grado es identificar, construir, evaluar y validar un modelo de previsión de las ventas de los productos suministrados por la Gaviota Alimentación que contemple la estacionalidad.

Del anterior objetivo general se derivan los siguientes objetivos específicos:

- Estudiar, aplicar y evaluar modelos de estimación de ventas.
- Identificar el mejor modelo.
- Validar y construir un prototipo que incorpore los modelos estudiados en este trabajo y que facilite la evaluación de los modelos y la identificación, por tanto, del más adecuado.

## **1.3 Fases**

El proyecto se divide en 6 fases:

En la primera fase se aborda lo relativo a los datos y su preprocesado, es decir, el estudio, la representación y la eliminación de atributos o datos nulos o que se creen no son relevantes para el trabajo. La segunda fase ha sido el estudio de los diferentes modelos de estimación de ventas, teniendo en cuenta los datos para ver cuál se puede adaptar mejor a ellos. En la tercera fase, se aplican los modelos a los datos facilitados y se evalúa su rendimiento. En la cuarta fase, a partir de los resultados obtenidos, se eligen los modelos que mejor se adaptan para cada artículo. En la quinta fase, se construye un prototipo de estimación de ventas, basado en los resultados anteriores. Por último, en la sexta fase, se lleva a cabo la redacción de la memoria.

## 2. Los datos y su tratamiento

La empresa ha facilitado datos históricos referentes a las ventas de cuatro productos representativos de su cartera de productos. Los datos abarcan el periodo histórico que va desde enero de 2014 hasta febrero de 2020. Las entradas del fichero de datos se encontraban reflejadas por orden de venta, no agrupadas ni separadas por artículo.

Los atributos disponibles de cada artículo son:

- Clave documento ventas: identificador de la venta.
- Fecha contable: fecha de la venta.
- Código artículo: artículo al que corresponde la venta.
- Unidades vendidas: número de artículos vendidos.
- Código almacén: almacén en el que se encontraba el producto vendido.
- Promoción: existía promoción o no para la operación.

Se ha usado “Código artículo” para separar el fichero de datos y obtener uno por cada uno de los artículos. No se han empleado durante el análisis los atributos “Clave documento ventas”, “Código almacén” y “Promoción”. Estos atributos podrían usarse en el futuro como atributos para la predicción, aunque, en estos momentos, está fuera del alcance de este proyecto evaluar el impacto que tienen en la predicción de las ventas.

La “Fecha contable” ha permitido agrupar los datos de manera mensual, para así visualizar mejor la posible estacionalidad de los mismos.

Es importante destacar que las ventas se han visto afectadas, en mayor o menor medida, por las restricciones impuestas como consecuencia de la pandemia debida al Covid-19. El comportamiento de las ventas durante ese periodo difiere enormemente del comportamiento histórico previo. Por ello, se tomó la decisión de eliminar los datos de ventas posteriores a diciembre de 2019.

El conjunto de datos se ha dividido en dos. El periodo que va de enero de 2014 a diciembre de 2018 constituye el conjunto de entrenamiento. Como conjunto de validación se ha tomado el periodo que va de enero de 2019 a diciembre del mismo año.

Para comprobar la estacionalidad de las ventas se ha usado la función “IsSeasonal”, de la librería “seastest” [9] de R, en esta se lleva a cabo el análisis de la varianza, mediante el test de friedman, con esto se puede definir si es estacional o no.

### 2.1. Los artículos

Como se ha señalado anteriormente, se cuenta con los datos de ventas de cuatro productos representativos de la cartera de producto La Gaviota Alimentación. A continuación se describen brevemente estos artículos. La descripción será útil para contextualizar y explicar los resultados obtenidos en el análisis.

- **Primer artículo.** Los datos de los que se disponen indican que es vendido en hoteles, restaurantes y cafeterías. Esto nos hace pensar que va a ser un producto muy afectado por la pandemia. La gráfica de los datos es la siguiente:

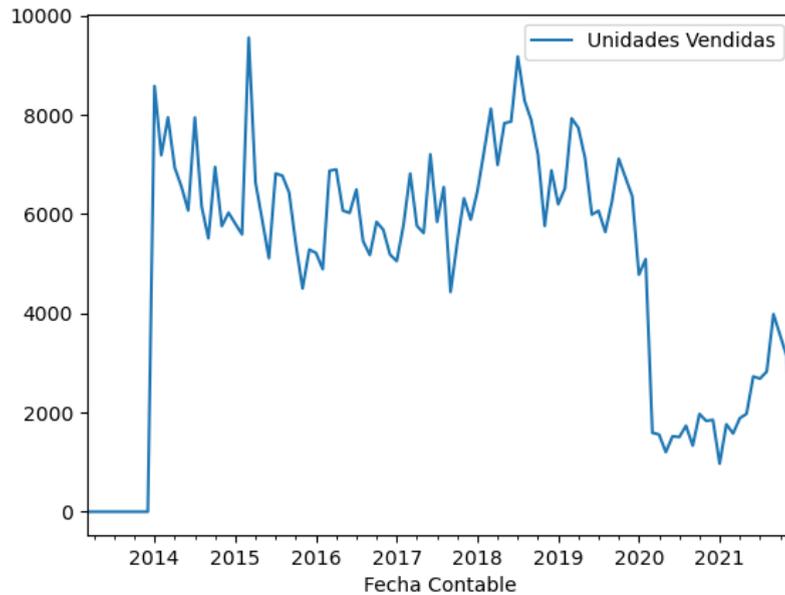


Figura 1.1: Artículo 1, ventas con covid.

A partir de la gráfica de la figura 1.1, vemos que por el cierre de los lugares a los que se comercializaba este producto, la reducción en las ventas ha sido más que evidente.

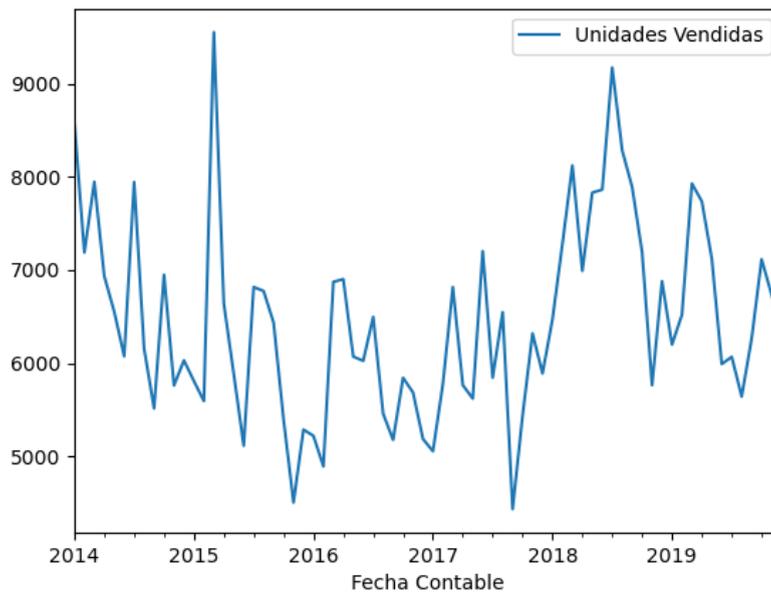


Figura 1.2: Artículo 1, ventas sin covid.

Los datos representados en la figura 1.2, obtenidos tras eliminar las ventas posteriores a diciembre de 2019, son los usados para el estudio, ya que en ellos podemos ver el comportamiento normal de las ventas.

Por otro lado, en este artículo no apreciamos una estacionalidad clara, ya que los datos de meses pasados no guardan relación con los posteriores. Esta afirmación es confirmada por el contraste de hipótesis de que se lleva a cabo por medio del test de Friedman.

- **Segundo artículo.** Atendiendo a la descripción dada por la empresa, se trata de un artículo de la marca principal que se vende en hipermercados y supermercados.

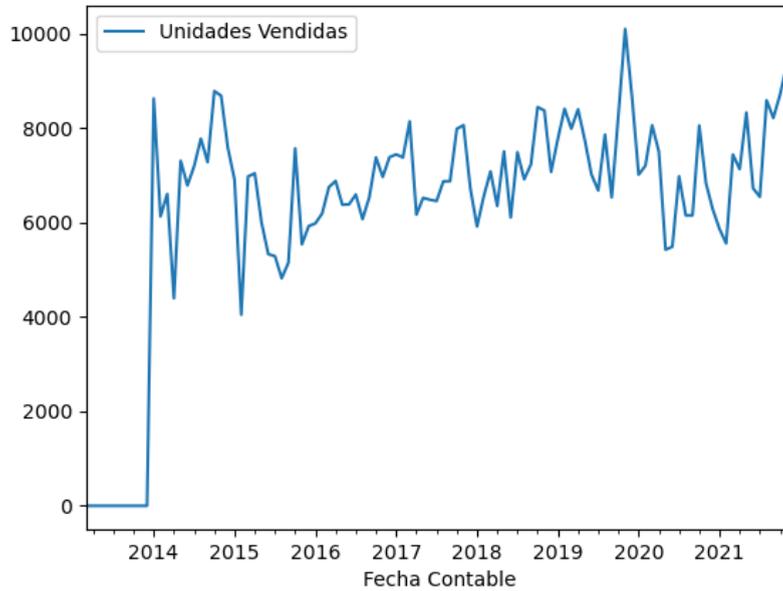


Figura 1.3: Artículo 2, ventas con covid.

En este artículo los efectos de la pandemia no son tan evidentes, a pesar de ello, hemos decidido eliminarlos para estudiar las ventas en el mismo período de tiempo para todos los productos.

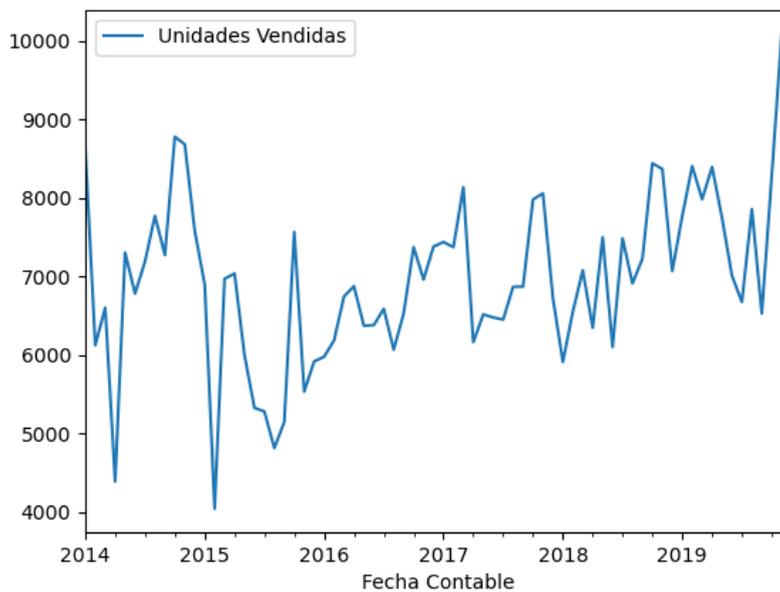


Figura 1.4: Artículo 2, ventas sin covid.

El posible motivo por el que quizás no se ha visto tan afectado por la pandemia es que se comercializa en hipermercados y supermercados, comercios que siguieron abiertos y demandados durante ese periodo.

A pesar de no poder identificar estacionalidad marcada a priori, tras aplicar el test de Friedman, podemos asegurar que sí hay estacionalidad.

• **Tercer artículo.** Este artículo es de la marca principal y se vende en hipermercados y supermercados.

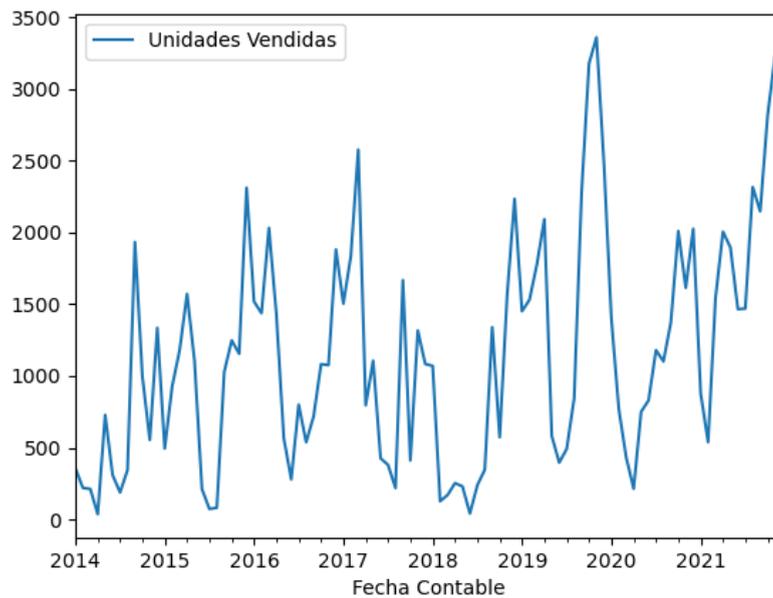


Figura 1.5: Artículo 3, ventas con covid.

Igual que el producto anterior, no presenta un cambio significativo tras el covid, ya que los clientes son los mismos.

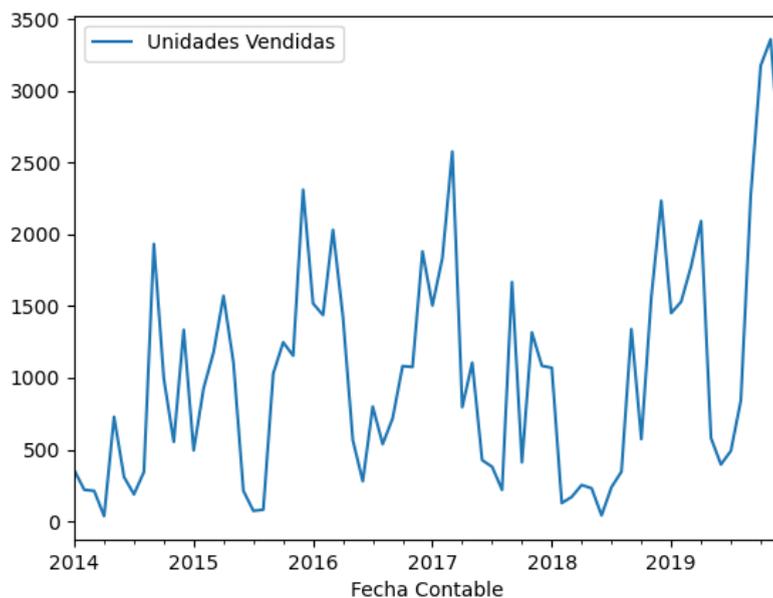


Figura 1.6: Artículo 3, ventas sin covid.

En este artículo podemos ver una estacionalidad muy marcada, ya que hay un aumento de las ventas en los últimos meses del año. Esto se repite todos los años de los que hay registro. Para corroborarlo, aplicamos el contraste de hipótesis usando Friedman y el resultado nos indica que existe estacionalidad.

- **Cuarto artículo.** Este artículo es de una marca secundaria y es el que más se vende en hipermercados y supermercados.

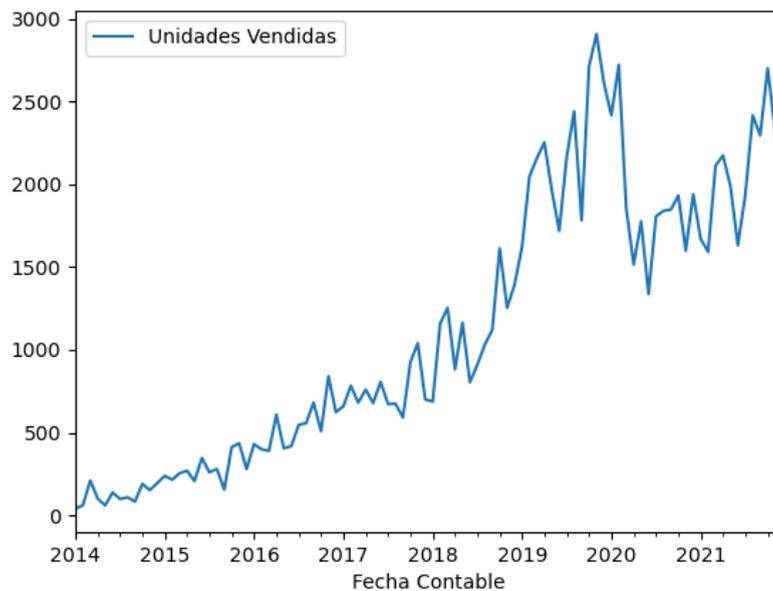


Figura 1.7: Artículo 4, ventas con covid.

Tras el covid vemos que las ventas se recuperan, pero aún así presentan una bajada significativa.

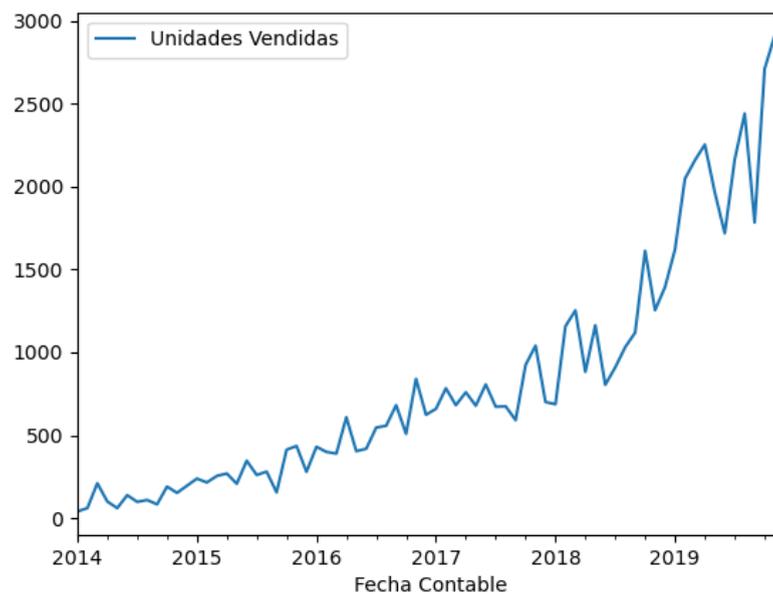


Figura 1.8: Artículo 4, ventas sin covid.

Podemos ver que tiene una tendencia creciente, además cuenta con cierta estacionalidad, ya que dentro de su crecimiento repite picos en meses clave. Para corroborar la estacionalidad, aplicamos el contraste de hipótesis usando el test de Friedman, que nos permite comprobar que existe estacionalidad.

# 3. Los modelos de predicción

El proceso de predicción de las ventas consta de dos fases. En la primera se identifica un modelo que se ajuste o explique los datos de ventas históricos; en la segunda, se emplea dicho modelo para predecir el valor futuro de las ventas en el horizonte de predicción considerado.

En el presente trabajo se emplean dos paradigmas de predicción que nos permiten identificar el modelo. En el primero, que podríamos llamar clásico, los datos de ventas se conciben como una serie temporal cuyo modelo debe ser identificado. En este paradigma se asume que solo las ventas pasadas explican las ventas presentes y futuras. Al segundo paradigma pertenecen las técnicas de Machine Learning que construyen un modelo de predicción a partir de varias variables explicativas o atributos. Entre los modelos del primer paradigma destacan la Autoregresión, SARIMA y Prophet. Al segundo pertenece, entre otros, Random Forest.

La teoría clásica establece que los modelos básicos de una serie temporal tienen potencialmente tres componentes: tendencia, estacionalidad y ruido o error. Es decir, modeliza la serie según la siguiente expresión

$$y(t) = g(t) + s(t) + e(t)$$

donde  $y(t)$  es el valor de la variable en el instante  $t$ ,  $g(t)$  es el valor de esa variable explicado por la tendencia,  $s(t)$  el valor explicado por la estacionalidad y  $e(t)$  el error.

Dependiendo de cómo se obtengan las anteriores componentes se obtiene un modelo de serie temporal u otro.

A continuación se describen brevemente los modelos usados en el trabajo.

## 3.1. Autoregresión

Es un modelo en el que la variable de interés depende de sus valores pasados, es decir, es la misma que la explicativa pero en un momento de tiempo posterior. Por este motivo se dice que existe una dependencia lineal entre ambas variables.

En este método se debe indicar el número de intervalos de tiempo en los que se va a estudiar la variable(lags). La función utilizada viene implementada por `statsmodels.tsa.ar_model` [20].

## 3.2. SARIMA

S: Seasonal, AR: Autoregressive, I: Integrated, MA: Moving Average. Es un modelo de predicción usado en series temporales. Este toma en cuenta los valores pasados de una variable para hacer una predicción. A diferencia de ARIMA, también tiene en cuenta la estacionalidad. Este modelo cuenta con 7 hiperparámetros, que podemos dividir en no estacionales y estacionales:

ARIMA (p,d,q) (P,D,Q)m

AR(p) Es el entero que indica cuántos valores antiguos van a ser utilizados.

I(d) Es el número de valores que usaremos para calcular la diferencia.

MA(q) Es el número de valores usados para la media móvil.

m Es la estacionalidad(12 es anual).

El resto de valores representa lo mismo, pero se escogen en múltiplos de m. La función que se ha usado para la predicción viene dada por el paquete "statsmodels.tsa.statespace.sarimax" [21] y se llama "SARIMAX".

Puede optarse por dos procedimientos alternativos para ajustar el valor de los parámetros que gobiernan el modelo Sarima. El primero sería el ajuste manual en el que, tras fijar diferentes valores para cada uno de los parámetros, se mediría la calidad del ajuste a los datos del modelo con esos valores. Se trata de un procedimiento laborioso que consume una gran cantidad de tiempo. En el segundo procedimiento se confía en los desarrollos de otros autores, que han implementado estrategias para variar el valor de los parámetros y escoger la combinación que presenta mejor ajuste. En este trabajo se ha optado por la segunda de las alternativas. La función `auto_arima`, del paquete `pmdarima` [3], a partir de unos valores iniciales entre los que se encuentra el de la estacionalidad y el número máximo para cada parámetro, prueba diferentes combinaciones para ofrecer la más adecuada para los datos dados.

### 3.3. Prophet

Es una herramienta de código abierto de Facebook usado para estimación de datos de series temporales. Basado en un modelo aditivo donde las tendencias son ajustadas a la estacionalidad y también toma en cuenta las vacaciones. La ecuación se ajusta de la siguiente manera:

$$y(t) = g(t) + s(t) + h(t) + e(t)$$

$g(t)$  tendencia

$s(t)$  estacionalidad

$h(t)$  efecto de las vacaciones

$e(t)$  término de error

Prophet exige una estructura específica para la entrada de datos, en la que la fecha debe ser denominada 'ds' y los valores 'y'.

### 3.4. Random Forest

Es un algoritmo que entrena varios árboles de clasificación y propone como predicción el valor medio de las predicciones hechas por cada árbol individual. Cada árbol se construye restringiendo el conjunto de atributos que pueden usar en un procedimiento que incrementa la probabilidad de que los atributos formen parte de algún árbol. De este modo, cada árbol suele especializarse en la tarea de clasificación usando un conjunto restringido de atributos. La construcción se hace en paralelo y sin interacción entre los árboles. Un parámetro relevante del procedimiento, que debe ser indicado, es el número de árboles que se construirán. En la figura 2.1 se muestra esquemáticamente el proceso de predicción usando Random Forest.

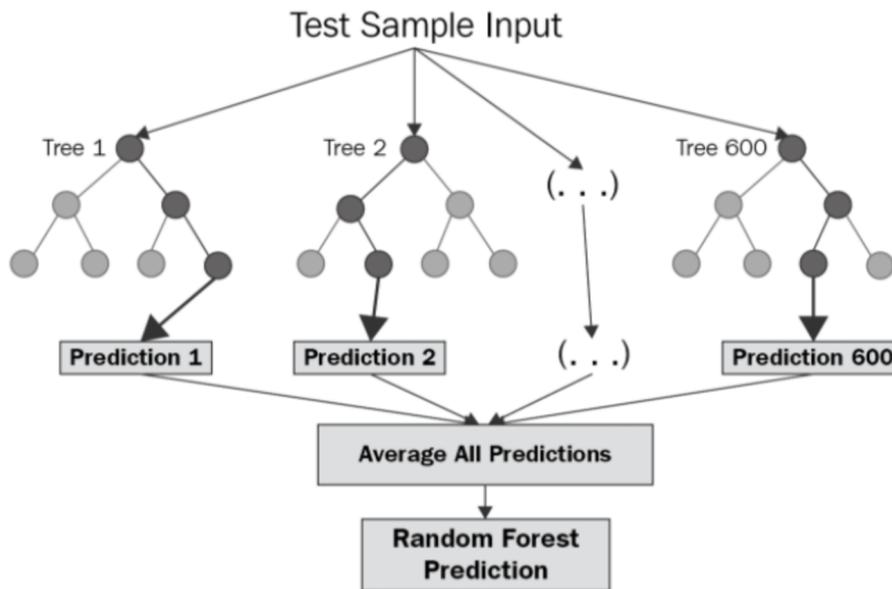


Figura 2.1: Funcionamiento Random Forest. [4]

Este modelo viene implementado por “RandomForestRegressor”, en la función “sklearn.ensemble.randomForestRegressor” [19]. En este se debe indicar la variable target, que en nuestro caso es el número de ventas y los features que son todas aquellas variables que vamos a usar para hacer la predicción.

Para este estudio se han decidido añadir los siguientes atributos numéricos:

- Media de los últimos 3 meses de ventas.
- Crecimiento en los últimos dos meses.
- Media de las ventas en el período de un año.
- Año de la venta.
- Mes de la venta.

Es importante destacar que algunos de estos atributos no están disponibles para algunas predicciones, en este trabajo se han podido usar ya que el objetivo era validar el modelo.

# 4. Resultados obtenidos

Para evaluar los modelos considerados en este trabajo la serie histórica de ventas se divide en dos conjuntos: uno de entrenamiento y otro de validación. El primero se usa para ajustar el modelo y el segundo para evaluar la calidad del mismo.

Ajustado el modelo, este puede usarse para predecir las ventas futuras. El número de predicciones determina la ventana de predicción. Se habla de una ventana anual cuando se predicen las ventas de los siguientes 12 meses. Del mismo modo pueden considerarse ventanas semestrales, trimestrales y mensuales. En todos los casos, la serie de entrenamiento empieza en enero de 2014 y finaliza un año, seis meses, tres meses o un mes antes del inicio del periodo de validación.

A continuación se describen los resultados obtenidos. Estos se encuentran en las tablas del apéndice 2. Para cada ventana de predicción se muestran las ventas reales, las predicciones que hace cada modelo y el porcentaje de desviación entre el valor real y la predicción. Además, se muestran también las estimaciones de la media móvil (método usado por la empresa) para llevar a cabo una comparación efectiva.

## 4.1. Artículo 1

Los resultados de las predicciones de venta del artículo 1 se encuentran en las tablas 2.1, 2.2, 2.3 y 2.4. En ellas se muestran las ventas reales y las predicciones con ventanas de predicción anual (tabla 2.1), semestral (tabla 2.2), trimestral (tabla 2.3) y mensual (tabla 2.4). En todos los escenarios se muestran las predicciones y desviaciones, con respecto a las ventas reales, obtenidas al aplicar los métodos considerados en el presente trabajo.

### ■ Evaluación por método

Estimaciones **media móvil**:

En el escenario anual, la desviación ha sido de 8.36%, en el semestral 8.87%, por lo que no hay un gran cambio en la desviación del modelo, pero en el estudio trimestral es de 10.59% y en el mensual 10.87%. Por lo que al pasar a una ventana de estudio más corta, vemos cómo empeora el rendimiento del modelo en casi un 2%.

Estimaciones **Autoregresión**:

En el escenario anual, la desviación es de 11.18%, en el semestral 11.02%, la diferencia en estos dos primeros escenarios es relativamente pequeña, pero al estudiarlo de manera trimestral, desciende un 1.41%, hasta el 9.61% y en el mensual se reduce un 1.39% hasta llegar al 8.42%. Por lo que podemos decir que conforme se añaden más datos al entrenamiento los resultados van mejorando.

Estimaciones **SARIMA**:

En el escenario anual, la desviación es de 11.89%, en el semestral 10.10%, entre estas dos primeras ventanas mejora 1.79%, al aplicarle la semestral mejora sutilmente, estableciéndose en un 9.86%, en la última ventana la desviación disminuye en un 1.15%, confirmando así que cuanto más pequeña sea la ventana, mejores resultados se obtienen.

### Estimaciones **Prophet**:

En el anual, la desviación es de 23.51%, en el semestral disminuye notablemente, hasta el 18.5%, sin embargo, en el trimestral mejora solo un 1%, hasta el 17.49%, al igual que en mensual, que disminuye hasta el 16.39%. Por todo esto podemos decir, que conforme más pequeña es la ventana, es mejor la predicción, pero con todo esto, la desviación sigue siendo elevada.

### Estimaciones **Random Forest**:

La desviación anual es del 12.04%, al ver la semestral, sorprendentemente, lejos de disminuir, aumenta un 2%, quedando en 14.09%, hecho que cambia cuando miramos los datos trimestrales, que mejoran y se posicionan con una desviación similar a la predicción anual, 12.13%, rendimiento que sigue mejorando, aunque no de manera muy notoria en el estudio mensual, 11.8%.

#### ■ Comparación entre modelos por escenario

##### Estimación **anual**:

En esta estimación el método que mejores resultados ofrece es la media móvil, superando en un 2.8% a la autoregresión, este a SARIMA en un 3.53%, este a random forest en un 3.68%, pero sin duda alguna, el método que ha presentado peor rendimiento ha sido prophet, con una diferencia de 15.15% del último.

##### Estimación **semestral**:

Sigue liderando la media móvil, pero ahora el resto de métodos se han acercado más, no de manera muy notoria, pero ahora en lugar de ser la autoregresión el segundo método con mejores resultados, es SARIMA el que posee este lugar, seguido por la autoregresión, el random forest y por último Prophet.

##### Estimación **trimestral**:

En este escenario, el método más eficaz ha sido la autoregresión, seguido con una diferencia muy pequeña(0.3%) por SARIMA, luego de este vendría la media móvil, que ha empeorado su rendimiento, seguida por el random forest y por último Prophet.

Fecha Contable	Unidades Vendidas	Media Móvil		Autoregresión		Sarima		Prophet		Random Forest	
		Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación
31/1/19	6199	6612.33	6.67	6216.85	0.29	7171.58	15.69	7426.05	19.79	7099.28	14.52
28/2/19	6519	6280.67	3.66	6035.02	7.42	6772.49	3.89	7686.97	17.92	6416.14	1.58
31/3/19	7930	6532.67	17.62	8142.60	2.68	6827.46	13.90	9899.97	24.84	7058.18	10.99
30/4/19	7738	6882.67	11.05	6441.43	16.76	7329.08	5.28	7300.41	5.66	7368.24	4.78
31/5/19	7122	7395.67	3.84	6841.45	3.94	7303.45	2.55	7163.06	0.58	6820.33	4.24
30/6/19	5990	7596.67	26.82	6860.68	14.54	7155.40	19.46	6646.38	10.96	7606.77	26.99
31/7/19	6067	6950.00	14.55	6456.53	6.42	6707.64	10.56	8445.36	39.20	7330.99	20.83
31/8/19	5641	6393.00	13.33	5836.23	3.46	6661.02	18.08	7727.37	36.99	7961.80	41.14
30/9/19	6264	5899.33	5.82	4985.49	20.41	6331.12	1.07	7171.39	14.49	6633.36	5.90
31/10/19	7116	5990.67	15.81	6031.56	15.24	6488.04	8.82	6677.16	6.17	6864.84	3.53
30/11/19	6737	6340.33	5.89	6111.66	9.28	6786.67	0.74	5459.48	18.96	6942.06	3.04
31/12/19	6364	6705.67	5.37	6399.71	0.56	6646.78	4.44	6437.28	1.15	6621.24	4.04
			10.87		8.42		8.71		16.39		11.80

Tabla 2.4: Artículo 1, Estimación mensual

#### Estimación **mensual**:

En esta ocasión se repiten los resultados, además todos los modelos, a excepción de la media móvil, han mejorado las predicciones.

Como conclusión podemos decir que en caso de querer llevar a cabo predicciones en periodos de tiempo más prolongados (anuales, semestrales), la mejor opción es la media móvil, pero si las predicciones van a ser en periodos cortos (trimestrales, mensuales) la autoregresión o en su defecto SARIMA, serían los métodos más adecuados.

## 4.2. Artículo 2

Para el artículo 2, los resultados de las predicciones y desviaciones respecto de las ventas reales tras aplicar los métodos estudiados, se muestran por ventana de predicción anual (tabla 2.5), semestral (tabla 2.6), trimestral (tabla 2.7) y mensual (tabla 2.8).

#### ■ Evaluación por método

##### Estimaciones **media móvil**:

En el escenario anual, la desviación ha sido de 9.06%, en el semestral 9.52%, vemos que no ha aumentado demasiado, pero al comprobar con el trimestral, 11.45%, ha aumentado un 1.93%, lo llamativo, es que al hacer la predicción mensual, lejos de mantener este aumento, disminuye 1.6% para situarse en 9.85%. No podemos asegurar si aumenta o disminuye su efectividad según hacemos más pequeña la ventana.

### Estimaciones **Autoregresión:**

En la predicción anual, presenta una desviación de 16.51%, que disminuye a 15.22% en el semestral, lo que supone una disminución de menos de 1%, esta situación se mantiene en las estimaciones trimestrales, en las que la desviación es de 14.36%, pero curiosamente, en el escenario mensual, ha disminuido en 3.17%, situándose en 11.19%, lo que implica que ha mejorado la predicción de manera notoria en esta última ventana.

### Estimaciones **SARIMA:**

En la estimación anual, presenta una desviación de 8.9%, en la semestral aumenta a 9.8%, casi un 1% más, sigue creciendo hasta 9.66% en el trimestral y 10.37% en el mensual. Sin duda alguna, cuanto más pequeña sea la ventana, peor rendimiento tendrá éste método, a pesar de que la diferencia no es muy grande.

### Estimaciones **Prophet:**

En la ventana anual, tiene una desviación de 12.02%, en la semestral, crece muy poco, hasta 12.35%, en la trimestral baja a 12.04% y en la mensual baja a 11.98%. No presenta variaciones muy elevadas, pero donde tiene una mayor eficacia es en la estimación mensual.

### Estimaciones **Random Forest:**

En el escenario anual, hay una desviación de 16.07%, en el semestral aumenta poco, hasta 16.46%, en el trimestral disminuye casi un 2%(1.89%) y en el mensual continúa disminuyendo hasta el 11.94%. Sin duda alguna, podemos afirmar que este modelo tiene un mejor rendimiento para ventanas de predicción pequeñas(trimestral y mensual).

#### ■ Comparación entre modelos por escenario

##### Estimación **anual:**

En este escenario, el método que presenta un mejor desempeño es SARIMA, con un 8.9% de desviación, le sigue la media móvil, por no decir que están casi igualados, con un 9.6%, superando a prophet que presenta una desviación algo mayor, 12.02%, casi igualados en última posición, tenemos a Random Forest 16.07% y la autoregresión con 16.51% y una diferencia de casi 7.5% con respecto al líder.

##### Estimación **semestral:**

Con una desviación algo mayor, sigue SARIMA como el modelo con mejores predicciones, todos mantienen el mismo orden, sin embargo, el único que disminuye es la autoregresión.

##### Estimación **trimestral:**

Segue aumentando la desviación, pero se mantiene líder SARIMA, es importante destacar que la autoregresión, Prophet y el random forest, han mejorado sus predicciones, sobre todo el random forest. A pesar de ello, se mantiene el orden.

Fecha	Unidades	Media Móvil		Autoregresión		SARIMA		Prophet		Random Forest	
		Vendidas	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción
31/1/19	7767	7958.67	2.47	7009.41	9.75	7789.15	0.29	7195.16	7.36	6174.44	20.50
28/2/19	8404	7734.33	7.97	6398.03	23.87	6960.30	17.18	5919.34	29.57	6663.96	20.70
31/3/19	7983	7746.67	2.96	8483.41	6.27	8541.27	6.99	7659.20	4.06	7765.25	2.73
30/4/19	8393	8051.33	4.07	6981.59	16.82	7440.94	11.34	6644.56	20.83	7546.43	10.09
31/5/19	7747	8260.00	6.62	7253.27	6.37	8350.92	7.80	7384.37	4.68	8109.48	4.68
30/6/19	7010	8041.00	14.71	6828.18	2.59	7624.17	8.76	6416.99	8.46	7891.74	12.58
31/7/19	6675	7716.67	15.61	6733.07	0.87	7738.22	15.93	7012.13	5.05	7319.84	9.66
31/8/19	7857	7144.00	9.07	6231.40	20.69	7195.24	8.42	6903.53	12.14	7298.04	7.11
30/9/19	6530	7180.67	9.96	7239.24	10.86	7612.74	16.58	7002.82	7.24	7704.24	17.98
31/10/19	8285	7020.67	15.26	7885.01	4.83	8463.67	2.16	9213.33	11.21	7234.20	12.68
30/11/19	10094	7557.33	25.13	7216.00	28.51	7900.42	21.73	8115.07	19.60	7881.41	21.92
31/12/19	8678	8303.00	4.32	8427.07	2.89	8052.60	7.21	7498.38	13.59	8905.18	2.62
			9.85		11.19		10.37		11.98		11.94

Tabla 2.8: Artículo 2, Estimación mensual

#### Estimación **mensual**:

En este escenario, la media móvil es el método que mejores predicciones presenta(9.85%), con menos de 1% de diferencia con respecto a SARIMA(10.37%), seguido por la autoregresión(11.19%), por menos de 1% y por último situaremos a prophet(11.98%) y random forest(11.94%) con un rendimiento casi idéntico.

Podemos concluir, que a pesar de que la media móvil presenta el mejor rendimiento en las estimaciones mensuales, todos se mantienen muy cerca. Además, en todos los otros escenarios, SARIMA sería la mejor elección.

### 4.3. Artículo 3

Los resultados de las predicciones del artículo 3 se encuentran en las tablas 2.9 a 2.12. En ellas se muestran las ventas reales y las predicciones con ventanas de predicción anual (tabla 2.9), semestral (tabla 2.10), trimestral (tabla 2.11) y mensual (tabla 2.12). En todos los escenarios se muestran las predicciones y desviaciones, con respecto a las ventas reales, obtenidas al aplicar los métodos considerados en el presente trabajo.

#### ■ Evaluación por método

##### Estimaciones **media móvil**:

En el escenario anual, tiene una desviación de 90.28%, en el semestral baja un 6.87% hasta el 83.41%, es una bajada considerable, en el trimestral presenta otra gran bajada, disminuye un 8.86%, llegando al 74.55% de error y en el mensual la bajada es mucho más pequeña(2.18%) fijando el valor en 72.37%. Es evidente que la desviación es

excesivamente grande en todos los escenarios, sin embargo, disminuye considerablemente a medida que la ventana es más pequeña.

#### Estimaciones **Autoregresión**:

En la estimación anual, muestra una desviación del 34.26%, en el semestral de 34.57%, por lo que aumenta de manera casi imperceptible, en la trimestral, disminuye un 1.95%, quedando en 32.62% y en la mensual, aumenta un 6.03%, estableciéndose en 38.65%. En las tres primeras observaciones el modelo llega a bajar el porcentaje de desviación, pero en la predicción mensual tiene un rendimiento notablemente peor.

#### Estimaciones **SARIMA**:

En la predicción anual, la desviación es del 34.83%, en la semestral casi no varía, 34.55%, en la trimestral aumenta notablemente, un 10.09%, situándose en 44.64% y en la mensual disminuye casi un 2%, quedando en 42.99%. Es notable que conforme la ventana es más pequeña, peor es el rendimiento de este método.

#### Estimaciones **Prophet**:

En la ventana anual, la desviación es de 51.37%, en la semestral disminuye casi un 6% hasta 46.59%, en la trimestral aumenta un 1.97%, situándose en 48.56% y en la mensual, sube poco más del 1%, quedando en 49.13%. Podemos afirmar, que conforme se hace más pequeña la ventana empeora la capacidad de predicción.

#### Estimaciones **Random Forest**:

En el escenario anual, la desviación es de 49.71%, en la semestral aumenta menos de 1% llegando a 51.16%, en la trimestral baja poco más de 1%, situándose en 49.11% y por último en la mensual disminuye un 1.22%. Con todo esto, vemos que la desviación es similar en cualquier escenario, pero su mejor desempeño es en el mensual.

#### ■ Comparación entre modelos por escenario

##### Estimación **anual**:

En este escenario, la autoregresión con un 34.26%, presenta el mejor desempeño junto con SARIMA(0.57% de diferencia), luego vendría el random forest con casi un 15% más de desviación(49.71), prophet le sigue con 1.66% más, por lo que podemos decir que tienen un rendimiento similar. El método cuya predicción ha sido muy poco acertada, ha sido la media móvil, con 56% más de error que la autoregresión, situándose en 90.28%.

##### Estimación **semestral**:

En esta ventana el mejor desempeño lo presenta SARIMA y la autoregresión, con una desviación de 34.55%, les sigue prophet con un 46.59%, adelantando al random forest, que queda con un 51.16%. En esta, repite la media móvil como el método con la peor predicción, con 83.41% de error.

### Estimación **trimestral**:

En este, el mejor resultado lo ofrece la autoregresión(32.63%), seguido con muy poca diferencia entre ellos por SARIMA, que ha empeorado la predicción,44.64%, prophet con casi un 4% más de desviación y random forest, con casi 1% más. La media móvil sigue presentando un mal desempeño, 74.55% de error.

Fecha Contable	Unidades Vendidas	Media Móvil		Autoregresión		SARIMA		Prophet		Random Forest	
		Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación
31/1/19	1452	1457.00	0.34	1361.46	6.24	1613.36	11.11	1167.05	19.62	1208.01	16.80
28/2/19	1531	1750.00	14.30	1371.79	10.40	1441.63	5.84	777.38	49.22	1644.70	7.43
31/3/19	1778	1739.33	2.17	1936.47	8.91	1795.60	0.99	873.41	50.88	1838.73	3.42
30/4/19	2093	1587.00	24.18	1259.87	39.81	1387.64	33.70	1151.20	45.00	1290.14	38.36
31/5/19	581	1800.67	209.93	1386.78	138.69	1695.11	191.76	1267.87	118.22	1680.03	189.16
30/6/19	396	1484.00	274.75	216.99	45.20	655.12	65.43	553.79	39.85	803.99	103.03
31/7/19	491	1023.33	108.42	456.61	7.00	596.03	21.39	299.54	38.99	511.74	4.22
31/8/19	842	489.33	41.88	335.14	60.20	526.90	37.42	363.93	56.78	394.23	53.18
30/9/19	2276	576.33	74.68	1418.46	37.68	1668.73	26.68	1635.97	28.12	875.29	61.54
31/10/19	3178	1203.00	62.15	1342.57	57.75	1532.54	51.78	1061.01	66.61	1428.43	55.05
30/11/19	3361	2098.67	37.56	2505.77	25.45	2774.83	17.44	1196.92	64.39	2274.30	32.33
31/12/19	2488	2938.33	18.10	3146.82	26.48	3790.57	52.35	2191.78	11.91	2739.95	10.13
			72.37		38.65		42.99		49.13		47.89

Tabla 2.12: Artículo 3, Estimación mensual

### Estimación **mensual**:

En esta última estimación, se mantiene como mejor método la autoregresión, 38.65%, seguido de SARIMA con una diferencia de 5.34%, random forest con diferencia de casi un 5%, seguido por una diferencia de 1.24% de Prophet. Una vez más, a pesar de mejorar, la media móvil se sitúa en último lugar con un 72.37% de desviación.

Podemos afirmar que en general la autoregresión es el método que mejor desempeño presenta, aunque no podemos olvidarnos de SARIMA que también ha presentado buenos resultados, sobre todo en los períodos más largos, random forest y prophet se mantienen como métodos con un desempeño similar entre ellos y entre todas las ventanas temporales. Por último, en este artículo la media móvil ha dado unos resultados muy poco acertados.

## 4.4. Artículo 4

Para el artículo 4, los resultados de las predicciones y desviaciones respecto de las ventas reales tras aplicar los métodos estudiados se muestran por ventana de predicción anual (tabla 2.13), semestral (tabla 2.14), trimestral (tabla 2.15) y mensual (tabla 2.16).

## ■ Evaluación por método

### Estimaciones **media móvil**:

En la predicción anual, tiene una desviación de 34.95%, en la semestral baja un 10.17%, situándose en 24.78%, es una gran mejora, en la predicción trimestral, a 18.87%, otra bajada importante de casi el 6%, en la predicción mensual se sitúa en 17.18%, bajando solo poco más de 1%. Visto esto, es evidente que el desempeño de la media móvil mejora cuando la ventana de predicción es más pequeña.

### Estimaciones **Autoregresión**:

En el escenario anual, la desviación es de 29.56%, el semestral, baja un 5.41% hasta el 24.15%, en el trimestral baja casi un 4% hasta 20.16% y por último en el mensual vemos otra bajada de más de 4% para terminar con una desviación de 16.57%. Conforme más pequeña es la ventana, mejor es la predicción de este método

### Estimaciones **SARIMA**:

En la estimación anual, presenta una desviación de 25.38%, número que baja 6.59% hasta el 18.79% en la predicción semestral, en la trimestral baja casi un 5% hasta 13.88%, sorprendentemente, en la estimación mensual sube menos de un 1%, situándose en 14.14%. Conforme hacemos el escenario más pequeño, mejores resultados se obtienen, a excepción del mensual que sube el error, pero muy poco.

### Estimaciones **Prophet**:

En el escenario anual, la desviación es de 38.71%, esto baja considerablemente, un 12% en la estimación semestral, quedando en 26.71%, baja un 5% en la trimestral, teniendo un 21.74% y baja nuevamente, pero en esta ocasión casi un 2%, para un error de 19.86%. A medida que la ventana se reduce es mejor la predicción, a pesar de ello, no hay gran diferencia entre la estimación trimestral y mensual.

### Estimaciones **Random Forest**:

En la predicción anual, la desviación es de 39.65%, en la semestral baja hasta 28.66%, casi un 11%, en la trimestral baja más de 7%, situándose en 21.97%, para finalmente quedar con una desviación de 17.34% en la mensual. Sin duda alguna, cuanto más corta es la ventana de predicción, mejor es el resultado.

## ■ Comparación entre modelos por escenario

### Estimación **anual**:

En este escenario, el modelo con mejores resultados es SARIMA, con un 25.38%, seguido con un 4.18% más de desviación la autoregresión(29.56%), seguido por la media móvil con 5.39% más para situarse en 34.95%, seguida por el prophet con 38.71% que tiene menos de 1% de diferencia con random forest.

### Estimación **semestral**:

En esta se mantienen las posiciones, sin embargo, es importante destacar que la media móvil se sitúa con solo 0.63% más de desviación, también recorta la distancia prophet y en menor medida random forest.

### Estimación **trimestral**:

En ella sigue SARIMA como la que ofrece mejores resultados con un 13.88% de desviación, esta vez seguida por la media móvil(18.87%) y luego por la autoregresión(20.16%), por último prophet y random forest tienen porcentajes de desviación muy similares.

Fecha Contable	Unidades Vendidas	Media Móvil		Autoregresión		SARIMA		Prophet		Random Forest	
		Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación
31/1/19	1620	1420.00	12.35	1301.90	19.64	1173.16	27.58	1168.72	27.86	1259.80	22.23
28/2/19	2047	1422.67	30.50	1734.49	15.27	1747.20	14.65	1289.13	37.02	1486.02	27.41
31/3/19	2158	1687.00	21.83	1582.43	26.67	2102.67	2.56	1511.70	29.95	1818.38	15.74
30/4/19	2253	1941.67	13.82	1919.21	14.82	1789.58	20.57	1437.30	36.20	1968.09	12.65
31/5/19	1964	2152.67	9.61	1928.65	1.80	2244.86	14.30	1621.00	17.46	2141.98	9.06
30/6/19	1718	2125.00	23.69	1758.00	2.33	1843.43	7.30	1753.02	2.04	2021.80	17.68
31/7/19	2164	1978.33	8.58	1680.72	22.33	1932.98	10.68	1760.41	18.65	1862.76	13.92
31/8/19	2440	1948.67	20.14	2127.62	12.80	2014.50	17.44	1942.51	20.39	2079.38	14.78
30/9/19	1783	2107.33	18.19	2538.20	42.36	2324.49	30.37	1936.18	8.59	2082.07	16.77
31/10/19	2710	2129.00	21.44	2078.38	23.31	2710.29	0.01	2369.15	12.58	1807.08	33.32
30/11/19	2908	2311.00	20.53	2565.01	11.79	2366.92	18.61	2321.30	20.18	2295.59	21.06
31/12/19	2612	2467.00	5.55	2760.31	5.68	2464.70	5.64	2418.87	7.39	2521.36	3.47
		17.18		16.57		14.14		19.86		17.34	

Tabla 2.16: Artículo 4, Estimación mensual

### Estimación **mensual**:

En esta ventana, a pesar de que aumenta un poco la desviación, SARIMA se mantiene primero, con algo más de 2% sigue autoregresión, seguido por media móvil con casi 1% más de desviación, luego random forest con 17.34% y prophet con 19.86%.

Podemos concluir en que SARIMA es el método con mejores resultados en todos los escenarios, la autoregresión y la media móvil tienen un desempeño similar y por su parte prophet y el random forest también.

# 5. Prototipo sistema de predicción de ventas

Este prototipo es un programa de consola que ha sido desarrollado en python siguiendo la programación orientada a objetos. Cuenta con los diferentes modelos que se han estudiado en este proyecto. Dispone de dos funcionalidades: validación, y predicción.

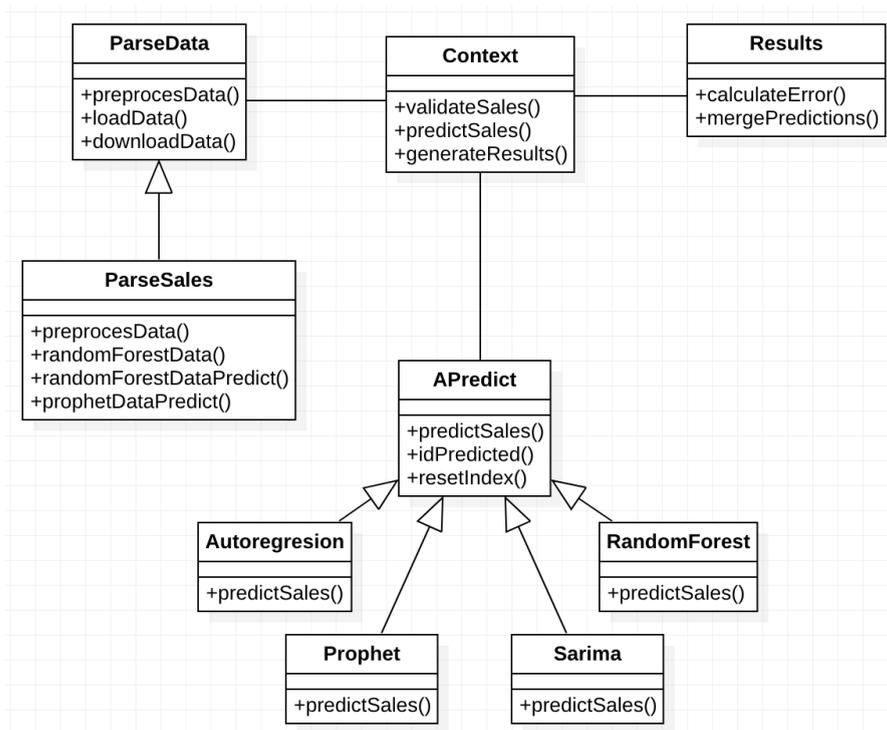


Figura 3.1: Diagrama de clases del prototipo [2].

Como podemos observar en la figura 3.1 (diagrama de clases), el prototipo se divide en 3 módulos: "Data", en el que está todo lo relativo a la carga y el preprocesado de los datos, implementado mediante un patrón estrategia, por si es necesario añadir algún otro método de carga; "Predict", en el que se ha implementado un patrón estrategia para la predicción de cada uno de los métodos; "Results" que se encarga de lo referente a los resultados de la validación. Toda la lógica es manejada en "Context", que es el que inicializa todo lo necesario para el funcionamiento del programa. Por último "Program", que contiene el menú con las diferentes opciones y el objeto del contexto para llevar a cabo lo que requiera el usuario.

## Validación

En esta modalidad el objetivo es conocer el porcentaje de error de cada modelo para los datos facilitados. En él se tomará un año como validación y el resto como entrenamiento, los resultados son generados en una tabla que indica el porcentaje de error de cada método. De esta manera, se puede ver el desempeño de cada algoritmo para los datos facilitados.

Es importante mencionar, que para el random forest, debemos introducir el nombre del fichero que contiene los datos, ya que estos llevan un preprocesado diferente al necesario para el resto de métodos. En este debe tener la siguiente estructura:

Unidades Vendidas	Media_3mes	Media_12mes	Crecimiento	Año	Mes
-------------------	------------	-------------	-------------	-----	-----

Figura 3.2: Formato fichero validación random forest

Puede tener más características de las ilustradas, pero es importante que el target sea incluido como “Unidades Vendidas”.

### Predicción

En esta función, a partir de los datos de la validación, que son tomados como entrenamiento, se debe introducir el número de meses que se quieren predecir, el fichero con el formato requerido por prophet para la predicción, es decir, una columna denominada “ds”, con las fechas a predecir, además de un fichero con las características requeridas para llevar a cabo la predicción con el random forest.

Con todo esto, se lleva a cabo la predicción con todos los métodos del estudio.

### Demstración del funcionamiento del programa

Al iniciar el programa, pide la ruta en la que se encuentran todos los archivos que se usarán para la ejecución, mencionados anteriormente. Tanto para validar como para predecir, será necesario introducir los hiper parámetros de SARIMA que se desean.

```
Programa de predicción
Inserte la ruta del archivo:
/Users/angelcapote/Downloads/TFG/CSV3
Inserte el nombre del archivo:
Art1.csv
Introduzca el nombre con los datos de random forest.
Art1_rf.csv
```

Figura 3.3: Demo iniciación programa.

Una vez introducida la ruta y los ficheros necesarios se muestran las opciones de la figura 3.4.

```
1. Hacer una predicción.
2. Imprimir el resultado de las validaciones.
3. Salir
Elige una:
```

Figura 3.4: Menú de opciones.

Si la elección es “Imprimir el resultado de las validaciones” el resultado será el ilustrado en la figura 3.5.

Fecha Contable	Unidades Vendidas	Autoregresión	Prophet	SARIMA	Random Forest	Error_autoreg%	Error_sarima%	Error_Prophet%	Error_RandomForest%
2019-01-31	6199	6216.846877	7426.052566	7643.282185	7099.275	0.287899	19.794363	23.298632	14.522907
2019-02-28	6519	6045.578246	7832.378519	7557.051062	6605.642	7.262184	20.146932	15.923471	1.329069
2019-03-31	7930	7867.300620	10186.067210	9122.401035	7575.568	0.790661	28.449776	15.036583	4.469508
2019-04-30	7738	6405.670943	7824.483435	8032.224180	7388.819	17.218003	1.117646	3.802329	4.512548
2019-05-31	7122	5941.808585	7767.586010	7793.460597	7100.474	16.571067	9.064673	9.427978	0.302247
2019-06-30	5990	5960.526038	7336.609484	7850.400486	7291.062	0.492053	22.480960	31.058439	21.720568
2019-07-31	6067	6355.640040	9246.964391	8566.699704	7233.784	4.757541	52.414116	41.201577	19.231647
2019-08-31	5641	5959.383346	8779.344091	8000.054624	7871.716	5.644094	55.634534	41.819795	39.544691
2019-09-30	6264	5183.924283	8374.478348	7291.561329	7173.560	17.242588	33.692183	16.404236	14.520434
2019-10-31	7116	5375.851183	8016.220975	7495.849111	7502.797	24.454031	12.650660	5.337958	5.435596
2019-11-30	6737	5115.574808	6876.787239	6926.091361	7329.299	24.067466	2.074918	2.806759	8.791732
2019-12-31	6364	5384.586212	7932.778634	7031.608631	7012.140	15.389909	24.650827	10.490393	10.184475

Figura 3.5: Resultado validaciones

Si se desea realizar una predicción, se deben insertar los ficheros necesarios, así como los valores de SARIMA y el resultado es el de la figura 3.6.

```
<Predict.Sarima.Sarima object
2020-01-31      6755.823613
2020-02-29      6751.945445
2020-03-31      8298.029161
2020-04-30      7340.524057
2020-05-31      7045.102963
2020-06-30      6913.525527
2020-07-31      7530.320587
2020-08-31      6985.400899
2020-09-30      6483.851070
2020-10-31      6793.344838
2020-11-30      6244.779132
2020-12-31      6283.751754
```

Figura 3.6: Resultado predicción SARIMA

Si se desea volver a probar con parámetros diferentes de SARIMA, con características diferentes para Random Forest, o en un intervalo diferente, solo es necesario seleccionar la opción de “Hacer una predicción” e insertar los valores.

La ventaja de este programa es que te permite comparar cuál es el modelo que presenta un mejor desempeño para tus datos y también te permite realizar predicciones con todos los modelos. En caso de que en un futuro se deseen añadir más métodos, es posible, ya que solo se debe heredar de “Apredict” e implementar la función de predicción, lo mismo es aplicable para la carga de datos.

## 6. Conclusiones y líneas futuras

En este trabajo se ha abordado el problema de la predicción de ventas de la empresa La Gaviota Alimentación, a partir de 4 de sus productos, empleando diversas técnicas de estimación de series temporales y algoritmos de machine learning.

Tras el estudio de los datos y los modelos disponibles, se ha evaluado la calidad de estos como medio para predecir las ventas de los productos considerados en este estudio. Se muestran, por tanto, para cada producto la previsión de ventas y la desviación con respecto al valor real.

Los resultados obtenidos indican que el desempeño de cada modelo depende de las características del producto y, por tanto, de la naturaleza de los datos. En líneas generales, SARIMA y la autoregresión han tenido el mejor rendimiento, ya que en productos con estacionalidad muy marcada la media móvil presenta una desviación muy elevada. Prophet y Random Forest no presentan el mejor desempeño, pero en el producto con la estacionalidad muy marcada, presentan mejor desempeño que la media móvil. Además, tanto Random Forest como Prophet, lograron estar a la altura de SARIMA en múltiples escenarios.

Una posible línea de investigación y análisis futuros sería la de considerar nuevos atributos de predicción para modelos como el Random forest. También se podrían estudiar otros algoritmos de Machine Learning. Sin embargo, una de las más interesantes sería combinar la agrupación de productos con la predicción de las ventas en cada grupo. En primer lugar, se crearían grupos de productos que comparten características similares para establecer, luego, el modelo que mejor predice en cada grupo. Una vez hecho esto, se aplicaría una técnica de clustering para asignar cada artículo a uno de los clusters. Así, cada artículo sería evaluado por el modelo que mejor se adapte a él y mejor desempeño presente.

# 7. Summary and Conclusions

In this work, the problem of predicting sales of the company La Gaviota Alimentación has been addressed, based on 4 of its products, using various time series estimation techniques and machine learning algorithms.

After studying the available data and models, their quality has been evaluated as a way of predicting the sales of the products considered in this study. Therefore, the sales forecast and the deviation from the actual value are shown for each product.

The results obtained indicate that the performance of each model depends a lot on the nature of the data, but in general terms, SARIMA and the autoregression have had the best performance, since in products with very marked seasonality the moving average presents a very high deviation. Prophet and Random Forest do not present the best performance, but in the product with very marked seasonality, they present better performance than the moving average. Furthermore, both Random Forest and Prophet managed to hold their own against SARIMA in multiple scenarios.

A possible line to continue with the investigation would be to have more features from the products for Random Forest. Other machine learning algorithms could also be studied. However, one of the most interesting would be to create groups of products that share similar characteristics to establish the model that works best for each group. Once this is done, apply a clustering technique to assign each item to one of the clusters. Thus, each article would be evaluated by the model that best suits it and presents a better performance.

# 8. Tecnologías

Para el desarrollo de este proyecto se han usado las siguientes tecnologías:

El entorno de desarrollo para el estudio ha sido jupyter notebook [25], en los entornos de anaconda, distribución de Python [15] y R [17] para computación científica. Entre los motivos por los que se escogió esta herramienta son: la simplicidad de la gestión de paquetes y el uso de los recursos locales. Además se ha usado Visual Studio code, también bajo los environment de conda [25], para el desarrollo del prototipo.

Entre las librerías usadas en este proyecto destacan:

- **Pandas [13]**, librería de python usada para la manipulación y análisis de datos, es la más usada en el ámbito de Machine Learning.
- **Scikit-learn [19]**, es una librería de python que implementa algoritmos de clasificación de machine learning, usada para el Random Forest.
- **Matplotlib [12]**, es una librería de representación de python, usada para el estudio de los datos y análisis de los resultados.
- **Statsmodels [20][21]**, es una librería de python con funciones de estimación de modelos estadísticos, usada para la autoregresión y el SARIMA.
- **Prophet [16]**, es una librería de python, open source que implementa el método de estimación Prophet.
- **Seastest [9]**, es una librería de R que tiene diversos test para comprobar la estacionalidad de una serie temporal.

## 9. Presupuesto

Todas las tecnologías usadas en este proyecto son open source, por lo que lo único necesario para llevarlo a cabo es un ordenador y el tiempo de trabajo de un ingeniero informático junior.

Para desarrollar este Trabajo de Fin de Grado han sido necesarias cerca de 230 horas, por lo que si establecemos que el salario medio de un junior es de 20€ la hora, el coste del proyecto sería de 4600€.

# 10. Apéndice 1: Datos de las ventas

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
2014	8580	7189	7948	6935	6552	6074	7945	6156	5513	6949	5760	6029
2015	5807	5592	9556	6642	5884	5112	6818	6774	6435	5396	4502	5285
2016	5218	4890	6873	6902	6070	6024	6496	5455	5175	5842	5684	5184
2017	5053	5776	6817	5763	5619	7203	5844	6546	4431	5435	6319	5889
2018	6469	7280	8124	6992	7832	7865	9176	8287	7892	7194	5763	6880
2019	6199	6519	7930	7738	7122	5990	6067	5641	6264	7116	6737	6364

Tabla 1.1: Artículo 1, ventas

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
2014	8622	6125	6602	4393	7303	6782	7192	7771	7274	8778	8681	7581
2015	6901	4045	6970	7038	5999	5329	5283	4818	5149	7565	5535	5920
2016	5979	6186	6745	6876	6374	6382	6589	6069	6530	7374	6962	7380
2017	7438	7374	8137	6165	6518	6479	6452	6868	6872	7980	8057	6728
2018	5913	6551	7078	6348	7500	6104	7485	6912	7226	8440	8367	7069
2019	7767	8404	7983	8393	7747	7010	6675	7857	6530	8285	10094	8678

Tabla 1.2: Artículo 2, ventas

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
2014	356	220	212	36	728	309	188	344	1934	996	554	1335
2015	495	928	1179	1572	1107	212	73	81	1030	1248	1155	2312
2016	1519	1438	2032	1430	565	279	800	538	718	1082	1076	1882
2017	1505	1835	2578	796	1106	426	380	219	1668	411	1317	1084
2018	1070	127	170	253	230	41	238	346	1340	573	1563	2235
2019	1452	1531	1778	2093	581	396	491	842	2276	3178	3361	2488

Tabla 1.3: Artículo 3, ventas

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
2014	40	60	210	100	60	138	98	109	83	190	152	195
2015	238	215	254	269	207	346	260	280	155	412	435	279
2016	430	399	389	608	404	418	546	557	681	508	840	623
2017	659	783	681	759	678	806	672	674	591	924	1040	700
2018	687	1157	1253	882	1163	804	909	1031	1119	1612	1254	1394
2019	1620	2047	2158	2253	1964	1718	2164	2440	1783	2710	2908	2612

Tabla 1.3: Artículo 4, ventas

# 11. Apéndice 2: Tablas con la estimación realizada

## 11.1. Artículo 1

Fecha Contable	Media Móvil			Autoregresión		SARIMA		Prophet		Random Forest	
	Unidades Vendidas	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación
31/1/19	6199	6612.33	6.66	6216.84	0.28	7171.58	15.69	7426.05	19.79	7099.27	14.52
28/2/19	6519	6418.44	1.54	6045.57	7.26	7170.51	9.99	7832.37	20.14	6605.64	1.32
31/3/19	7930	6636.92	16.30	7867.30	0.79	7170.51	9.58	10186.06	28.44	7575.56	4.46
30/4/19	7738	6555.90	15.27	6405.67	17.21	7170.51	7.33	7824.48	1.11	7388.81	4.51
31/5/19	7122	6537.09	8.21	5941.80	16.57	7170.51	0.68	7767.58	9.06	7100.47	0.30
30/6/19	5990	6576.63	9.79	5960.52	0.49	7170.51	19.71	7336.60	22.48	7291.06	21.72
31/7/19	6067	6556.54	8.06	6355.64	4.75	7170.51	18.19	9246.96	52.41	7233.78	19.23
31/8/19	5641	6556.75	16.23	5959.38	5.64	7170.51	27.11	8779.34	55.63	7871.71	39.54
30/9/19	6264	6563.31	4.77	5183.92	17.24	7170.51	14.47	8374.47	33.69	7173.56	14.52
31/10/19	7116	6558.87	7.82	5375.85	24.45	7170.51	0.77	8016.22	12.65	7502.797	5.43
30/11/19	6737	6559.64	2.63	5115.57	24.06	7170.51	6.43	6876.78	2.07	7329.299	8.79
31/12/19	6364	6560.61	3.08	5384.58	15.38	7170.51	12.67	7932.77	24.65	7012.14	10.18
			8.36		11.18		11.89		23.51		12.04

Tabla 2.1: Artículo 1, Estimación anual

Fecha Contable	Media Móvil			Autoregresión		Sarima		Prophet		Random Forest	
	Unidades Vendidas	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación
31/1/19	6199	6612.33	6.67	6216.85	0.29	7171.58	15.69	7426.05	19.79	7099.28	14.52
28/2/19	6519	6418.44	1.54	6045.58	7.26	7170.51	9.99	7832.38	20.15	6605.64	1.33
31/3/19	7930	6636.93	16.31	7867.30	0.79	7170.51	9.58	10186.07	28.45	7575.57	4.47
30/4/19	7738	6555.90	15.28	6405.67	17.22	7170.51	7.33	7824.48	1.12	7388.82	4.51
31/5/19	7122	6537.09	8.21	5941.81	16.57	7170.51	0.68	7767.59	9.06	7100.47	0.30
30/6/19	5990	6576.64	9.79	5960.53	0.49	7170.51	19.71	7336.61	22.48	7291.06	21.72
31/7/19	6067	6950.00	14.55	6456.53	6.42	6707.64	10.56	8445.36	39.20	7330.99	20.83
31/8/19	5641	6687.33	18.55	6115.48	8.41	6926.63	22.79	7907.09	40.17	8100.09	43.59
30/9/19	6264	6542.44	4.45	5327.76	14.95	6926.63	10.58	7466.29	19.19	7375.28	17.74
31/10/19	7116	6726.59	5.47	5467.48	23.17	6926.63	2.66	7099.00	0.24	7763.33	9.10
30/11/19	6737	6652.12	1.26	5214.16	22.60	6926.63	2.81	5973.13	11.34	7689.87	14.14
31/12/19	6364	6640.39	4.34	5466.45	14.10	6926.63	8.84	7052.75	10.82	7433.87	16.81
			8.87		11.02		10.10		18.50		14.09

Tabla 2.2: Artículo 1, Estimación semestral

		Media Móvil		Autoregresión		Sarima		Prophet		Random Forest	
Fecha	Unidades										
Contable	Vendidas	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación
31/1/19	6199	6612.33	6.67	6216.85	0.29	7171.58	15.69	7426.05	19.79	7099.28	14.52
28/2/19	6519	6418.44	1.54	6045.58	7.26	7170.51	9.99	7832.38	20.15	6605.64	1.33
31/3/19	7930	6636.93	16.31	7867.30	0.79	7170.51	9.58	10186.07	28.45	7575.57	4.47
30/4/19	7738	6882.67	11.05	6441.43	16.76	7329.08	5.28	7300.41	5.66	7368.24	4.78
31/5/19	7122	7110.56	0.16	5985.76	15.95	7137.39	0.22	7160.99	0.55	6816.42	4.29
30/6/19	5990	7307.74	22.00	6005.05	0.25	7137.39	19.16	6645.03	10.94	7460.13	24.54
31/7/19	6067	6950.00	14.55	6456.53	6.42	6707.64	10.56	8445.36	39.20	7330.99	20.83
31/8/19	5641	6687.33	18.55	6115.48	8.41	6926.63	22.79	7907.09	40.17	8100.09	43.59
30/9/19	6264	6542.44	4.45	5327.76	14.95	6926.63	10.58	7466.29	19.19	7375.28	17.74
31/10/19	7116	5990.67	15.81	6031.56	15.24	6488.04	8.82	6677.16	6.17	6864.84	3.53
30/11/19	6737	5965.22	11.46	5515.48	18.13	6500.16	3.52	5439.13	19.26	6854.90	1.75
31/12/19	6364	6073.30	4.57	5669.99	10.91	6500.16	2.14	6385.31	0.33	6633.58	4.24
		10.59		9.61		9.86		17.49		12.13	

Tabla 2.3: Artículo 1, Estimación trimestral

## 11.2. Artículo 2

		Media Móvil		Autoregresión		SARIMA		Prophet		Random Forest	
Fecha	Unidades										
Contable	Vendidas	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación
31/1/19	7767	7958.67	2.47	7009.41	9.75	7789.15	0.29	7195.16	7.36	6174.44	20.50
28/2/19	8404	7798.22	7.21	6079.31	27.66	6973.77	17.02	5904.88	29.74	6335.33	24.62
31/3/19	7983	7608.63	4.69	7274.09	8.88	7910.48	0.91	7672.04	3.90	6387.44	19.99
30/4/19	8393	7788.51	7.20	6626.70	21.04	7067.54	15.79	6608.49	21.26	6411.51	23.61
31/5/19	7747	7731.79	0.20	6354.32	17.98	7574.08	2.23	7301.28	5.75	6407.00	17.30
30/6/19	7010	7709.64	9.98	6070.93	13.40	7106.51	1.38	6256.78	10.74	6353.63	9.36
31/7/19	6675	7743.31	16.00	6201.70	7.09	7452.97	11.66	6710.65	0.53	6328.40	5.19
31/8/19	7857	7728.25	1.64	5968.37	24.04	7363.22	6.28	6604.13	15.95	6858.16	12.71
30/9/19	6530	7727.07	18.33	6191.75	5.18	7500.35	14.86	6695.89	2.54	7038.62	7.79
31/10/19	8285	7732.87	6.66	7662.57	7.51	8790.17	6.10	8860.52	6.95	7412.67	10.53
30/11/19	10094	7729.40	23.43	6891.22	31.73	8299.66	17.78	7813.95	22.59	7666.96	24.04
31/12/19	8678	7729.78	10.93	6609.45	23.84	7596.49	12.46	7213.09	16.88	7190.46	17.14
		9.06		16.51		8.90		12.02		16.07	

Tabla 2.5: Artículo 2, Estimación anual

		Media Móvil		Autoregresión		SARIMA		Prophet		Random Forest	
Fecha	Unidades										
Contable	Vendidas	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación
31/1/19	7767	7958.67	2.47	7009.41	9.75	7789.15	0.29	7195.16	7.36	6174.44	20.50
28/2/19	8404	7798.22	7.21	6079.31	27.66	6973.77	17.02	5904.88	29.74	6335.33	24.62
31/3/19	7983	7608.63	4.69	7274.09	8.88	7910.48	0.91	7672.04	3.90	6387.44	19.99
30/4/19	8393	7788.51	7.20	6626.70	21.04	7067.54	15.79	6608.49	21.26	6411.51	23.61
31/5/19	7747	7731.79	0.20	6354.32	17.98	7574.08	2.23	7301.28	5.75	6407.00	17.30
30/6/19	7010	7709.64	9.98	6070.93	13.40	7106.51	1.38	6256.78	10.74	6353.63	9.36
31/7/19	6675	7716.67	15.61	6733.07	0.87	7738.22	15.93	7012.13	5.05	7319.84	9.66
31/8/19	7857	7491.22	4.66	6263.61	20.28	7635.64	2.82	6962.82	11.38	7505.04	4.48
30/9/19	6530	7405.96	13.41	6356.07	2.66	7781.49	19.17	7083.50	8.48	7708.16	18.04
31/10/19	8285	7537.95	9.02	7771.29	6.20	9064.96	9.41	9241.40	11.54	7152.73	13.67
30/11/19	10094	7478.38	25.91	6944.97	31.20	8596.08	14.84	8155.08	19.21	7849.93	22.23
31/12/19	8678	7474.10	13.87	6702.76	22.76	7878.82	9.21	7485.37	13.74	7453.93	14.11
		9.52		15.22		9.08		12.35		16.46	

Tabla 2.6: Artículo 2, Estimación semestral

		Media Móvil		Autoregresión		SARIMA		Prophet		Random Forest	
Fecha	Unidades										
Contable	Vendidas	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación
31/1/19	7767	7958.67	2.47	7009.41	9.75	7789.15	0.29	7195.16	7.36	6174.44	20.50
28/2/19	8404	7798.22	7.21	6079.31	27.66	6973.77	17.02	5904.88	29.74	6335.33	24.62
31/3/19	7983	7608.63	4.69	7274.09	8.88	7910.48	0.91	7672.04	3.90	6387.44	19.99
30/4/19	8393	8051.33	4.07	6981.59	16.82	7440.94	11.34	6644.56	20.83	7546.43	10.09
31/5/19	7747	8146.11	5.15	6530.61	15.70	7949.63	2.62	7399.79	4.48	7766.28	0.25
30/6/19	7010	8060.15	14.98	6152.95	12.23	7464.18	6.48	6412.63	8.52	7487.82	6.82
31/7/19	6675	7716.67	15.61	6733.07	0.87	7738.22	15.93	7012.13	5.05	7319.84	9.66
31/8/19	7857	7491.22	4.66	6263.61	20.28	7635.64	2.82	6962.82	11.38	7505.04	4.48
30/9/19	6530	7405.96	13.41	6356.07	2.66	7781.49	19.17	7083.50	8.48	7708.16	18.04
31/10/19	8285	7020.67	15.26	7885.01	4.83	8463.67	2.16	9213.33	11.21	7234.20	12.68
30/11/19	10094	7135.89	29.31	7006.39	30.59	7973.71	21.01	8129.37	19.46	7775.95	22.96
31/12/19	8678	6895.52	20.54	6765.80	22.04	7269.94	16.23	7460.01	14.04	7302.46	15.85
		11.45		14.36		9.66		12.04		13.83	

Tabla 2.7: Artículo 2, Estimación trimestral

## 11.3. Artículo 3

Fecha Contable	Unidades Vendidas	Media Móvil		Autoregresión		SARIMA		Prophet		Random Forest	
		Predicción anual	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación
31/1/19	1452	1457.00	0.34	1361.46	6.24	1366.11	5.92	1167.05	19.62	1208.01	16.80
28/2/19	1531	1751.67	14.41	1322.84	13.60	1014.93	33.71	768.92	49.78	1572.18	2.69
31/3/19	1778	1814.56	2.06	1824.79	2.63	1109.47	37.60	870.96	51.01	1765.94	0.68
30/4/19	2093	1674.41	20.00	1275.54	39.06	1239.68	40.77	1141.47	45.46	1125.76	46.21
31/5/19	581	1746.88	200.67	981.98	69.02	885.61	52.43	1229.80	111.67	1505.39	159.10
30/6/19	396	1745.28	340.73	388.06	2.01	488.07	23.25	463.47	17.04	725.79	83.28
31/7/19	491	1722.19	250.75	463.10	5.68	656.71	33.75	173.19	64.73	495.63	0.94
31/8/19	842	1738.11	106.43	327.02	61.16	656.03	22.09	218.59	74.04	389.60	53.73
30/9/19	2276	1735.19	23.76	1192.54	47.60	1360.13	40.24	1501.00	34.05	840.47	63.07
31/10/19	3178	1731.83	45.51	911.73	71.31	1364.91	57.05	951.66	70.05	879.97	72.31
30/11/19	3361	1735.05	48.38	1231.03	63.37	1277.49	61.99	1150.38	65.77	1331.44	60.39
31/12/19	2488	1734.02	30.30	1755.32	29.45	2258.69	9.22	2159.87	13.19	1559.29	37.33
			90.28		34.26		34.83		51.37		49.71

Tabla 2.9: Artículo 3, Estimación anual

Fecha Contable	Unidades Vendidas	Media Móvil		Autoregresión		SARIMA		Prophet		Random Forest	
		Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación
31/1/19	1452	1457.00	0.34	1361.46	6.24	1366.11	5.92	1167.05	19.62	1208.01	16.80
28/2/19	1531	1751.67	14.41	1322.84	13.60	1014.93	33.71	768.92	49.78	1572.18	2.69
31/3/19	1778	1814.56	2.06	1824.79	2.63	1109.47	37.60	870.96	51.01	1765.94	0.68
30/4/19	2093	1674.41	20.00	1275.54	39.06	1239.68	40.77	1141.47	45.46	1125.76	46.21
31/5/19	581	1746.88	200.67	981.98	69.02	885.61	52.43	1229.80	111.67	1505.39	159.10
30/6/19	396	1745.28	340.73	388.06	2.01	488.07	23.25	463.47	17.04	725.79	83.28
31/7/19	491	1023.33	108.42	456.61	7.00	687.46	40.01	299.54	38.99	511.74	4.22
31/8/19	842	666.78	20.81	321.18	61.86	683.99	18.77	364.46	56.72	383.44	54.46
30/9/19	2276	695.37	69.45	1204.46	47.08	1291.54	43.25	1645.42	27.71	885.30	61.10
31/10/19	3178	795.16	74.98	889.03	72.03	1449.90	54.38	1073.34	66.23	926.23	70.85
30/11/19	3361	719.10	78.60	1199.77	64.30	1334.20	60.30	1234.25	63.28	1076.18	67.98
31/12/19	2488	736.54	70.40	1740.82	30.03	2383.93	4.18	2198.89	11.62	1328.93	46.59
			83.41		34.57		34.55		46.59		51.16

Tabla 2.10: Artículo 3, Estimación semestral

Fecha Contable	Unidades Vendidas	Media Móvil		Autoregresión		SARIMA		Prophet		Random Forest	
		Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación
31/1/19	1452	1457.00	0.34	1361.46	6.24	1366.11	5.92	1167.05	19.62	1208.01	16.80
28/2/19	1531	1751.67	14.41	1322.84	13.60	1014.93	33.71	768.92	49.78	1572.18	2.69
31/3/19	1778	1814.56	2.06	1824.79	2.63	1109.47	37.60	870.96	51.01	1765.94	0.68
30/4/19	2093	1587.00	24.18	1259.87	39.81	1734.17	17.14	1151.20	45.00	1290.14	38.36
31/5/19	581	1632.00	180.90	973.24	67.51	1319.48	127.10	1270.53	118.68	1530.49	163.42
30/6/19	396	1665.67	320.62	384.14	2.99	871.68	120.12	537.90	35.83	785.08	98.25
31/7/19	491	1023.33	108.42	456.61	7.00	687.46	40.01	299.54	38.99	511.74	4.22
31/8/19	842	666.78	20.81	321.18	61.86	683.99	18.77	364.46	56.72	383.44	54.46
30/9/19	2276	695.37	69.45	1204.46	47.08	1291.54	43.25	1645.42	27.71	885.30	61.10
31/10/19	3178	1203.00	62.15	1342.57	57.75	2005.58	36.89	1061.01	66.61	1428.43	55.05
30/11/19	3361	1440.33	57.15	1397.28	58.43	1716.45	48.93	1247.22	62.89	1430.57	57.44
31/12/19	2488	1639.78	34.09	1828.14	26.52	2643.63	6.26	2242.79	9.86	1572.00	36.82
			74.55		32.62		44.64		48.56		49.11

Tabla 2.11: Artículo 3, Estimación trimestral

## 11.4. Artículo 4

Fecha Contable	Unidades Vendidas	Media Móvil		Autoregresión		SARIMA		Prophet		Random Forest	
		Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación
31/1/19	1620	1420.00	12.35	1301.90	19.64	1173.16	27.58	1168.72	27.86	1259.80	22.23
28/2/19	2047	1356.00	33.76	1615.31	21.09	1647.21	19.53	1275.83	37.67	1254.60	38.71
31/3/19	2158	1390.00	35.59	1433.75	33.56	1805.22	16.35	1373.16	36.37	1276.75	40.84
30/4/19	2253	1388.67	38.36	1452.64	35.52	1386.95	38.44	1194.26	46.99	1257.09	44.20
31/5/19	1964	1378.22	29.83	1311.88	33.20	1683.62	14.28	1289.32	34.35	1270.90	35.29
30/6/19	1718	1385.63	19.35	1185.61	30.99	1327.78	22.71	1297.29	24.49	1279.07	25.55
31/7/19	2164	1384.17	36.04	1283.32	40.70	1425.53	34.13	1247.06	42.37	1277.06	40.99
31/8/19	2440	1382.67	43.33	1451.47	40.51	1551.92	36.40	1307.91	46.40	1278.87	47.59
30/9/19	1783	1384.16	22.37	1738.53	2.49	1638.93	8.08	1229.86	31.02	1275.22	28.48
31/10/19	2710	1383.67	48.94	1878.34	30.69	2131.29	21.35	1613.04	40.48	1375.49	49.24
30/11/19	2908	1383.50	52.42	1902.66	34.57	1774.08	38.99	1426.02	50.96	1342.11	53.85
31/12/19	2612	1383.78	47.02	1782.92	31.74	1913.70	26.73	1422.18	45.55	1335.26	48.88
			34.95		29.56		25.38		38.71		39.65

Tabla 2.13: Artículo 4, Estimación anual

		Media Móvil		Autoregresión		SARIMA		Prophet		Random Forest	
Fecha	Unidades										
Contable	Vendidas	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación
31/1/19	1620	1420.00	12.35	1301.90	19.64	1173.16	27.58	1168.72	27.86	1259.80	22.23
28/2/19	2047	1356.00	33.76	1615.31	21.09	1647.21	19.53	1275.83	37.67	1254.60	38.71
31/3/19	2158	1390.00	35.59	1433.75	33.56	1805.22	16.35	1373.16	36.37	1276.75	40.84
30/4/19	2253	1388.67	38.36	1452.64	35.52	1386.95	38.44	1194.26	46.99	1257.09	44.20
31/5/19	1964	1378.22	29.83	1311.88	33.20	1683.62	14.28	1289.32	34.35	1270.90	35.29
30/6/19	1718	1385.63	19.35	1185.61	30.99	1327.78	22.71	1297.29	24.49	1279.07	25.55
31/7/19	2164	1978.33	8.58	1680.72	22.33	1932.98	10.68	1760.41	18.65	1862.76	13.92
31/8/19	2440	1886.78	22.67	1764.26	27.69	1946.43	20.23	1878.52	23.01	1951.86	20.01
30/9/19	1783	1861.04	4.38	1983.32	11.23	2070.05	16.10	1842.57	3.34	1897.29	6.41
31/10/19	2710	1908.72	29.57	2173.82	19.79	2574.75	4.99	2248.78	17.02	1795.64	33.74
30/11/19	2908	1885.51	35.16	2263.56	22.16	2197.58	24.43	2063.16	29.05	1889.83	35.01
31/12/19	2612	1885.09	27.83	2283.39	12.58	2346.37	10.17	2044.31	21.73	1881.18	27.98
			24.78		24.15		18.79		26.71		28.66

Tabla 2.14: Artículo 4, Estimación semestral

		Media Móvil		Autoregresión		SARIMA		Prophet		Random Forest	
Fecha	Unidades										
Contable	Vendidas	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación	Predicción	Desviación
31/1/19	1620	1420.00	12.35	1301.90	19.64	1173.16	27.58	1168.72	27.86	1259.80	22.23
28/2/19	2047	1356.00	33.76	1615.31	21.09	1647.21	19.53	1275.83	37.67	1254.60	38.71
31/3/19	2158	1390.00	35.59	1433.75	33.56	1805.22	16.35	1373.16	36.37	1276.75	40.84
30/4/19	2253	1941.67	13.82	1919.21	14.82	1789.58	20.57	1437.30	36.20	1968.09	12.65
31/5/19	1964	2048.89	4.32	1686.08	14.15	2064.91	5.14	1572.65	19.93	1967.44	0.18
30/6/19	1718	2049.52	19.30	1492.02	13.15	1708.67	0.54	1622.29	5.57	1956.29	13.87
31/7/19	2164	1978.33	8.58	1680.72	22.33	1932.98	10.68	1760.41	18.65	1862.76	13.92
31/8/19	2440	1886.78	22.67	1764.26	27.69	1946.43	20.23	1878.52	23.01	1951.86	20.01
30/9/19	1783	1861.04	4.38	1983.32	11.23	2070.05	16.10	1842.57	3.34	1897.29	6.41
31/10/19	2710	2129.00	21.44	2078.38	23.31	2710.29	0.01	2369.15	12.58	1807.08	33.32
30/11/19	2908	2117.33	27.19	2174.64	25.22	2366.91	18.61	2204.15	24.20	1955.72	32.75
31/12/19	2612	2009.78	23.06	2202.76	15.67	2319.66	11.19	2207.24	15.50	1862.05	28.71
			18.87		20.16		13.88		21.74		21.97

Tabla 2.15: Artículo 4, Estimación trimestral

# Bibliografía

- [1] Aayush Bajaj (2022) [ARIMA & SARIMA: Real-World Time Series Forecasting](#)
- [2] Ángel Emilio Capote Pérez (2022) [Sales Forecasting Program](#)
- [3] [ARIMA estimators for Python — pmdarima 1.8.5 documentation](#)
- [4] Chaya Bakshi (2020) [Random Forest Regression, Level Up Coding](#)
- [5] Dimitris Effrosynidis (2020) [Forecasting Wars: Classical Forecasting Methods vs Machine Learning, Towards Data Science](#)
- [6] Encora (2019) [Fundamentals of Time Series Data and Forecasting, Better Programming](#)
- [7] [Friedman Test - StatsTest.com](#)
- [8] Hyndman, R.J., & Athanasopoulos, G. (2021) [Forecasting: principles and practice. 3rd edition. OTexts: Melbourne, Australia.](#)
- [9] [isSeasonal: Testing the seasonality of series in seastests](#)
- [10] Joaquín Amat Rodrigo (2021) [Skforecast: forecasting series temporales con Python y Scikit Learn](#) –revisar
- [11] Liudmyla Taranenka (2021) [Machine Learning Demand Forecasting Methods For Sales Prediction in 2021](#)
- [12] [matplotlib.pyplot — Matplotlib 3.1.2 documentation](#)
- [13] [pandas 1.4.2 documentation](#)
- [14] Paula Rodó (2019) [Modelo autoregresivo \(AR\). Economedia.com](#) -revisar

- [15] [Python Docs](#)
- [16] [Quick Start | Prophet](#)
- [17] [R para principiantes](#)
- [18] Ravindra Elicherla (2019) [\*Retail sales forecast using Facebook's Prophet\*](#)
- [19] [sklearn.ensemble.RandomForestRegressor — scikit-learn 1.1.1 documentation](#)
- [20] [statsmodels.tsa.ar\\_model.AutoReg](#)
- [21] [statsmodels.tsa.statespace.sarimax.SARIMAX](#)
- [22] Taylor SJ, Letham B. (2017) [\*Forecasting at scale. PeerJ Preprints 5:e3190v2\*](#)
- [23] Thanakorn Panyapiang (2022) [\*Time-series forecasting using ordinary Machine Learning algorithms. Geek Culture\*](#)
- [24] Tony Yiu (2019) [\*Understanding Random Forest. How the Algorithm Works and Why Is So Effective, Towards Data Science\*](#)
- [25] [Using Jupyter Notebook — Anaconda documentation](#)
- [26] Will Koehrsen (2017) [\*Random Forest in Python. A Practical End-To-End Machine Learning Example, Towards Data Science\*](#)