



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Informática

Modelos para la fijación de precios en el mercado de  
cítricos y frutas.

Trabajo Fin de Grado

Grado en Ingeniería Informática

AUTOR/A: Palmero Roldán, Héctor

Tutor/a: Botti Navarro, Vicente Juan

Cotutor/a: Palanca Cámara, Javier

CURSO ACADÉMICO: 2021/2022



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



Escola Tècnica  
Superior d'Enginyeria  
Informàtica

Escola Tècnica Superior d'Enginyeria Informàtica

Universitat Politècnica de València

# Modelos para la fijación de precios en el mercado de cítricos y frutas

Trabajo Fin de Grado

Grado en Ingeniería Informática

**Autor:** Héctor Palmero Roldán

**Tutores:** Javier Palanca Camará

Vicente Juan Botti Navarro

2021/2022



# Resumen

---

La actual situación económica fruto de la salida de la pandemia y del complejo equilibrio geoestratégico ha dado lugar a un panorama complejo lleno de incertidumbres para todos los sectores de la economía. Esta preocupación afecta en especial al ámbito de la agricultura, ya castigado durante los últimos años, y genera un gran impacto en la Comunidad Valenciana a través de un producto estrella, la naranja.

A su vez y en relación con hacer frente a la incertidumbre, durante los últimos años han ido apareciendo nuevas herramientas informáticas, o adaptaciones de ya existentes, capaces de aportar con una relativa facilidad conocimientos muy relevantes que son capaces de aportar significativamente a la toma de decisiones en las empresas o entidades que los usen.

En este trabajo se presentarán algunas de estas interesantes herramientas, una explicación del mencionado sector de la naranja, así como sus problemas en el país y la comunidad, y una combinación de ambas cosas. La explicación del contexto y posterior puesta en común servirá tanto para estudiar las capacidades de estas tecnologías como para mostrar cómo pueden usarse para beneficiar a sectores relevantes para la agricultura y la economía.

**Palabras clave:** naranjas y cítricos, series temporales, predicción, ARIMA, LSTM.

# Abstract

---

The current economic situation resulting from the end of the pandemic and the complex geostrategic balance has given rise to a complex panorama full of uncertainties for all sectors of the economy. This concern particularly affects the agricultural sector, which has already been hard hit in recent years, and has had a major impact on the Valencian Community through its star product, the orange.

At the same time, and in relation to dealing with uncertainty, in recent years new computer tools have been appearing, or adaptations of existing ones, capable of providing very relevant knowledge with relative ease, which can make a significant contribution to decision-making in the companies or entities that use them.

This paper will present some of these interesting tools, an explanation of the orange sector, as well as its problems in the country and the community, and a combination of both. The explanation of the context and subsequent sharing will serve both to explore the capabilities of these technologies and to show how they can be used to benefit sectors relevant to agriculture and the economy.

**Keywords :** oranges and citrus, time series, forecasting, ARIMA, LSTM.

# Tabla de contenidos

---

1. Introducción	7
1.1 Motivación	7
1.2 Objetivo del proyecto	8
1.3 Estructura de la Memoria	9
2. Estado del Arte	11
2.1 Elección del lenguaje: Python y sus herramientas	11
2.2 Tecnologías de fijación tradicionales	14
2.3 Tecnologías de fijación vanguardistas	23
3. Análisis del problema	30
3.1 Evaluación del mercado de cítricos	30
3.2 Problemas relacionados con el objetivo de fijación de precios	32
3.3 Solución propuesta	34
4. Diseño y desarrollo de los Modelos	36
4.1 Consideraciones y tratamiento previo	36
4.2 Modelos tradicionales	37
4.3 Modelos vanguardistas	43
5. Pruebas y análisis de resultados	49
5.1 Análisis y testeo de los modelos obtenidos	49
5.2 Análisis sobre resultados	55
6. Conclusiones y Trabajo Futuro	59
6.1 Conclusiones	59
6.2 Relación con estudios cursados	60
6.3 Trabajo Futuro	61
Referencias bibliográficas	62
Anexo: Objetivos de desarrollo sostenible	66





## 1. Introducción

El mercado de la producción de frutas, y más concretamente de cítricos, es uno de los pilares fundamentales del sector primario en España, y especialmente en la Comunidad Valenciana. A su vez, tanto el país como la región se posicionan como los principales exportadores del mundo al respecto, demostrando la importancia del sector en la economía regional.

Con esta enorme influencia en el entorno y las numerosas empresas de la zona que se dedican al mismo, se ha decidido que es el sector perfecto para desarrollar este trabajo, donde se aplicarán diversas técnicas informáticas de análisis y fijación de precios para intentar aportar valor y así beneficiar tanto a los implicados de forma directa como a la economía en general.

Esta cercanía tanto con los datos reales de la actividad como con los agentes implicados en la misma favorece realizar esta investigación y aplicación práctica de las distintas herramientas relacionadas con la carrera, así como ampliadas para este proyecto, dejando un marco idóneo para intentar hacer un trabajo de calidad.

---

### 1.1 Motivación

---

La motivación para realizar este trabajo se descompone en distintas partes que han servido como hilo para ir componiendo a través del proceso previo de idear el mismo.

La motivación inicial y básica ha sido un interés personal por las distintas técnicas y herramientas informáticas capaces de aportar información valiosa sobre un sector de la economía y ayudar en la toma de decisiones. Al ser este uno de los dos TFG para el doble grado junto con Administración y Dirección de Empresas, el interés en conseguir que lo aprendido en este grado de informática tenga un valor palpable en el apartado económico y empresarial ha sido desde siempre un punto fundamental a la hora de pensar cómo enfocar este proyecto.

Dentro de esta amplia idea de aportar a la parte estratégica, rápidamente se abandonó la idea de los sistemas de soporte como ERP y similares en favor de un campo poco explorado en la rama del doble grado, pero fundamental en lo relativo a la toma de decisiones, como son las herramientas predictivas. La idea de emplear distintas técnicas informáticas para conseguir algo tan importante y que aporte tanta ventaja a la hora de desarrollar la actividad económica, como es tener estimaciones precisas y fiables de cómo va a encontrarse el panorama en el futuro, ganó peso instantáneamente.

Y después partiendo de esto, el ya mencionado entorno favorable para desarrollar el trabajo sirvió para concretar que la mejor forma de explorar esto era enfocando a unos de los sectores económicos más importantes de la región, las naranjas y cítricos. En este marco el acceso a los datos y la capacidad de poder beneficiar no solo a una persona o empresa concreta sino a prácticamente la región en su conjunto permitió concretar finalmente los objetivos de este trabajo.

Por tanto, un resumen de la motivación es el deseo de emplear herramientas informáticas tan interesantes e importantes como las técnicas predictivas para influir de forma significativa en la toma de decisiones de un sector esencial para la Comunidad Valenciana.

---

## 1.2 Objetivo del proyecto

---

El objetivo general de este proyecto es **la obtención de un modelo útil que permita obtener unas estimaciones a futuro de los precios de los cítricos fiables**. Esta memoria recogerá el proceso de idear, plantear, implementar y testear los distintos modelos posibles y sus resultados con intención de compararlos entre ellos para obtener la mejor respuesta posible a la pregunta de cómo obtener unas predicciones fiables y efectivas para la toma de decisiones.

El objetivo último del trabajo es obtener el modelo de precios planteado, pero siempre teniendo claro que su uso estará destinado a una aplicación práctica en el mercado real y deberá cumplir con unos estándares mínimos para ello.

Como las distintas partes del trabajo, así como los pasos para conseguir este objetivo general, se tienen una serie de objetivos específicos que servirán para desmenuzar la idea general e ir avanzando por esta. Estos objetivos específicos son:

- Exponer y comprender las distintas herramientas que pueden emplearse para llevar a cabo este tipo de tareas, comparándolas entre sí a nivel teórico.
- Aprender a aplicar estas herramientas de forma práctica sobre datos reales y compararlas una vez se han obtenido resultados.
- Familiarizarse con el sector de los cítricos y analizarlo hasta un nivel adecuado para poder aplicar estas herramientas y entender los resultados.
- Generar un modelo predictivo para el precio de algunos productos del sector cítrico
- Entender y utilizar las metodologías de testeo para poder comprobar el correcto funcionamiento de los modelos planteados.

---

### 1.3 Estructura de la Memoria

---

Para finalizar el capítulo de introducción, se va a proceder a exponer brevemente qué estructura seguirá la memoria en cuanto a lo que se pretende exponer, para informar al lector de lo que espera y guiarle en cuanto a cómo va a desarrollarse el trabajo.

Primero y a continuación de esta presentación va a exponerse el estado del arte actual de las distintas herramientas que pueden emplearse para la fijación de precios mediante técnicas informáticas. Se empezará con algo básico pero necesario para entender el entorno de trabajo general, como es explicar brevemente el lenguaje de programación Python, aunque solamente de forma superficial en lo general y concretamente los entornos que se emplearán para el trabajo. Después se revisarán las tecnologías algo más “tradicionales”, es decir herramientas informáticas y librerías que pese a ser actuales y estar a la orden del día en cuanto a funcionamiento y optimización sus bases teóricas son conceptos estadísticos ya muy establecidos. Por último, se expondrán modelos menos establecidos en este ámbito y considerables como más vanguardistas, basados generalmente en inteligencia artificial y aprendizaje y no tanto en estimaciones estadísticas clásicas.



Después y antes de pasar a la aplicación práctica de este estado del arte planteado, se expondrá el análisis del problema. En este capítulo se expondrá brevemente el mercado en el que tienen lugar el trabajo, sus oportunidades y desafíos más importantes, la relación y contacto con este trabajo y la relevancia para el mismo que tiene una estimación correcta de los precios y porqué es necesario hacer este trabajo para ellos. Además, se expondrá cuáles son las propuestas de aplicación de las técnicas explicadas en el estado del arte que puedan hacer frente a estos problemas y retos a los que hace frente el sector, más aún en el marco macroeconómico actual.

Una vez expuestas las herramientas en el estado del arte y la situación del mercado de cítricos junto a sus desafíos y como se pretende relacionar ambos conceptos en el análisis del problema, se procederá en el capítulo cuatro a realizar la aplicación práctica de los distintos modelos con la intención de obtener las estimaciones de precios de las que se ha hablado hasta ahora. Se aplicarán de forma separada, pero sobre los mismos datos los modelos expuestos en el estado del arte y se obtendrán aquí las estimaciones claves para el trabajo.

Con las fijaciones de precios ya efectuadas, el capítulo cinco consistirá en las pruebas pertinentes para observar cómo de fiables son los diferentes modelos por separado. Después también será en este capítulo donde se realizará el análisis final sobre los mismos y se harán las consideraciones y decisiones pertinentes.

Finalmente, se procederá con las conclusiones finales que se han extraído del proyecto tanto en resultados del propio como por parte del alumno, y procediendo después para terminar con las referencias empleadas durante el trabajo, que por la estructura del mismo se centrarán más en los capítulos de análisis dos y tres.

## 2. Estado del Arte

Como ya se ha explicado en la estructura de la memoria, en este capítulo se va a proceder a plantear el estado de las herramientas y tecnologías relacionadas con el proyecto para dar un contexto sobre qué se va a emplear en el mismo.

---

### 2.1 Elección del lenguaje: Python y sus herramientas

---

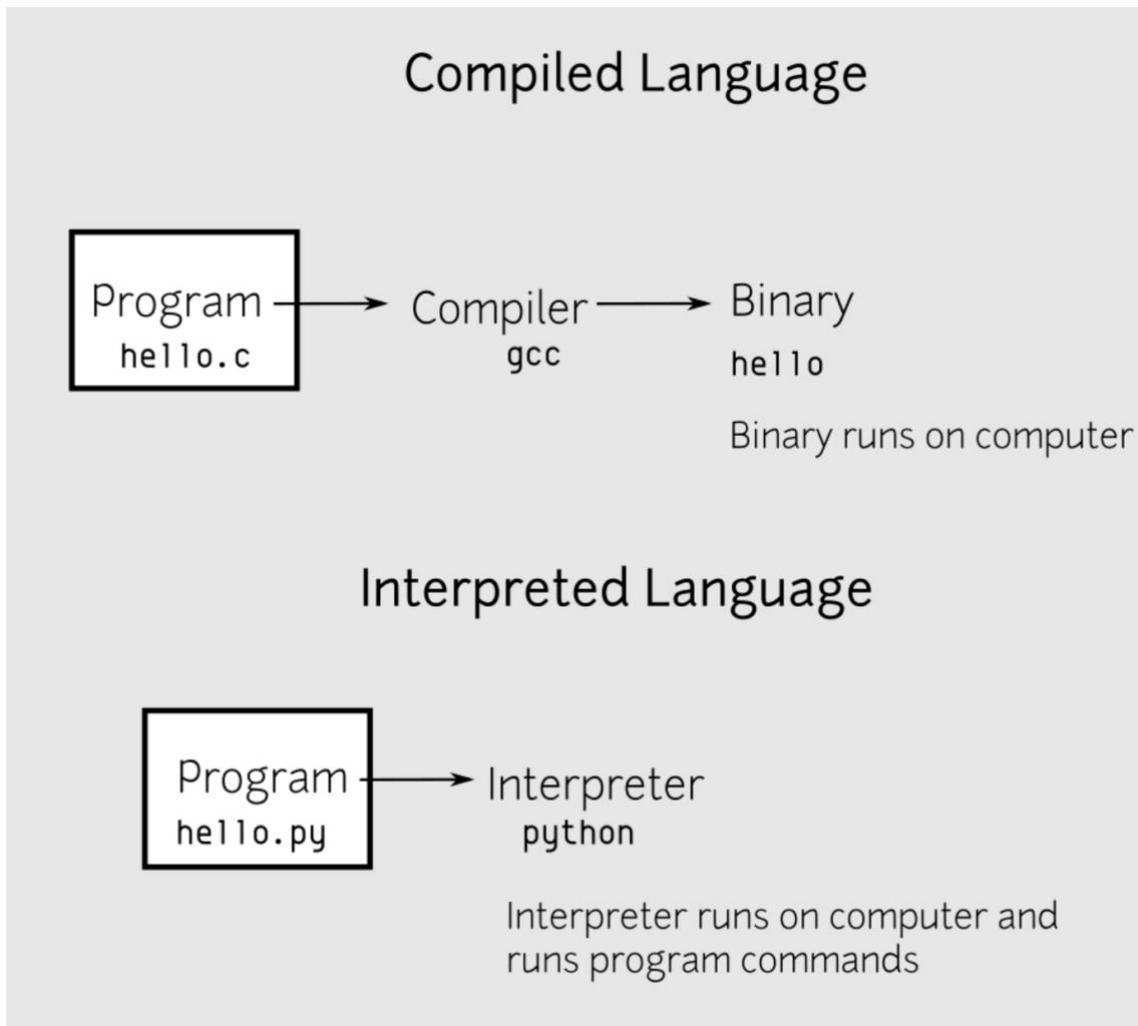
El lenguaje de programación y entorno que se empleará para el trabajo y el diseño de los modelos será Python. El que a día de hoy es uno de los lenguajes más populares y usados del mundo debe esta importancia y crecimiento reciente al factor por el cual se ha elegido para este trabajo, su óptimo funcionamiento para las tareas de análisis de datos.

Primero se van a exponer algunas de las características principales del lenguaje, para introducirlo a aquellos que no lo conozcan y establecer las bases para luego explicar por qué es el más adecuado para este proyecto.

Para empezar, es necesario entender que Python es un lenguaje de programación multiparadigma, siendo estos principalmente la orientación a objetos y el funcional, y es considerado como de alto nivel. Este concepto de multiparadigma le otorga al lenguaje una gran versatilidad, permitiéndole ser muy útil tanto para componer proyectos a mayor escala con una estructura de clases como para trabajar con múltiples scripts pequeños para tareas concretas. Este lenguaje y su creación tienen lugar en 1991 por el informático Guido Van Rossum.



Una de las características más diferenciales que tiene es que es un lenguaje interpretado, en contraposición a los compilados, lo cual le otorga una ventaja en varios campos, como por ejemplo la independencia del equipo en el que se encuentra o la portabilidad del código del mismo. Que sea interpretado significa que no se produce un archivo binario ejecutable para el sistema en el que se encuentra, sino que existe un intérprete que se encarga en tiempo de ejecución de transmitir el programa al ordenador de forma adecuada, la diferencia puede apreciarse en la Figura 1. El intérprete en cuestión de Python es lo que se denomina un intérprete interactivo que también se denomina REPL (Read Evaluate Print Loop), y cómo sus siglas en inglés indican funciona como un bucle infinito de lectura de las instrucciones (Harrison, M. , 2017).



**Figura 1. Diferencia básica entre lenguaje compilado e interpretado. Fuente Harrison, M. (2017)**

El motivo por el que este lenguaje es el elegido para el proyecto es principalmente porque es considerado en general por la comunidad como el más adecuado para el análisis de datos, da igual desde qué perspectiva se enfoque, y como consecuencia de esta concepción no son solo sus buenas capacidades base lo que nos interesa, sino que también se han ido desarrollando multitud de librerías y documentación encofradas para hacer este proceso más eficiente. De base, lo que ya se ha planteado de Python como un lenguaje multiparadigma e interpretado lo hacen muy versátil tanto en uso como en los equipos en los que se empleen, además de poder relacionarse fácilmente con otros lenguajes como C. Python permite a su vez establecer una serie de clases que vertebran un proyecto gracias a su orientación a objetos a la vez que tiene scripts funcionales muy generalizados que permite realizar procesamientos de datos a gran velocidad. Esto lleva a la que muchos consideran la gran ventaja de Python, y es su eficiencia en cuanto al tiempo de programación, ya que este es un lenguaje diseñado para ser elegante, ligero y poco indentado, haciendo que, con muy pocas líneas de código, y por tanto horas de programación, puedan hacerse cosas realmente productivas.

Todo esto le ha conferido a Python una gran popularidad y crecimiento en los últimos años, donde cada vez se ha valorado más la eficiencia en cuanto al tiempo de los desarrolladores. Como también se ha dicho, esta popularidad no solo es buena como tal, sino que ha fomentado la creación de numerosos recursos tanto generales como específicos que favorecen mucho el desarrollo en Python.

En resumen, Python es un lenguaje innovador, versátil, muy eficiente para los programadores y con una gran comunidad que ha mejorado aún más si cabe el alcance del lenguaje. Esto, junto a la adaptación concreta de recursos aportados a Python con las técnicas que se explicarán a continuación, hacen al lenguaje la mejor alternativa a elegir para desarrollar este trabajo.

Dentro de Python, hay una gran variedad de entornos de desarrollo, así como gestores varios que permiten y facilitan el desarrollo en el mismo, lo cual es otro gran motivo para elegirlo como opción. Sin entrar en profundidad en toda esta variedad, se ha optado finalmente por dos elementos clave: Conda y PyCharm. Conda es un gestor de paquetes y entornos para Python que permite trabajar con facilidad y llevar un control adecuado de las distintas librerías que representan las herramientas a emplear en el trabajo, las cuales se expondrán a continuación. Por su parte, PyCharm es el IDE (Integrated Development Environment) que se empleará para escribir y estructurar el código, ya que es uno de los más utilizados, con múltiples facilidades e integrado en el propio Conda. En las figuras 2 y 3 pueden apreciarse sus logos y acceder a sus páginas de referencia.



**Figura 2. Gestor de environments Conda. Fuente <https://www.conda.io>**



**Figura 3. IDE de desarrollo en Python PyCharm. Fuente <https://www.jetbrains.com/es-es/pycharm/>**

Estás dos herramientas son de las más empleadas en el desarrollo con Python y en su uso analítico, pues cuentan con una gran sinergia entre ellas. Cubren las necesidades básicas necesarias para el proyecto y aportan una gran variedad de utilidades que facilitan la vida a los programadores, motivo por el cual son tan ampliamente empleadas y apreciadas por la extensa comunidad de Python. Además, y como ya se ha planteado, son perfectas para incorporar las tecnologías que se explicarán a continuación sin suponer ninguna complicación técnica o de gestión.

---

## 2.2 Tecnologías de fijación tradicionales

---

Una vez establecido el lenguaje de programación como herramienta básica para el trabajo es necesario introducir cuáles son las tecnologías como tales y la base teórica detrás de las mismas que servirán de sustancia para este proyecto. A fin de cuentas, los motivos que han llevado a la selección de Python como lenguaje es su capacidad de incorporar estas de la mejor forma posible.

Como ya se adelantó en la introducción del trabajo, estas tecnologías se separarán en un primer grupo basado en conceptos y métodos estadísticos que sirvan de base para las distintas predicciones, y luego en una propuesta más diferenciada mediante el uso de

redes neuronales y evoluciones aplicadas de las mismas que, aun teniendo también implicación estadística, mostrará claramente la diferenciación con el primer grupo. Este primer bloque es el que pasaremos a explicar en este apartado del capítulo.

Para este trabajo, cuentan con una serie de datos obtenidos a partir de la web oficial de Eurostat (Eurostat, 2022), que es la fuente de datos estadísticos más importante del continente al proceder de la institución misma de la Unión Europea. Estos datos son históricos sobre distintos elementos clave del mercado en los últimos años, con información sobre las ventas y los precios con distintas catalogaciones. Estos datos entran dentro de lo que se conoce comúnmente como series temporales: registros de determinados valores, precios en este caso, relacionados con el momento en que fueron medidos. A lo largo de los años se han desarrollado multitud de modelos de análisis diseñados para tratar con estas series temporales, aportar información relevante sobre las mismas y, en última instancia, hacer estimaciones de qué valores tomará dicha serie en instantes futuros.

Como se ha comentado en el apartado anterior, la versatilidad y eficiencia de Python no solo lo hacen adecuado para trabajar con estas series temporales, sino que ha atraído a mucha gente y entidades que han desarrollado ya multitud de herramientas para implementar de forma eficiente estos modelos que se han comentado dentro del lenguaje, lo que proporciona un punto de partida sólido desde el cual aproximarse a este análisis que se plantea para el proyecto.



A continuación, se procederá a presentar y explicar algunas de estas herramientas seleccionadas para el proyecto, generalmente, y en este caso concreto también, representadas en Python a través de unas librerías que emplear. Cabe señalar que la intención de este capítulo del estado del arte es presentar las herramientas, no aplicarlas aún al caso, de modo que aún no se enseñará código empleando dichas librerías, solo se las introducirá y explicará su trasfondo, para posteriormente adjuntar ejemplos de su uso para los datos del trabajo en el capítulo cuatro.

## **Statsmodels**

Hablando de herramientas que tienen su base en la estadística es inevitable hablar de la que probablemente sea la herramienta más empleada a la hora de hacer análisis estadísticos en Python, la librería Statsmodels.

Esta librería ofrece funciones para multitud de tareas de análisis que permiten extraer información de un conjunto de datos y que puede ayudar tanto a la investigación científica como a la toma de decisiones empresarial. Statsmodels forma parte del denominado kit científico de Python, y tiene una alta integración con sus otros elementos, que también son algunas de las librerías más empleadas de todo Python como pueden ser Matplotlib, Pandas o Numpy, y que desde luego serán usadas en el desarrollo del trabajo.

Fue concebida inicialmente como un modelo de la librería científica SciPy fue desarrollada en 2009 por el programador Jonathan Taylor durante unas jornadas de Google, que sirvió como promotor y financiador del proyecto que desembocará en la organización actual (Statsmodels, 2022). Statsmodels está registrada y gestionada por la organización homónima, que se encarga principalmente del mantenimiento, desarrollo de mejoras y gestión de la marca como tal, puede verse el logo representativo de la librería en la Figura 4.



**Figura 4. Logo representativo de Statsmodels. Fuente**  
**<https://www.statsmodels.org/stable/about.html#>**

Se ha comentado que es fundamental para el análisis estadístico, pero aún no se ha especificado por qué en concreto. Sin entrar en el uso que se le dará con datos como tal, Statsmodel permite aplicar una serie de métodos muy importantes a las series temporales de las que se dispongan. Entre estos se destacan:

- El filtro Hodrick-Prescott: permite separar fácilmente los datos de las series temporales en sus dos factores característicos: la tendencia y las variaciones cíclicas. Esto permite analizar las series temporales no para realizar estimaciones futuras directamente pero sí para obtener información que ayude a entender los datos, otorgar una base sobre la que enfocar otros métodos e incluso aportar a la toma de decisiones
- El objeto Unobserved Components: un objeto aportado por la librería que permite almacenar la serie temporal en él para posteriormente usar distintos métodos del mismo que devuelva datos estadísticos de esta serie, como sumarios o descomposiciones de la serie por distintos parámetros. De nuevo resulta una alternativa muy interesante para explorar la serie de forma más profunda y no solo a ojo.
- Métodos de descomposición estacional: otra alternativa en forma de funciones que permite descomponer la serie para extraer componentes cíclicos o estacionales.
- El test de Dickey-Fuller: un importante test estadístico que permite comprobar directamente si una serie temporal tiene o no componentes estacionales importantes. Esto es clave para saber si pueden aplicarse modelos de estimación concretos que ya serán usados para obtener las predicciones
- Otras herramientas estadísticas: la librería cuenta además con una gran cantidad de utilidades varias relacionadas con el pack científico de Python del que ya se ha hablado además de algunas propias. Destacan pruebas de correlación para los errores, test de normalidad o funciones básicas para la representación gráfica.



- ARIMA básico: Statsmodels no destaca por su aplicación de este modelo, que se explicará más en profundidad a continuación, pero ofrece unas funcionalidades básicas para aplicarlo y realizar predicciones estándar sobre las series temporales en cuestión.

En definitiva, Statsmodels no es una librería centrada en la predicción de valores, pero si se la estudia por su análisis estadístico es la más relevante de todas las disponibles en Python así como la más empleada, y básica para cualquier labor relacionada. Además, sí que cuenta con herramientas predictivas base que puedan servir de punto de referencia. En definitiva, se trata de un pilar de la comunidad y una de las librerías más usadas del mundo y más aún del sector, un referente indispensable en un trabajo de estas características.

## Modelos ARIMA en Pmdarima

Como se ha comentado, Statsmodels no tiene su punto fuerte en las estimaciones a futuro como tal. Aun así, se ha incluido como una de las utilidades la capacidad de aplicar un modelo ARIMA a las series de datos, ya que es cierto que tiene funciones para ello, pero cuando lo que se quiere es realizar estimaciones con estos modelos, la librería necesaria es **pmdarima**.

Pmdarima surge como una iniciativa de software libre desarrollada por la ya mencionada amplia comunidad de Python como respuesta a la necesidad de aplicar los modelos ARIMA de predicción de forma rápida y eficiente desde este entorno de programación.

Esta idea surge a raíz de comparar Python con el otro gran lenguaje de programación para el análisis de datos, R. R contaba desde hace bastante tiempo con herramientas de este tipo, concretamente la función `auto.arima`, y surgió la idea de la comunidad de desarrollar un equivalente en Python. Es de software abierto y colaborativo, contando con múltiples participantes, y la idea original, así como el mantenimiento y la mayoría de aportaciones vienen de parte del informático Taylor G Smith (Pmdarima, 2022).

Tanto en la explicación de Statsmodels como en esta de pmdarima se ha mencionado repetidamente el modelo ARIMA, sin explicar lo que es. El modelo ARIMA (Autoregressive integrated moving average) es un modelo estadístico capaz de trabajar con series temporales cuyo objetivo no es la obtención de datos o conocimiento de la

misma, como se explicó en las funcionalidades de statsmodels, sino simple y directamente la realización de predicciones sobre los valores futuros de la serie.

El modelo ARIMA es un estándar a día de hoy en el mundo de la estadística y la técnica más empleada sin duda a la hora de trabajar con series temporales. No es el objetivo de este trabajo explicar su funcionamiento y demostraciones internas, ya que algunas son bastante complejas, pero como es la base de este apartado del trabajo sí que es necesaria una pequeña introducción para entender cómo las diferencias entre algunas de las herramientas y cómo se procederá en capítulos posteriores. Fue propuesto por los estadistas George Box y Gwilym Jenkins y pese a no ser tenido en cuenta en sus primeros años de vida, finalmente es un pilar clave a la hora de trabajar con series temporales en la actualidad

Sin entrar en mucho detalle como ya se ha comentado, el modelo ARIMA trata de realizar estimaciones a futuro de los valores de la serie. En el fondo, el concepto mismo del modelo ARIMA es el de un modelo de regresión lineal donde el factor explicativo son los valores anteriores de la propia variable explicada, es decir se considera que los valores de la serie vendrán explicados por sus valores anteriores. Esta es la parte que le confiere la primera parte de su nombre, el componente autorregresivo AR. Este concepto de autoexplicación basándose en los valores previos puede verse claramente en la Figura 5, donde aparecen las fórmulas que los propios Box y Jenkins plantearon. En la figura 5 también aparece uno de los parámetros característicos del modelo, la  $p$ , que representa como de atrás en la serie hay que ir para explicar cada uno de sus valores, se dice que los modelos son AR( $p$ ), autorregresivos con una ventana de  $p$  periodos.

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \phi_2 \dot{Z}_{t-2} + \dots + \phi_p \dot{Z}_{t-p} + \varepsilon_t$$

**Figura 5. Fórmula del modelo AR. Fuente: Time series analysis: Forecasting and control**

El otro componente clave del modelo es la parte de la media móvil MA, que hace referencia a la media de los errores cometidos por la regresión y a cómo se tienen que tener en cuenta para que el modelo sea lo más preciso posible. Esta parte trata de explicar el valor de la serie en un instante basándose en los errores de estimación cometidos en instantes anteriores, puede verse la fórmula en la Figura 6. De nuevo, en esta figura 6 encontramos otro de los parámetros característicos del modelo, la  $q$  que representa la longitud de la ventana de la media para considerar los errores, que no tiene por qué ser igual a la  $p$ , este modelo también se define con este de forma MA( $q$ ).

$$\dot{Z}_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

**Figura 6. Fórmula del modelo MA. Fuente: Time series analysis: Forecasting and control**

Por último, estas dos partes pueden ponerse en común para formar un modelo ARMA( $p, q$ ), que aúna a la vez ambos enfoques, puede verse en la Figura 7. Finalmente, para la última parte que falta se encuentra la integración y a que hace referencia. Los modelos ARMA tienen la característica que han de aplicarse a series considerables como estacionarias, así que para asegurarse que las series con las que se trabaja cumplen con esto es necesario diferenciarlas en un orden de  $d$ , siendo  $d=0$  si ya era estacionaria de base y algo mayor si la serie era considerable como no estacionarias. Extraer el componente determinista, y en definitiva estacionar la serie a través del parámetro  $d$  nos deja un modelo integrado, último componente del modelo ARIMA( $p, d, q$ ), que puede verse en su totalidad en la Figura 8 con la  $B$  representando el retardo de un instante hacia el anterior. Cabe destacar que hay variaciones de ARIMA como ARIMAX, SARIMA y demás que añaden nuevos parámetros característicos a los tres básicos, sobre todo el SARIMA de estacionalidad, pero ya que todo esto es una introducción básica para entender el modelo general y el uso de las herramientas que Python ofrece, así como las diferencias entre ellas, no se entrará en mucho más detalle.

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \phi_2 \dot{Z}_{t-2} + \dots + \phi_p \dot{Z}_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

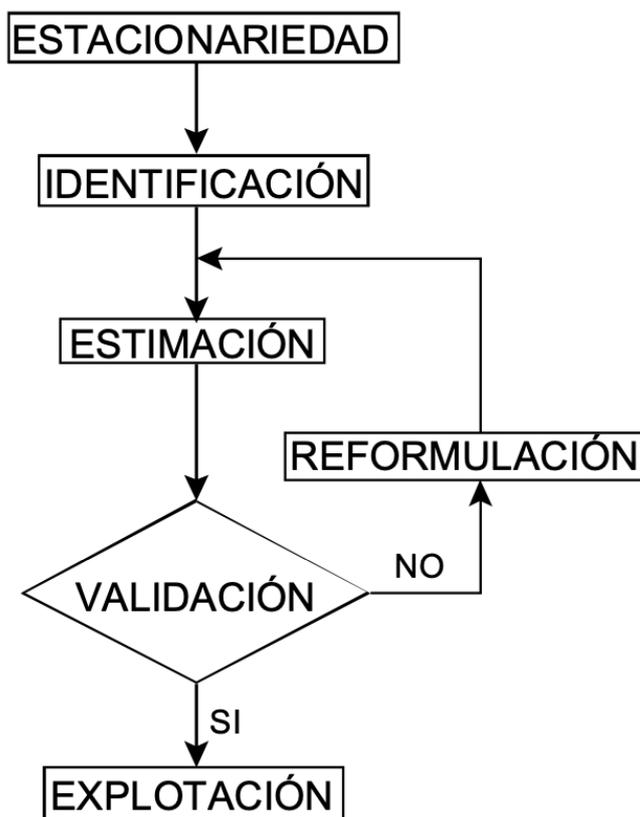
**Figura 7. Fórmula del modelo ARMA. Fuente: Time series analysis: Forecasting and control**

$$(1 - \phi_1 B^1 - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d Z_t = (1 + \theta_1 B^1 + \theta_2 B^2 + \dots + \theta_q B^q) \varepsilon_t$$

**Figura 8. Fórmula del modelo ARIMA. Fuente: Time series analysis: Forecasting and control**

Ahora que ya se ha presentado la base teórica y el funcionamiento básico de un modelo ARIMA como parte fundamental de esta parte del trabajo, puede explicarse por qué están importante la librería pmdarima. Pese a que statsmodels ofrece la posibilidad de aplicar un modelo ARIMA a una serie temporal, los parámetros necesarios para aplicar dichas funciones o métodos tienen que ser propuestos por el programador de forma manual. Las diferencias de precisión y buen funcionamiento de un modelo ARIMA cambian radicalmente entre unos parámetros y otros para los mismos datos, por lo que elegir la p, d y q, además de otros en caso de necesitarlos, es fundamental para el modelo. El proceso clásico para estudiar los distintos modelos ARIMA para una misma serie temporal con el objetivo de encontrar los válidos y el óptimo se denomina metodología Box-Jenkins (Box & Jenkins et al., 1976), cuyo procedimiento se ve reflejado en la Figura 9. Como se ve, esta metodología consiste en una iteración constante buscando los mejores valores para los parámetros en la parte de identificación, así como para los coeficientes de las ARMA en la estimación, llevando tareas como el análisis de errores, los p-valores y otros estadísticos. No se entrará en detalle, ya que se considera que es un proceso costoso, a repetir para cada serie temporal y muy específico, pero es necesario mencionarlo, ya que su automatización es la gran aportación de pmdarima. Como ya se ha comentado, esta librería pretendía imitar el auto.arima de R, y por lo tanto lo que pmdarima permite es que solamente pasando la serie temporal junto a alguna otra especificación, sean las herramientas de esta la que lleven a cabo el método de forma interna buscando los mejores parámetros y coeficientes que minimicen los criterios seleccionados, que como dato suele ser el ampliamente utilizado en el sector AIC (Akaiken Information Criterion) (Akaike, 1974).





**Figura 9. Diagrama de flujo de la metodología Box-Jenkins. Fuente: Apuntes Econometría\***

En definitiva, todo este apartado ha tratado de explicar en qué consiste el modelo ARIMA como base teórica de estas herramientas que se emplearán en la fijación de precios, así como la importancia de la librería pmdarima como herramienta en sí capaz de automatizar la compleja metodología de Box-Jenkins, necesaria para sacarle todo el provecho al método más relevante para tratar con las series temporales.

Ahora solo queda una aproximación más, mucho más breve, antes de pasar al último apartado del capítulo del estado del arte.

### Facebook's Prophet library

La librería Prophet es una aportación realizada por el departamento de data science de Facebook con el fin de emplearla en sus proyectos propios, que más adelante abrieron. Es una librería que aporta automatización de muchas labores exploratorias al igual que pmdarima para que gente sin conocimientos estadísticos avanzados pueda usarla, pero con margen para que la aportación de conocimiento experto pueda sacarle más provecho.

Prophet es un modelo aditivo de regresión con una capacidad mejorada cuando se trata de series temporales con una alta estacionalidad, y emplea una gran cantidad de tecnologías internas para añadir potenciales mejoras a través de conceptos más profundos que no se van a desarrollar, pero sí mencionar como series de Fourier o variables Dummy (Pik & Ghosh, 2021). También cuenta con conceptos que se le escapan a pmdarima como la consideración de calendarios en la estacionalidad, como pueden ser festivos o similares.

Prophet es una alternativa interesante para realizar predicciones y producida de forma interna por una de las mayores empresas del sector como es Facebook. Ha sido añadida al estado del arte, ya que será empleada en capítulos posteriores para los datos del trabajo y se comparará con los modelos ARIMA tradicionales.

---

## 2.3 Tecnologías de fijación vanguardistas

---

Por último, en este capítulo sobre el estado actual de las técnicas y tecnologías a emplear en el trabajo se llega a este apartado denominado como vanguardista. Se ha decidido denominar así para hacer referencia a una tecnología concreta que, pese a haber sido desarrollada hace ya bastante tiempo, ha sido durante los últimos años donde ha empezado a hacerse un hueco en el campo de las predicciones de precios y a plantearse por algunos expertos como alternativa a los modelos anteriores basados en aproximaciones estadísticas puras a las series temporales. Se está hablando de las RNN (Redes Neuronales Recurrentes) en general, y de las denominadas como LSTM (Long-Short Term Memory) en particular.

Todas estas técnicas que entran dentro del amplio abanico de la inteligencia artificial han tenido desde hace décadas una gran atención en lo que a realizar predicciones de precios se refiere, principalmente debido al campo de las finanzas donde las predicciones de stocks han permitido desarrollar en profundidad muchas de estas herramientas. El camino de la investigación de estas tecnologías informáticas y matemáticas ha recibido numerosas aportaciones con el fin de ser incluidas en las predicciones bursátiles, pero ya que el fondo de esto es la estimación a futuro de una serie temporal, también es perfectamente incorporable al asunto de este trabajo, ya que es lo mismo que se pretende.



Las redes neuronales recurrentes tienen en su origen múltiples aproximaciones que abordan los conceptos teóricos que dieron pie a ellas, de las cuales destacan: las conocidas como Redes de Hopfield (Hopfield, 1982), cuando el científico estadounidense John Hopfield las comenzó a desarrollar con el objetivo de aportar en el tema del que trata este trabajo y las predicciones en series temporales y el origen de las RNN como tal se conoce hoy en día en general por parte de Rumelhart (Rumelhart, Hinton, & Williams, 1986). A su vez, otras aproximaciones vanguardistas a la utilización de redes neuronales en general para este tipo de tareas tuvieron lugar a finales del siglo pasado en Japón, donde diversos autores trataron de emplear estas herramientas para realizar predicciones concretas en el mercado real y compararlas con las equiparables al apartado anterior, donde se destaca uno de los más relevantes al tratar predecir el mercado de la bolsa de Tokio (Kimoto & Asakawa, 1990).

Posteriormente, y como evolución en la aplicación de estos avances, en 1997 aparece la herramienta que realmente será empleada en este proyecto, las LSTM (Hochreiter & Schmidhuber, 1997). Estas series, cuyo funcionamiento interno será explicado a continuación, buscaban mejorar el rendimiento de las RRN tradicionales y aumentar la precisión, así como la eficiencia y velocidad de las predicciones.

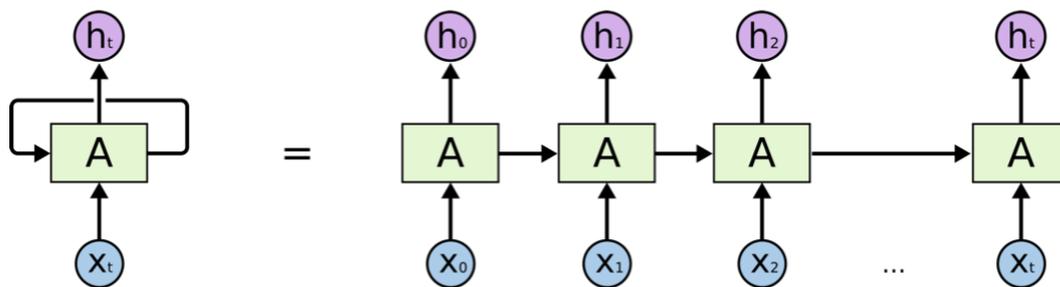
Todas estas herramientas de redes neuronales no llegaron a convencer a gran parte del sector, que las comparaba con las clásicas predicciones ARIMA que eran sólidas y no solían cometer fallos. Sin embargo, y como se ha adelantado al inicio de este apartado, durante los últimos años esto ha cambiado. A partir de 2018 y empezando por un estudio de Makridakis (Makridakis, 2018), numerosos autores afirman y demuestran en una gran cantidad de artículos como las estimaciones realizadas por los modelos basados en redes neuronales en general y por las LSTM en particular han sobrepasado a los clásicos y se han convertido en la mejor alternativa como tecnología de fijación de precios.

Para ejemplificar la clave de este apartado del trabajo, encontramos un artículo cuya intención era una comparación directa entre las técnicas ARIMA explicadas en el apartado anterior, y la tecnología más novedosa LSTM, donde se concluye que: “the empirical studies conducted and reported in this article show that deep learning-based algorithms such as LSTM outperform traditional-based algorithms such as ARIMA model. More specifically, the average reduction in error rates obtained by LSTM is between 84 - 87 percent when compared to ARIMA indicating the superiority of LSTM to ARIMA” (Siami-Namini & Siami Namin, 2018). Es por esto, que son este tipo de redes neuronales y deep learning la herramienta y modelo con el que se ha decidido comparar y complementar las técnicas anteriores y que veremos aplicada al mercado de cítricos en el capítulo cuatro.

Antes de finalizar con esto y al igual que con los modelos ARIMA, pese a que no es el objetivo de este trabajo explicar en profundidad el funcionamiento y optimización de este modelo, se va a realizar una breve introducción al mismo para entender los conceptos básicos que lo diferencian.

## Recurrent Neural Networks

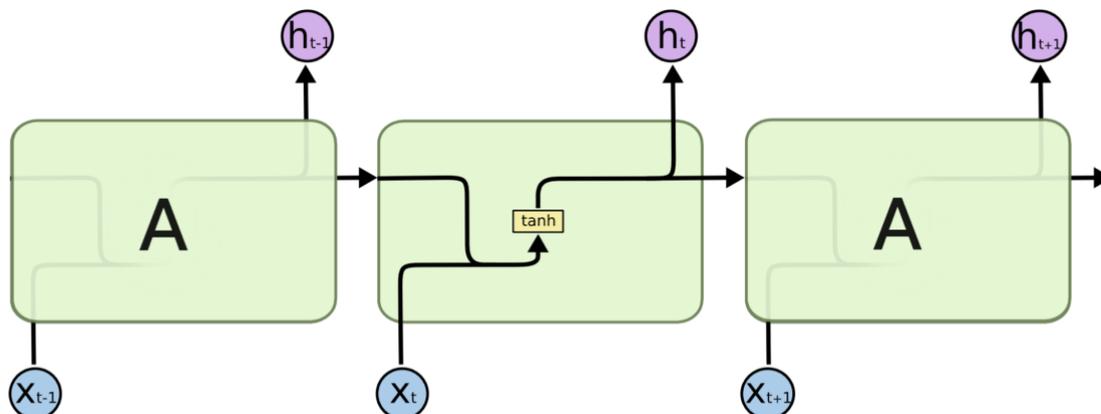
Las RNN son un tipo de red neuronal especializada en secuencias de datos. Mientras que otros tipos de redes neuronales son perfectamente funcionales con datos concretos como imágenes, cuando se trabaja con información en secuencia como un vídeo pierden capacidades y dejan de funcionar adecuadamente. Las RNN vienen a solucionar este problema incorporando un sistema de memoria, ya que es trabajando con este tipo de datos secuenciales como videos o música donde mejor funcionan. Esto también incluye obviamente las series temporales. Además, éstas también destacan no solo por ser capaces de clasificar con los datos ya disponibles en las secuencias, sino que incluyen la posibilidad de continuarlas, lo que en el caso de las series temporales significa hacer predicciones. Entrando en detalle, puede verse la estructura de una RNN en la Figura 10, la cual se explicará a continuación.



**Figura 10. Esquema general y desenvuelto de una RNN. Fuente: Olah 2015**

En la figura del esquema general se puede apreciar por qué las RNN se llaman recurrentes, ya que se ve como a la izquierda hay una retroalimentación que representa el concepto de memoria del que se ha hablado. Está flecha de recurrencia, que se transmite al desenrollado como las flechas entre celdas, es lo que se denomina como **estado oculto**, y permite transmitir información de un estado de la secuencia al siguiente para participar en la predicción de este. Estos estados de la secuencia, la  $t$  en la figura 10, son los componentes correlacionados que conforman la misma, y en el caso de las series temporales de este trabajo representan perfectamente los instantes temporales de las mismas, comúnmente meses. Para ver cómo influye en la predicción de cada estado y

posteriormente entender la diferencia con las LSTM, puede verse el interior de las celdas en la Figura 11.



**Figura 11. Módulo de repetición de una RNN. Fuente: Olah 2015**

Aquí se ve como cada celda coge la entrada para el instante  $t$  y el estado oculto proveniente de  $t-1$ , representando el estado de la memoria recurrente en ese momento, y a través de una función y unos coeficientes calculados durante el entrenamiento producirá un resultado. Este resultado será a partir de ahora el nuevo estado oculto y la salida para ese instante  $t$  también saldrá de este, ya sea directamente como en la figura 11 o pasando antes por una función conocida como Softmax. Este funcionamiento estándar de un modelo RNN permite introducir el concepto de retroalimentación y memoria para poder procesar secuencias, pero tiene un problema. El modelo RNN básico tan solo incorpora al cálculo interno el estado oculto inmediatamente anterior, por tanto, aunque esté tenga dentro información proveniente desde el principio de la secuencia, se puede decir que solo incorpora una memoria a corto plazo, lo cual puede ser realmente problemático para algunos casos. Esto es especialmente relevante para el caso de las series temporales, ya que, como se ha visto en la explicación del ARIMA, la influencia más relevante de una  $t$  anterior puede no venir de la inmediatamente anterior, y es para solucionar este problema de memoria a corto plazo donde entra el LSTM.

## Long Short Term Memory

Un modelo LSTM es un tipo de RNN especializado en solucionar este problema de la memoria a corto plazo, permitiendo tener en cuenta esta influencia de períodos anteriores sin problema. Además de esto, los modelos LSTM no solo son capaces de recordar, sino que también incorporan mecanismos para olvidar información obsoleta, y es esta versatilidad lo que las ha convertido en el tipo de RNN más empleado en la actualidad, especialmente en las series temporales. Puede verse las diferencias generales con el módulo de RNN estándar en la Figura 12.

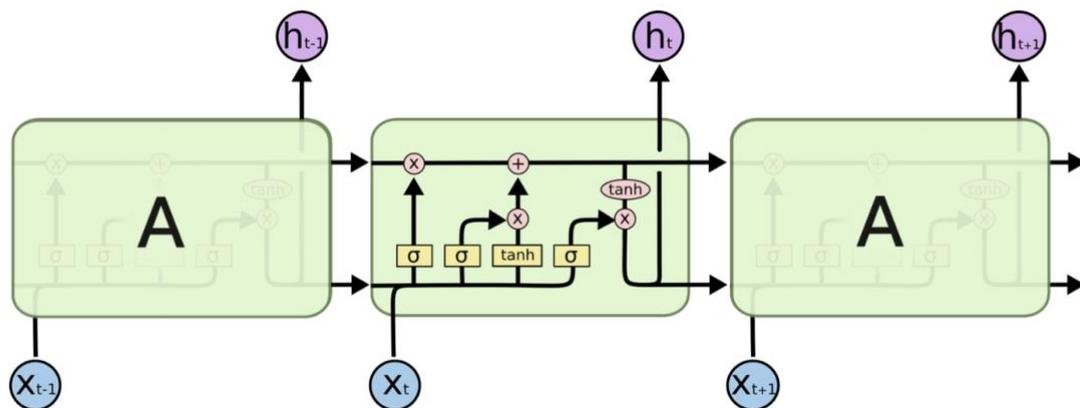


Figura 12. Módulo de repetición de una LSTM. Fuente: Olah 2015

Sin entrar en mucho detalle sobre el comportamiento interno o las funciones de activación, se considera necesario una breve explicación de esta figura y sus diferencias con la RNN. Primero, la línea superior del módulo supone la denominada **celda de estado**, que representa el concepto del que se ha hablado de memoria a largo plazo recorriendo todo el modelo siendo modificado solo cuando se considera necesario. Estas posibles modificaciones de la memoria se realizan en las conocidas como “**forget gate**” y “**update gate**”, que pueden verse en concreto en las Figuras 13 y 14.

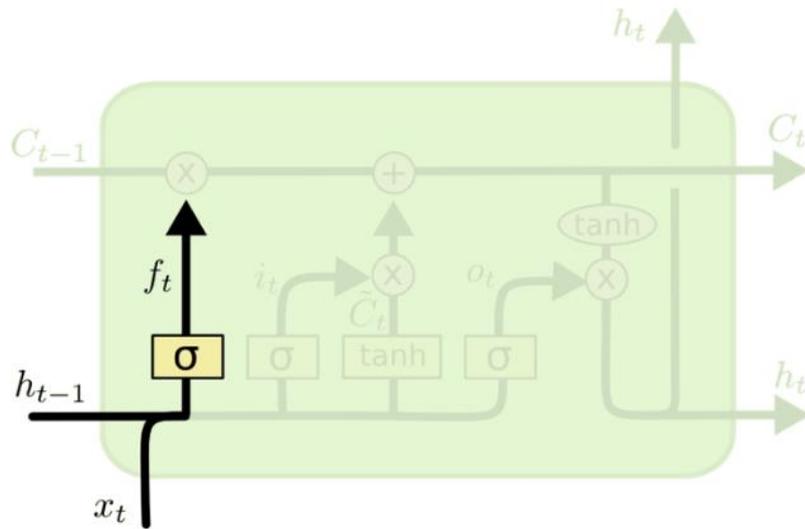


Figura 13. Forget gate del LSTM. Fuente: Olah 2015

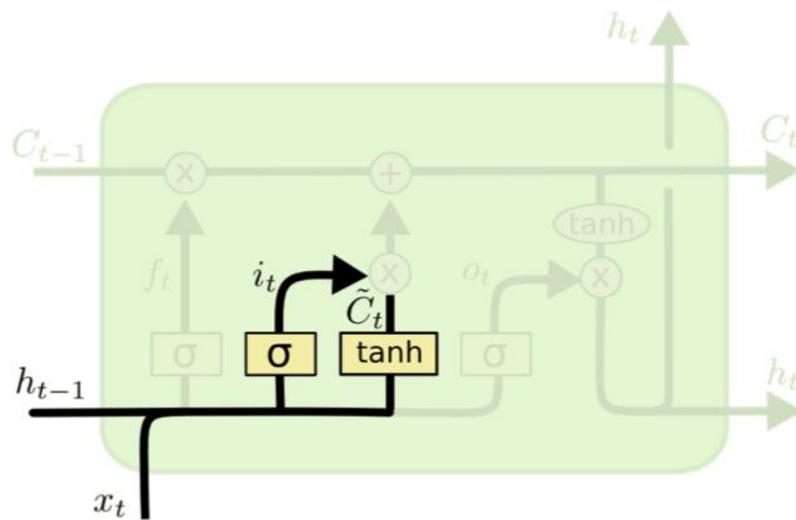
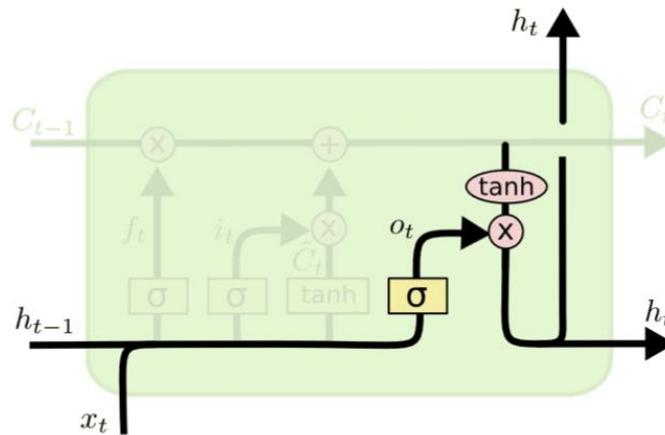


Figura 14. Update gate del LSTM. Fuente: Olah 2015

En la forget gate de la figura 13, la capa se encarga de decidir cuál es la información que quiere olvidarse debido a que un cambio en el estado de la secuencia ha hecho que deje de ser correcta, como un cambio en el tiempo verbal que hace que el anterior tiempo deje de ser adecuado, y con la operación X se garantiza que solo siga lo considerado como correcto. Después en la update gate de la figura 14 se decide qué información es interesante añadir a la vez del valor de esos datos, siguiendo con el ejemplo la primera pata decide que es importante añadir un nuevo tiempo verbal a una secuencia de texto y en la segunda pata indicar ese tiempo verbal.

Finalmente, en la Figura 15 se puede ver la última capa de salida, donde con la información del estado actual junto con la celda de estado ya actualizada, se decide cual será el resultado de esa celda, y se devuelve como resultado a la vez que se envía al siguiente instante para seguir con la secuencia.



**Figura 14. Capa de salida del LSTM. Fuente: Olah 2015**

Los modelos LSTM son una aportación cada vez más empleada y eficiente que ha llamado la atención en los últimos años. Es la herramienta de deep learning e inteligencia artificial en general más adecuada para procesar series temporales y un competidor nato para los modelos ARIMA.

Junto con el apartado anterior, se han expuesto las que son a día de hoy las herramientas más empleadas y efectivas a la hora de hacer predicciones en series temporales, que es de lo que trata este proyecto, así como su funcionamiento básico y diferencias entre ellas. El objetivo de esto es aportar una visión adecuada del estado del arte actual en el campo de la fijación de precios, para así entender cómo serán aplicadas estas herramientas en el capítulo cuatro.



### 3. Análisis del problema

Una vez planteado el estado del arte actual correspondiente a este trabajo, y por ende las metodologías a emplear, en este capítulo se hará una explicación básica del mercado de cítricos y la relevancia de este trabajo, sin entrar en muchos detalles, ya que el objetivo puro de este trabajo no es el apartado económico del mismo sino la aplicación práctica de las técnicas expuestas en el capítulo 2.

Aun así, con este capítulo se pretende dar al menos un pequeño punto de referencia para entender la aplicación que se hará más adelante, explicando la importancia del sector, su peso en la economía, y, sobre todo, cómo encaja con este el concepto de la fijación de precios. De este modo, aunque este capítulo será más breve al estar fuera de la aplicación de técnicas informáticas propia de este trabajo final del grado, es interesante para entender el marco del trabajo en el mundo real.

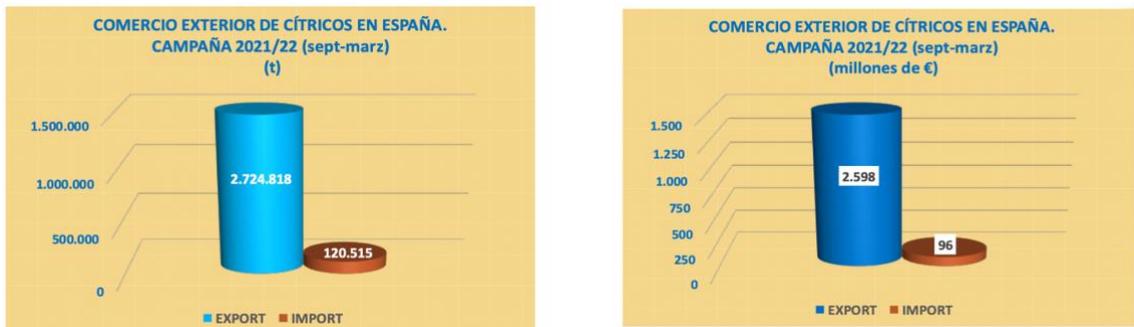
---

#### 3.1 Evaluación del mercado de cítricos

---

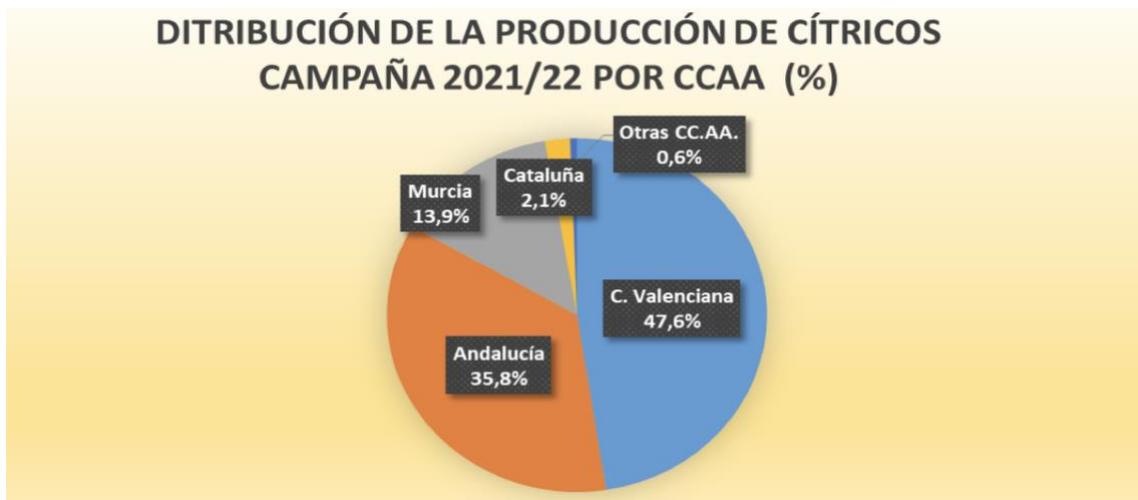
Para evaluar el mercado de cítricos en España y en toda Europa es necesario contar con los datos adecuados para entender cómo es el estado actual del mismo. Para la explicación de este apartado se ha recurrido a dos informes clave como son el informe sobre la mesa de precios de la Lonja de Valencia (Lonja de Valencia, 2022) y el Informe del desarrollo de la campaña 2021/2022 (Ministerio de Agricultura, Pesca y Alimentación [MAPA], 2022). Estos informes se basan a su vez en datos propios de las instituciones o también provenientes del mayor analista del comercio en la Unión Europea, Eurostat (Eurostat, 2022). Como ya se ha dicho, no es la intención de este capítulo ni de este trabajo hacer una descripción completa del sector de los cítricos, por lo que lo expuesto a continuación será lo que se considere más relevante de dichas fuentes, si se desea profundizar más puede accederse directamente a ellas donde se encuentra toda la información detallada.

Primero de todo, es necesario dimensionar el mercado para entender la importancia del mismo. En la actual campaña del 21/22 desde septiembre hasta marzo, España ha exportado un total de 2.724.818 toneladas de cítricos por un valor total de 2.600 millones de euros, junto con unas importaciones de los mismos muy pequeñas dejan un saldo comercial de más de 2.500 millones de euros, esto puede verse reflejado en la Figura 15. Los mayores aportadores de estas exportaciones son primero los denominados como pequeños cítricos, mandarinas por ejemplo, luego naranjas casi con igual peso y algo por detrás limones. Y el destino es en un 86% a países socios de la Unión Europea.



**Figura 15. Exportaciones e importaciones de cítricos. Fuente: MAPA**

Como se puede ver, es un mercado enorme que representa una cantidad enorme de ingresos y uno de los factores más positivos para la balanza comercial española. Así mismo, pese a que este sector engloba a todo el país, la Comunidad Valenciana tiene una relevancia especial, ya que es la principal contribuyente de todo el país, representando prácticamente la mitad con un 47,8%, como puede verse desgranado en la Figura 16. Esto reafirma lo que ya se comentó en la introducción de este trabajo sobre la importancia de este sector para la economía española, y sobre el peso de esta comunidad autónoma en dicho sector.



**Figura 16. Distribución de la producción de cítricos por Comunidad Autónoma. Fuente: MAPA**

Con todo esto, España en general y la Comunidad Valenciana en particular se posicionan como actores relevantes a escala internacional, no solo teniendo unas cifras altas propias, sino en comparación con el mercado global del sector. En dicho sector se dibuja un marco diferenciado para la producción y otro para la exportación como tal. Los principales productores puros del mundo son Brasil y China, pero estos destinan la mayor parte de esta producción al autoconsumo. Mientras, en el campo de las exportaciones destacan a nivel internacional sobre el total México, Sudáfrica y la Unión Europea. A nivel intra-europeo, España es el principal exportador de cítricos de la zona euro, muy

por delante de los siguientes Francia y Grecia, siendo responsable del 61% de las exportaciones extracomunitarias.

Para finalizar este apartado sobre el mercado de cítricos no solo con el peso del mismo en la economía española y la importancia de España como exportadora, es interesante aportar algo de información sobre la estructura interna del mismo. El sector frutícola español es bastante diverso y cuenta con empresas potentes que se dedican de forma más industrial y a gran escala, pero es necesario dejar claro que a día de hoy existe una parte importante de los productores que consisten en cooperativas de agricultores que, aunque se han ido modernizando durante los últimos años, tienen una estructura más tradicional que los otros.

Como resumen, el mercado de cítricos en España es una gran aportación a la economía nacional y su balanza de pagos comerciales. A su vez, España y la Comunidad Valenciana tienen un peso en el mercado internacional como uno de los mayores exportadores mundiales de cítricos, los primeros de la Unión Europea por amplio margen, todo esto con una estructura productiva diversa en cuanto a la organización de los mismos.

---

### 3.2 Problemas relacionados con el objetivo de fijación de precios

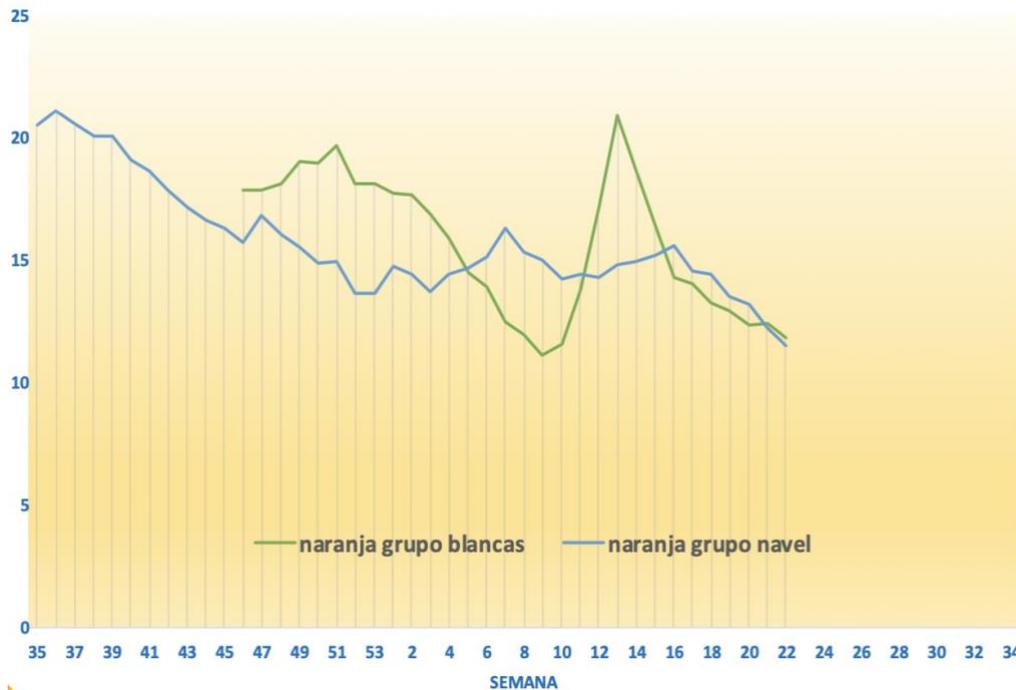
---

El mercado de cítricos en España tiene un gran peso e importancia como se ha visto en el apartado anterior, pero durante los últimos años y en la actualidad está haciendo frente a una serie de desafíos que pueden suponer un problema realmente grave si no se abordan correctamente.

Desde hace unos años el colectivo de agricultores y ganaderos en su totalidad, incluyendo frutícolas y especializados en cítricos, lleva haciendo numerosas reclamaciones en relación con los escasos márgenes que les deja la cadena de distribución. Este problema se centra en que hay una gran diferencia entre los precios a los que les compran a ellos las frutas y los que acaban pagando el consumidor final. Estos problemas y reclamaciones datan de antes de la pandemia donde ya hubo manifestaciones muy mediáticas porque las diferencias eran muy significativas. Esta brecha en el margen llegaba a un 400% más caro para el consumidor que el precio de origen, y un 500% para naranjas y mandarinas (Daniel, 2020). Estos escasos márgenes han originado muchos problemas de venta a pérdidas y una gran amenaza para el sector.

Aparte de los escasos márgenes, ha habido una gran bajada de precios los últimos años que han acuciado esta brecha aún más, bajada que ha continuado en esta campaña como puede verse en los datos del ministerio en la Figura 17. Está bajada no se ha

repercutido a los consumidores, por lo que es simplemente una pérdida para el sector productor, que cada vez ve durante los años posteriores a la pandemia más y más recortado su ya de por sí escaso margen.



**Figura 17. Evolución del precio medio de la naranja en la campaña 21/22. Fuente: MAPA**

Estos problemas vienen originados por diversos motivos, donde destacan dos que ya se han comentado en este capítulo.

- La competencia: pese a que España prácticamente no importa cítricos, que no otros productos agrarios en los que no se va a entrar, la competencia internacional ha ido creciendo durante los últimos años. Pese a que los grandes productores como China aún no están entrando fuertemente en la exportación, España cuenta con grandes competidores históricos como México o Egipto y una creciente competencia proveniente del hemisferio sur por parte de Argentina y Sudáfrica (MAPA, 2022). Estos competidores provienen de países con una mano de obra mucho más barata y con unos controles y requisitos sanitarios y de calidad mucho menores al no contar con la Unión Europea. Estas diferencias son lo que los agricultores nacionales esgrimen para denominarlo competencia desleal y exigir protección frente a ellos.

- Escaso poder de negociación: como también se comentó en el apartado anterior, parte del sector se encuentra muy fragmentada en multitud de cooperativas que negocian de forma separada con los distribuidores el precio de venta. Esto hace que en general carezcan del poder de negociación necesario para hacer valer sus condiciones y se vean obligados a aceptar lo que se les ofrece.

En definitiva, el mercado y sus miembros forman parte de un sector potente en términos de envergadura y demanda, pero con dificultades diversas que suponen un reto a tener en cuenta. Estos impedimentos fruto de la competencia internacional de países más baratos por el efecto de la globalización y de la fragmentación en los productores que les impide negociar con más fuerza en el interior del país han tenido un impacto significativo en los precios, que es de lo que trata este trabajo final de grado. Todo esto hace manifiesta la necesidad de algún punto de apoyo que dote a los participantes del sector de información relevante que les permita tomar decisiones adecuadas.

---

### 3.3 Solución propuesta

---

Para finalizar este capítulo, se expondrá cuál es la propuesta para hacer frente a estos retos que se han expuesto y que afronta el sector desde hace años y que amenaza un mercado tan importante para la economía nacional y regional.

Estos problemas oscilan principalmente en torno a unos precios no solo bajos, sino también muy variables y con cambios muy bruscos en los últimos años, como se puede ver en la anterior figura 17 donde en tan solo un año el cambio en el precio de las naranjas de distintas familias ha sido vertiginoso. En este contexto, los agricultores no saben cómo van a evolucionar los precios ni siquiera en el corto plazo, por lo que les es muy difícil realizar la toma de decisiones en cuanto a conceptos como el volumen de la explotación o si incluso los precios no van a poder cubrir costes y necesitaría cambiar algún aspecto de la actividad.

Como propuesta a este problema, este trabajo pretende en el siguiente capítulo plantear una serie de modelos de fijación de precios basados en las tecnologías expuestas en el estado del arte del capítulo 2. Se aplicarán por separado todas las opciones con la intención de obtener predicciones similares pero obtenidas con distintos métodos, además de luego realizar una labor de control sobre su efectividad. Además, la forma del mismo se pretende que sea sencillo de emplear para acceder a ella con facilidad.

Con esto, se pretende realizar una serie de predicciones de los precios que además puedan ser comparadas entre sí para ofrecer al sector en su conjunto una información útil en el proceso de toma de decisiones. De este modo, reducir la incertidumbre en tiempos tan cambiantes y aportar un soporte sencillo de emplear a un sector tan relevante para la economía.

## 4. Diseño y desarrollo de los Modelos

Una vez ya expuestos tanto el marco teórico y tecnológico como el entorno económico de la industria de los cítricos es en este capítulo donde se pasará a la aplicación práctica de los mismos, mediante el diseño de los modelos como tal y su estructura para obtener las predicciones concretas.

El objetivo de cada una de las herramientas será la obtención de un modelo que se ajuste lo mejor posible al marco histórico previo de la serie temporal, y posteriormente emplearlo para fijar el precio teórico esperado de acuerdo con cada modelo para el siguiente punto de la serie, es decir realizar la predicción del precio a un mes vista.

En este capítulo se pretende explicar que estructura se ha seguido para plantear y poner en conjunto todas las herramientas expuestas en el capítulo dos y los resultados obtenidos al aplicarlas a datos obtenidos del sector del capítulo tres, pero de momento no se harán comparaciones cruzadas ni se entrará en detalles sobre la precisión de estos.

Como se va a exponer los modelos obtenidos es previsible que sí se haga alguna apreciación sobre el aspecto obtenido, alguna comparación entre el valor predicho para el próximo mes o si el modelo parece adecuado en términos generales o no. Pero las mediciones sobre la precisión de estos y la evaluación, es decir el proceso de validación de los modelos como tal, tendrá lugar en el capítulo 5.

---

### 4.1 Consideraciones y tratamiento previo

---

Primero, es necesario especificar de donde surge la serie temporal exacta que se va a emplear a continuación. Los datos con los que se trabajará se han obtenido de una página web de datos estadísticos denominada IndexMundi, que a su vez solo ha hecho que seleccionar y plasmar unos datos extraídos de la base del Banco Mundial (IndexMundi, 2022). Estos datos son un histórico de los precios mensuales de los últimos 20 años del precio de consumo de las naranjas, y supondrá la serie temporal concreta a emplear en los modelos de fijación.

Ahora se expondrá el esquema general de las funciones y métodos que pondrán en práctica las herramientas. El esquema general es sencillo:

- Preparación de los datos: esta parte es la primera que tiene lugar y es general para todos los modelos. Se encarga de cargar los datos en el programa principal en forma de DataFrame de Pandas, que es la mejor forma con la que trabajar con datos en Python. Además, realizará labores básicas de limpieza en lo referente a los índices y nombres y ajustará los datos para establecer un índice del DataFrame de tipo fecha asegurándose que no da problemas de formato, lo cual es imprescindible para evitar problemas en los modelos estadísticos. Después facilitará el DataFrame preparado para el uso a los modelos.
- Apartado estadístico: aquí se incluye la aplicación de los tres modelos tradicionales de los que se habla en el capítulo 2: statsmodel, pmdarima y prophet. Cada uno de estos estará separado en una función propia y no interactuarán entre ellos, pero se agrupan de esta manera por sus características similares y su forma equivalente de presentar la información de salida. Recibirán el DataFrame y devolverán el modelo ajustado y una continuación que representará como consideran que continuará la serie en el futuro.
- Apartado LSTM: este apartado plasmará el modelo basado en redes neuronales recurrentes, pero con alguna diferencia importante con respecto a lo anterior. Devolverá el modelo, el tiempo que ha tardado en realizarlo mediante entrenamiento y la predicción estimada, pero aparte habrá más de un modelo como tal. Para ahondar en las capacidades de esta tecnología y compararla con los estándares que suponen los modelos ARIMA, se harán diversas modificaciones y plantearán múltiples modelos similares, pero diferentes para ver cuál se adapta mejor. Para este apartado se emplearán unas librerías especializadas llamadas Keras y Tensorflow.

---

## 4.2 Modelos tradicionales

---

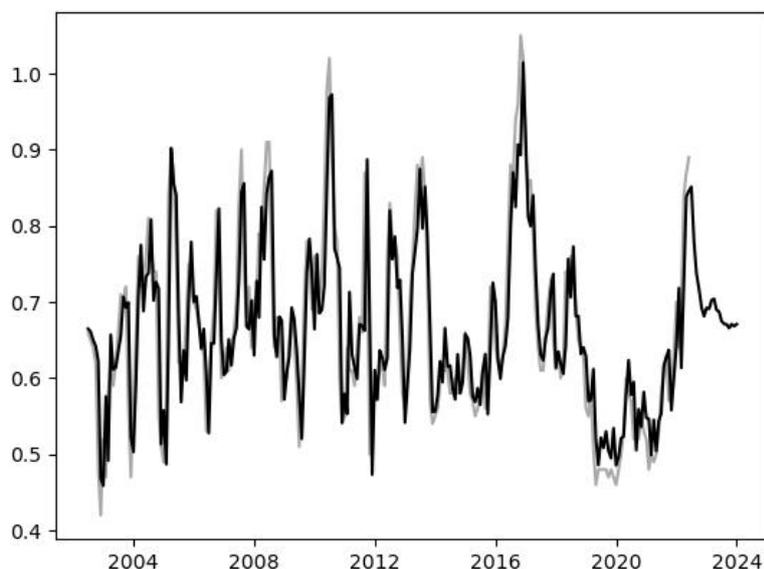
### Statsmodels

El primer modelo es el más directo y el más sencillo de implementar. A través de las funcionalidades descritas en el capítulo dos, se va a proceder directamente a aplicar la función del modelo ARIMA a la serie temporal ya expresada como serie temporal con un DateTime index adecuado al origen de los datos.



El modelo ARIMA realizable por statsmodel requiere que se le especifiquen los valores correspondientes a los parámetros característicos de los que se habló en el capítulo dos: la  $p$ , la  $d$  y la  $q$ . La selección de estos es imprescindible para mejorar lo máximo el modelo, ya que representan literalmente como se llevan a cabo las autorregresiones y el análisis de los datos. No es la intención de llevar a cabo el costoso proceso de identificar los más adecuados, ya que el siguiente apartado soluciona este problema, pero no pueden fijarse al azar si se desea un mínimo de buen funcionamiento. Las elecciones han sido:  $p = 10$ ,  $d = 0$  y  $q = 6$ .

Los valores elevados para la  $p$  y la  $q$  se han elegido para un enfoque más amplio, aunque previsiblemente menos preciso, que en el siguiente apartado se prevé que baje considerablemente de acuerdo con los valores habituales que se obtienen en estos modelos. Por otro lado, el 0 en la  $d$  es más determinante porque corresponde a la parte de Integración para conseguir estacionariedad en los datos, pero con un análisis superficial de los datos no se aprecia una tendencia fuerte así que se ha hecho una consideración inicial de que la serie ya es estacionaria de por sí y por tanto el grado de integración es 0, es decir que no es necesaria. Con estos parámetros ejecutamos el modelo de statsmodel y obtenemos el siguiente resultado, se puede ver el ajuste de modelo a nivel gráfico en la Figura 18.



**Figura 18. Modelo ARIMA (10, 0, 6). Fuente: Elaboración Propia**

Las herramientas ARIMA de statmodel proporcionan de forma automática y con mucha facilidad predicciones a mayor plazo del propuesto de un mes, como puede verse en la figura 19. Sin embargo, éstas empiezan a utilizar las propias estimaciones del modelo como dato así que la precisión decrece con el tiempo, sin embargo, es útil para apreciar cuál es la tendencia general a esperar para el futuro. Pero como se ha dicho, el objetivo es la predicción de valores a un mes vista, y la propuesta realizada por este modelo para la siguiente etapa de la serie temporal es de 0.851289 €/kg. De nuevo, tanto esta predicción como las siguientes que se darán en este capítulo tienen poca de relevancia sin realizar un análisis de la validación del modelo, lo cual se hará en el capítulo 5, sin embargo, se adelantan aquí ya que son la consecuencia y objetivo de la ejecución de estos.

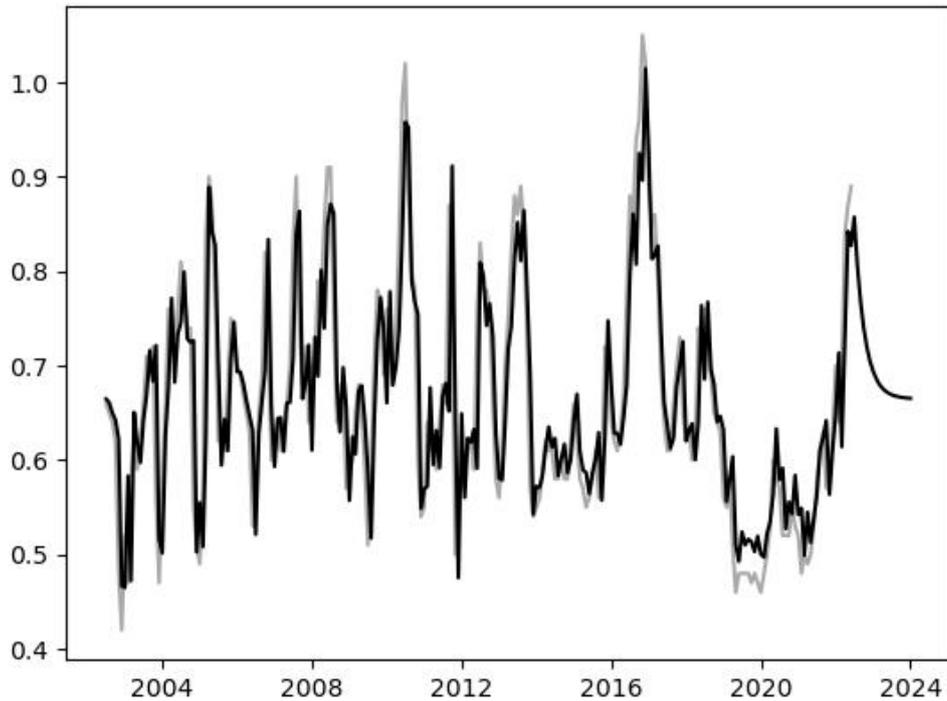
## **Pmdarima**

Como se ha comentado varias veces durante este trabajo, el proceso de optimizar el modelo ARIMA y obtener el más adecuado para la serie temporal con la que se trabajó es complejo y costoso en términos de recursos. Es en la necesidad de mejorar los estos modelos al máximo donde entra la librería pmdarima, que realizará este proceso de forma automática. Cabe destacar que esta herramienta encuentra el mejor modelo posible para la serie dada, pero este modelo no es algo distinto a un modelo ARIMA planteado con statmodel en lo que al nivel teórico se refiere. Esto significa que, si empleas los parámetros obtenidos por autoarima en el modelo anterior, obtendrás los mismos resultados.

El autoarima de pmd es, como su nombre indica, automático en cuanto a los parámetros  $p$ ,  $d$  y  $q$  que se han comentado antes, pero pueden indicarse conceptos como valores iniciales o incluso máximos que influirán en el desarrollo del proceso sin afectar al resultado final. Aun así, sí que se requiere una indicación y es si se quiere que el modelo sea o no estacional, es decir diferenciar entre si deseas un ARIMA o SARIMA, y en el segundo caso indicar el periodo. En este trabajo van a plantearse dos modelos para este punto, uno sin indicación de estacionalidad y otro con una anual, es decir de 12 meses.

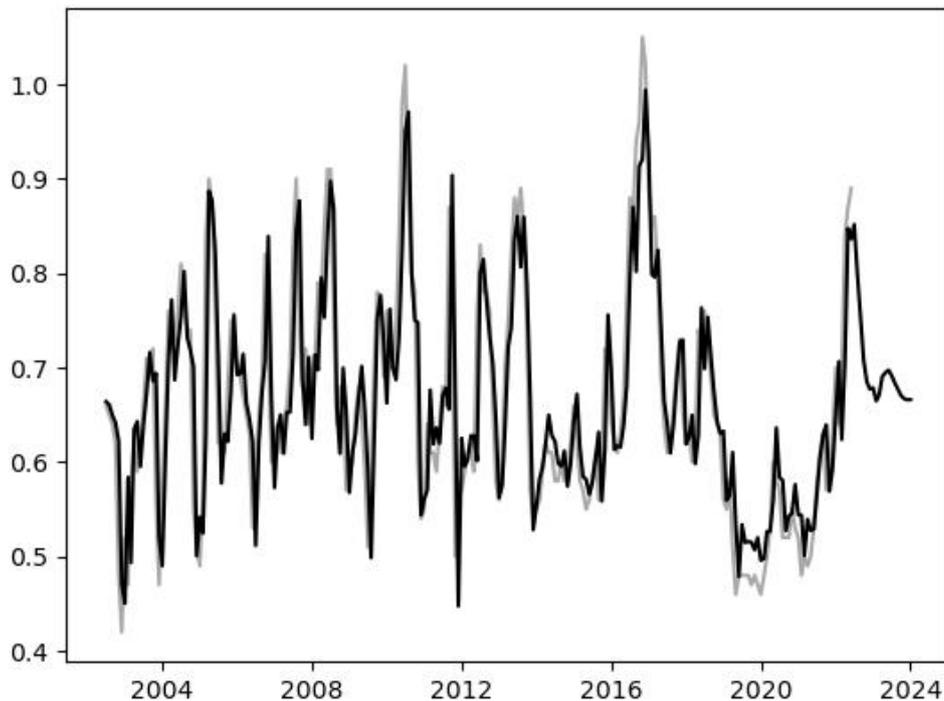
El primero de estos puede apreciarse en la Figura 19. El autoarima ha determinado lo que ya se suponía en las consideraciones del ARIMA base del punto anterior, una  $d$  de 0 debido a que la serie ya es estacionaria y unos valores de  $p$  y  $q$  mucho más bajos que los iniciales. En conjunto, el modelo ARIMA optimizado es  $(1, 0, 1)$  y proporciona una predicción a un mes vista de 0.857325 €/kg.





**Figura 19. Modelo ARIMA (1, 0, 1). Fuente: Elaboración Propia**

Como el objeto de estudio de este trabajo son los cítricos, y en concreto las naranjas para la serie temporal, es posible que al tratarse de un sector de la agricultura tenga una fuerte estacionalidad a nivel anual, dependiendo tanto de las estaciones como las temporadas de las naranjas. Para analizar esta posibilidad se ha aplicado el autoarima bajo esta suposición y se ha obtenido el segundo modelo, apreciable en la Figura 20. Este modelo propone un precio para la siguiente etapa de **0.851586 €/kg**, más similar la primero de los ARIMA.

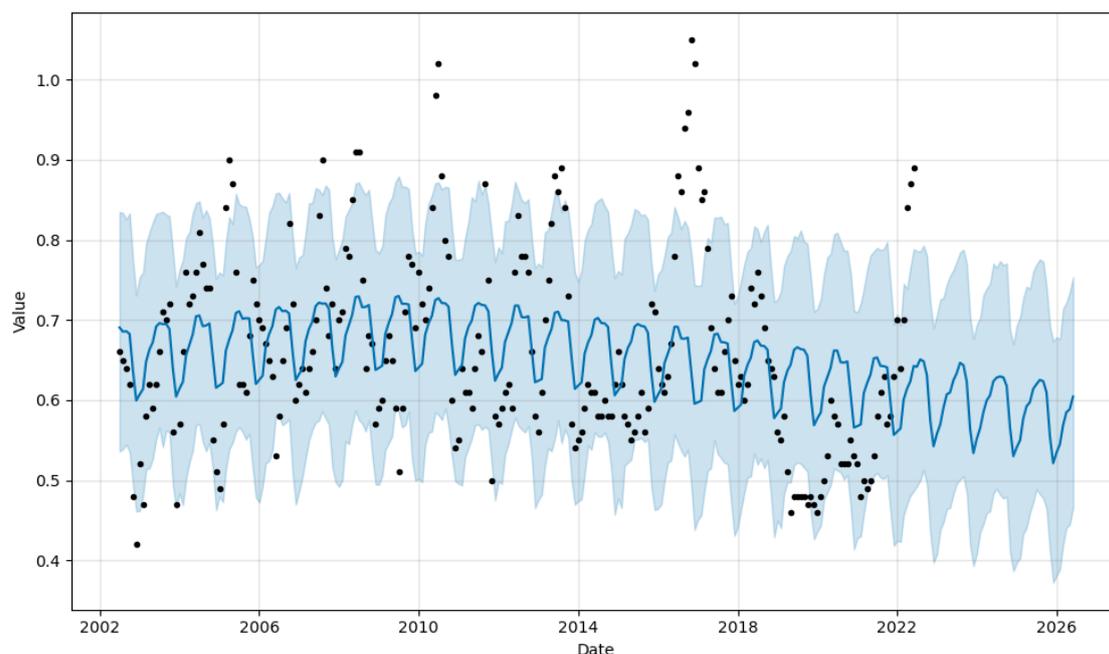


**Figura 19. Modelo SARIMA (2, 0, 0)x(1, 0, 0, 12) Fuente: Elaboración Propia**

Estos 3 modelos ARIMA Y Seasonal ARIMA son fruto del mismo concepto teórico, pero con diferentes consideraciones prácticas. Los resultados se presentan en formatos muy similares y extremadamente fácil compararlos, pero más allá de lo estimable a simple vista en las gráficas el resto se reserva para el siguiente capítulo.

## **Prophet**

Para finalizar con este apartado tenemos la herramienta más diferente de este bloque de 3. Prophet no se basa en ARIMA sino en un sistema propio especializado en componentes estacionales a distintos niveles, pero esto puede ser perjudicial dependiendo de la forma que tenga la serie como se verá a continuación en el modelo propuesto de la Figura 20. El modelo viene dado en otro formato que los anteriores, donde los datos existentes son los puntos negros, la predicción exacta es la línea azul y el área azulada los intervalos de confianza.



**Figura 20. Modelo Prophet Fuente: Elaboración Propia**

Aún sin entrar en la fase de control, puede verse fácilmente como este modelo plantea ya un nivel de fallos importantes, con un ajuste al histórico muy errático e incluso unos intervalos de confianza que suelen fallar, además de que casi ningún dato histórico como tal se encuentra sobre la línea del modelo. Además, la predicción a un mes vista del modelo es de 0.650733 €/kg, un valor extremadamente alejado no solo de las anteriores predicciones sino también de los valores recientes de la serie.

Esta visible imprecisión e incluso fallo en el modelo se debe a que Prophet es una herramienta poco adecuada para esta serie temporal. La especialización de Prophet reside en los componentes estacionales, especialmente cuando hay muchos de ellos, e incluso en factores de calendario como festivos. Esto hace referencia a por ejemplo una serie que presente a su vez: comportamientos marcados por meses del año, semanas concretas dentro de los meses, días específicos de la semana o incluso horas del día, varios de estos a la vez. Cuando esto tiene lugar es cuando Prophet se vuelve una herramienta realmente efectiva, pero este no es el caso. La serie con la que se trabaja no presenta estos componentes, como además se evidenciará más aún en el capítulo cinco, además que al tratarse nuestro dataset original de datos mensuales muchos de estos no son siquiera una opción.

En definitiva, solamente con este análisis lógico general y del modelo y predicción puede verse que Prophet, sin dejar de ser una potente herramienta en su campo, no es adecuada para este trabajo. Se ha decidido dejarla como ejemplo de que herramientas muy potentes para determinadas series pueden no aportar nada de valor en otras con un formato muy distinto y con las que la herramienta no se beneficie de trabajar.

### 4.3 Modelos vanguardistas

---

Ahora en este apartado va a plantear que estructura se seguirá con el modelo basado en LSTM y las distintas variaciones de este que se expondrán para compararlas entre sí.

Antes de explicar las diferencias en los modelos y los resultados de estos, se expondrán los pasos en la programación de la red generales para todos ellos. Como descripción general de los modelos a presentar y el del procedimiento de programación necesario para componerlos, entrenarlos y ejecutarlos tenemos los siguientes.

Para empezar, al tratarse de una red neuronal es recomendable estandarizar los datos para un mejor y más rápido entrenamiento, por tanto, se ha empleado la herramienta de sklearn DataScaler para normalizar la totalidad de los datos y poder trabajar con ellos antes de dirigirlos a los distintos modelos. Cabe destacar que los datos se devolverán a su forma original al presentar resultados, por lo que todas las conclusiones y predicciones presentadas serán transparentes en cuanto a este comportamiento, pero es necesario clarificarlo.

Después, los datos se presentarán en un esquema X vs. Y, donde X representará los datos de determinadas etapas anteriores e Y el valor predicho por el modelo para la etapa inmediatamente posterior, es decir en este caso el mes vista del que se ha estado hablando, todo esto con valores estarán aún normalizados en vez de los precios originales, ya que han de ser transmitidos a la red para el entrenamiento. Esta estructura teórica se reflejará a nivel de implementación en dos arrays que se compondrán dentro del programa, siendo el array X\_train una matriz donde cada posición será a su vez un array con los valores de los días anteriores y el array Y\_train en el que cada posición tendrá solo un valor correspondiente. Por ejemplo y para mostrarlo mejor, si se desea usar diez etapas para predecir el valor de la undécima etapa, el valor en la posición 0 del array X será un array con los valores de las diez primeras etapas y la posición 0 del array Y el valor único de la undécima etapa, para este ejemplo no podrían hacerse predicciones para los 10 primeros días y la longitud de los arrays sería solo de 230 valores. Como ejemplo, en la Figura 21 puede verse esta estructura, que supone los datos que se le pasarán al modelo para el entrenamiento.



```
[ [0.15873016]
  [0.07936508]
  [0.25396825]
  [0.31746032]
  [0.26984127]
  [0.31746032]
  [0.38095238]
  [0.46031746]
  [0.44444444]
  [0.47619048]] >> [0.22222222]
```

**Figura 21. Array X\_train >> Array Y\_train del modelo Fuente: Elaboración Propia**

Una vez se cuenta con estos dos arrays que compondrán los datos para el entrenamiento, y se extrae una muestra para el testeo que se explicará en el capítulo cinco, puede componerse la red como tal. Para ello, se carga desde keras el modelo base vacío al que se llamará regresor. A este modelo es necesario añadirle capa a capa todas las que se considere necesario de forma manual, indicando el número de neuronas por capa junto a otros parámetros como las funciones de las neuronas o el número de datos a esperar. Además, ya que las predicciones son a un mes vista para las etapas, los modelos contarán siempre con una capa de salida con una única neurona para dicho valor de predicción, que también se ha añadido en el código como tal.

Después de añadir por código todas las capas, se compila la red indicando factores clave del modelo, principalmente el optimizador a emplear por el modelo y la función clave del “loss”, parte fundamental que se explicará en el capítulo cinco. Se establece una unidad de tiempo de inicio para medir la duración de entrenamiento, y se procede a entrenar la red como tal pasando como parámetros los arrays explicados y el número de etapas, 100 para todos los casos, al método fit del modelo. Cuando se ejecute el código y este método finalice, el modelo ya estará entrenado y podrá tomarse otra medida temporal para saber cuánto ha tardado el proceso.

Por último, una vez se tiene ya modelo entrenado se especifican los valores de las últimas etapas designadas como base de la predicción y se le pasan como array simulando el formato del X\_train. El método predict del modelo ya entrenado en ejecución devolverá el resultado para la etapa futura desconocida, un mes vista en este caso, y una vez se revierta el proceso de normalización con el DataScaler invirtiéndolo, se tendrá ya la predicción.

Esto es una explicación de lo que se ha hecho para programar la base de todos los modelos que se diferenciarán a continuación. El uso de los modelos como tal consistirá solamente en diferenciar cuál de las variaciones quiere probarse y especificar el Excel con los datos.

Las principales variaciones entre estos modelos serán en función de tres hiperparámetros característicos, que consistirán en:

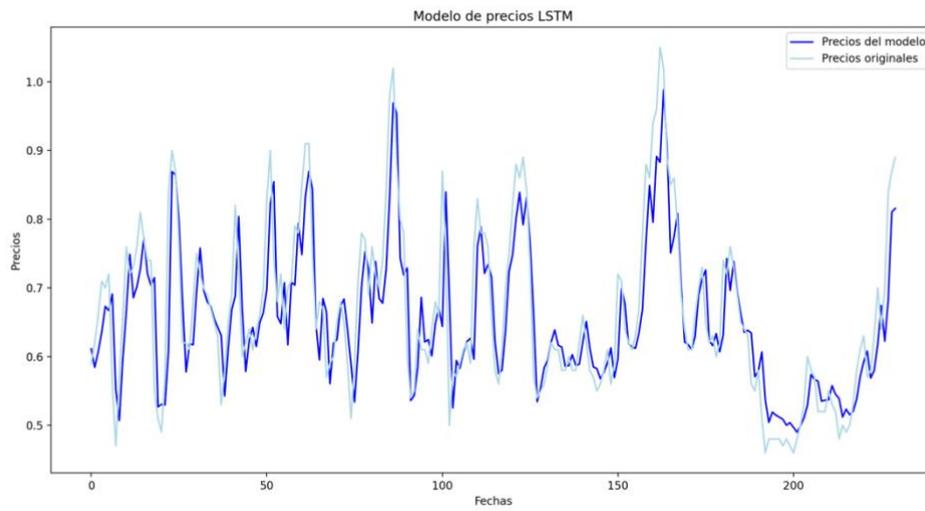
- El optimizador: keras y tensorflow ofrecen distintos optimizadores que engloban parámetros muy importantes para una red neuronal como pueden ser el learning rate o el decay. Se emplearán dos en concreto: Adam y Rmsprop.
- El número de capas ocultas internas del modelo: representará el número de niveles internos del modelo y sus valores serán de 2 y 3.
- El número de neuronas en estas capas: representa los datos procesados y transmitidos por dichas capas y sus valores serán de 10 y 15.

La clave será en encontrar un equilibrio, ya que aumentar cifras normalmente representa un incremento del coste de entrenamiento, pero no necesariamente de los resultados, pudiendo ser en teoría más costosa, pero obtener un peor ajuste final. Además, el optimizador empleado también modifica parámetros clave que definen la red como tal.

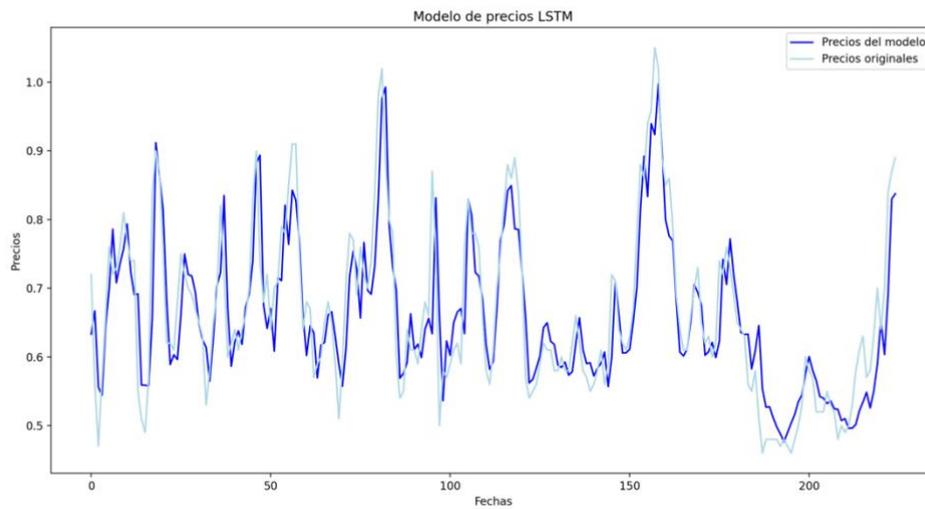
Para la realización de los modelos se separará una función para cada una de las opciones, ya que la modificación de los hiperparámetros supone un cambio en sí mismo del código a emplear y afecta a todo el modelo desde la base. Estos cambios afectan a la parte explicada antes como general, tanto al añadido una a una de las capas con sus características como a la compilación del modelo una vez añadidas todas.

Se han aplicado todas las versiones del programa LSTM a la serie temporal y se han conseguido una serie de resultados que se mostrarán a continuación. En las figuras 22 y 23 puede verse la totalidad del modelo comparado con la serie temporal original en dos casos de aplicación distintos. Puede apreciarse que, aunque el ajuste general es bueno en ambos, existen diferencias en algunas de las figuras, dando lugar a distintas predicciones y resultados.





**Figura 22. Modelo LSTM RMS\_2\_10 Fuente: Elaboración Propia**

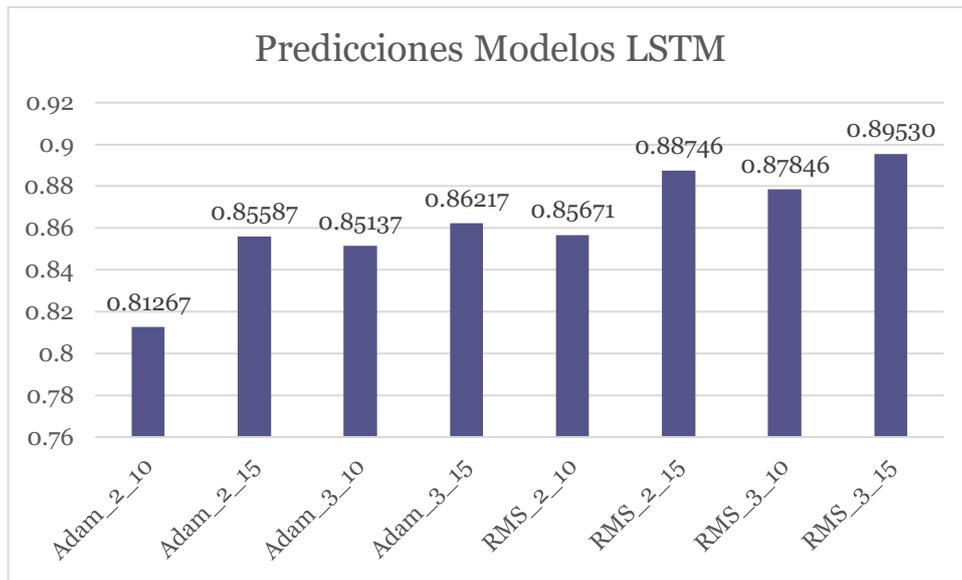


**Figura 23. Modelo LSTM ADAM\_3\_15 Fuente: Elaboración Propia**

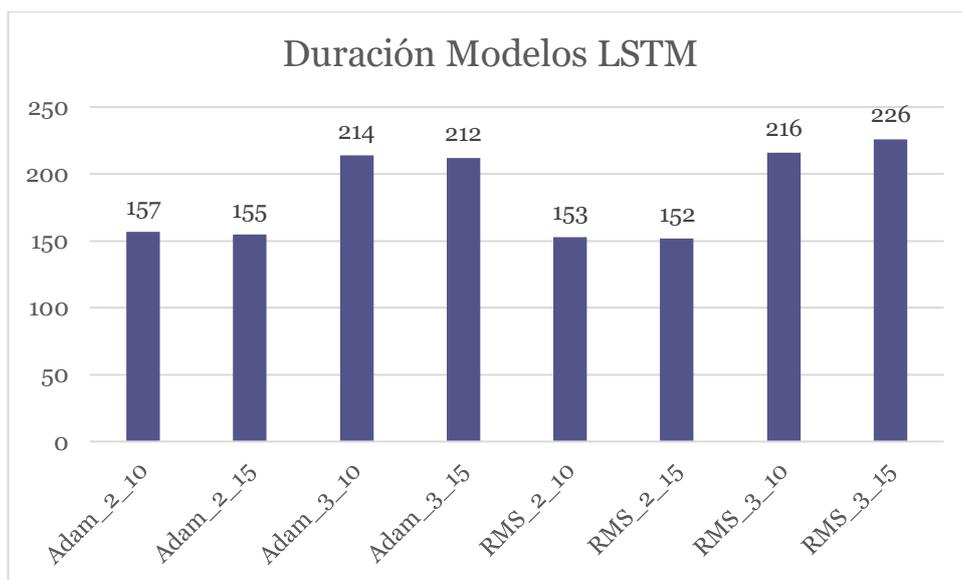
Para mostrar todos los modelos obtenidos, pero sin necesidad de mostrar todas las gráficas, se ha recopilado la tabla de la figura 24, donde se recogen tanto las predicciones de cada uno de los modelos como el tiempo que han tardado en aplicarse. Esta información puede verse también en las gráficas de las figuras 25 y 26.

	Adam				Rmsprop			
	2		3		2		3	
	10	15	10	15	10	15	10	15
Predicción	0,81267	0,85587	0,85137	0,86217	0,85671	0,88746	0,87846	0,89530
Tiempo de entrenamiento	157	155	214	212	153	152	216	226

**Figura 24. Tablas LSTM Fuente: Elaboración Propia**



**Figura 25. Predicciones a un mes vista LSTM Fuente: Elaboración Propia**



**Figura 26. Duración del entrenamiento LSTM Fuente: Elaboración Propia**



De estos datos iniciales pueden extraerse diferentes interpretaciones. A falta del análisis de precisión que se realizará en el siguiente capítulo, pueden apreciarse comportamientos relevantes dependiendo del modelo. Primero, en términos de las predicciones como tal, el optimizador RMS tiende ser más optimista que el Adam y el incremento tanto del número de capas como de neuronas por capa tiene el mismo efecto. En lo que al coste de entrenamiento se refiere, puede verse como el aumento del número de capas supone un incremento significativo, mientras que tanto el número de neuronas como el optimizador no parecen influir.

Para finalizar el capítulo, repetir que pese a tener su utilidad, comparar estos datos de las predicciones no es realmente significativo hasta que se averigua, basándose en otros criterios, cuáles son las más precisas, permitiendo comparar la fiabilidad de los modelos como tal. Esta labor final del trabajo de evaluar como de fiables son los modelos planteados tendrá lugar en el siguiente y último capítulo de este trabajo.

## 5. Pruebas y análisis de resultados

Como ya se ha adelantado, el objetivo de este capítulo final es exponer cuáles son los distintos criterios y los resultados para medir la fiabilidad y precisión de los modelos planteados y realizados en la práctica a lo largo de esta memoria.

Cabe mencionar que como ya se expuso en el capítulo anterior, ante la obvia falta de precisión en el modelo proveniente de Prophet al ser una herramienta especializada no adecuada para estos datos, se le considera de base impreciso por lo que no aparecerá más en este capítulo.

Después de presentar los distintos resultados, el apartado final del capítulo consistirá en un análisis de estos y en la elección de los mejores modelos con base en los criterios adecuados. Esto forma parte del análisis como tal de los modelos, por lo que las conclusiones como tal del trabajo quedan reservadas para la susodicha parte después de este capítulo.

---

### 5.1 Análisis y testeo de los modelos obtenidos

---

A continuación, se separará el análisis en dos bloques de acuerdo a la base teórica que les da sustento, primero los modelos ARIMA y después los LSTM. Para analizar y comparar los modelos, es interesante realizarlo desde dos perspectivas:

- **Parámetros propios:** se refiere a valores propios de la realización o entrenamiento del modelo como tal, y representa el buen ajuste del mismo en base a los datos empleados para él.
- **Predicción del testeo:** esto significa reservar una parte de los datos antes del entrenamiento del modelo como tal y posteriormente solicitar al modelo predicciones para esas etapas, para así comparar los resultados del modelo con los valores reales. Esto es equivalente a realizar el modelo en un periodo anterior y comprobar si sus predicciones han acertado, y es la mejor forma de estimar la precisión y fiabilidad del modelo.

Algunos de los parámetros propios pueden ser exclusivos del tipo de modelo, como se verá a continuación con los ARIMA, pero la precisión del testeo es un método general que permitirá comparar todas las opciones.



## ARIMA y SARIMA

A la hora de comparar modelos ARIMA entre sí hay diversos criterios en cuanto al ajuste del modelo se refiere, estos ya se adelantaron en el capítulo dos del estado del arte y son fundamentales a la hora de compararlos sobre todo entre sí. Estos son los conocidos como “Information Criterion” (IC) o criterios de información y consisten en los de: Akaike (AIC), el Bayesiano (BIC) y el de Hannan-Quinn (HQIC). Estos valores son estimados a partir de los componentes de los modelos ARIMA, como pueden ser las diferencias entre estimaciones y datos reales o el número de datos, y son lo que todos los manuales de estadística recomiendan emplear a la hora de elegir el mejor de los modelos ARIMA. El criterio de selección en relación con este es minimizar los tres criterios lo máximo posible, es decir que se considerará mejor al modelo con los más bajos.

Las herramientas empleadas, tanto statsmodel como pmdarima, ofrecen estos importantes criterios como parte de la información accesible del propio modelo a través del sumario. Esta información, entre otra del propio sumario, es clave para entender el ajuste de estos modelos. En la Figura 27 puede apreciarse la salida de los criterios en PyCharm para el modelo ARIMA inicial de statsmodel, y en la Figura 28 una tabla con estos criterios para los tres modelos tipo ARIMA del trabajo.

```

No. Observations:      240
Log Likelihood         334.989
AIC                   -633.978
BIC                   -571.326
HQIC                  -608.734
    
```

**Figura 27. Criterios de información ARIMA (10,0,6) Fuente: Elaboración Propia**

	statsmodel	pmd SARIMA	pmd ARIMA
AIC	-633,978	-647,526	-647,767
BIC	-571,326	-630,123	-633,844
HQIC	-608,734	-640,514	-642,157

**Figura 28. Criterios de información totales Fuente: Elaboración Propia**

De acuerdo con esta información extraída de los modelos y con los criterios establecidos, podemos comparar los criterios de información para ver que el que tiene los mejores valores, los menores en este caso, es el modelo ARIMA (1,0,1) propuesto por pmdarima sin componente estacional. Cabe destacar que se busca siempre los valores más bajos posibles, no los valores absolutos, de modo que en este caso con todos los valores negativos el mejor modelo es el más negativo.

Otro criterio interno útil para comparar los modelos ARIMA entre ellos de forma interna son los coeficientes del modelo generados, y sobre todo su fiabilidad a través del P-valor de los mismos. El P-valor de un coeficiente en un modelo de regresión representa la fiabilidad de dicho coeficiente en la representatividad del modelo, es siempre positivo y se busca que sea lo más cercano a 0 posible y por debajo del 0,05. En las figuras 29 y 30 pueden verse estos coeficientes, columna coef, y P-valores, columna P>|z| para los modelos ARIMA, el de statsmodel básico y el de pmdarima. Fácilmente se puede apreciar como el modelo arbitrario construido al principio tiene muchos valores y todos ellos bastante elevados, en cambio la propuesta optimizada de pmdarima ha conseguido unos valores mucho más correctos.

```

Covariance Type:          opg
=====
              coef  std err          z      P>|z|      [0.025   0.975]
-----
const          0.6648    0.025    26.161    0.000     0.615    0.715
ar.L1          0.8143    3.836     0.212    0.832    -6.704    8.332
ar.L2         -0.0863    1.376    -0.063    0.950    -2.782    2.610
ar.L3          0.1806    0.500     0.361    0.718    -0.800    1.161
ar.L4         -0.6758    0.686    -0.985    0.325    -2.021    0.669
ar.L5          0.7686    2.149     0.358    0.721    -3.444    4.982
  
```

**Figura 29. Tabla de coeficientes ARIMA (10,0,6) Fuente: Elaboración Propia**

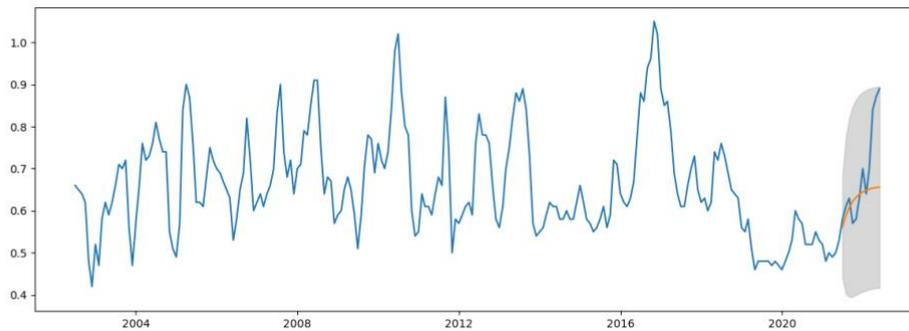
```

Covariance Type:          opg
=====
              coef  std err          z      P>|z|      [0.025   0.975]
-----
const          0.6647    0.022    29.649    0.000     0.621    0.709
ar.L1          0.7315    0.049    15.031    0.000     0.636    0.827
ma.L1          0.4427    0.053     8.416    0.000     0.340    0.546
sigma2         0.0038    0.000    13.383    0.000     0.003    0.004
  
```

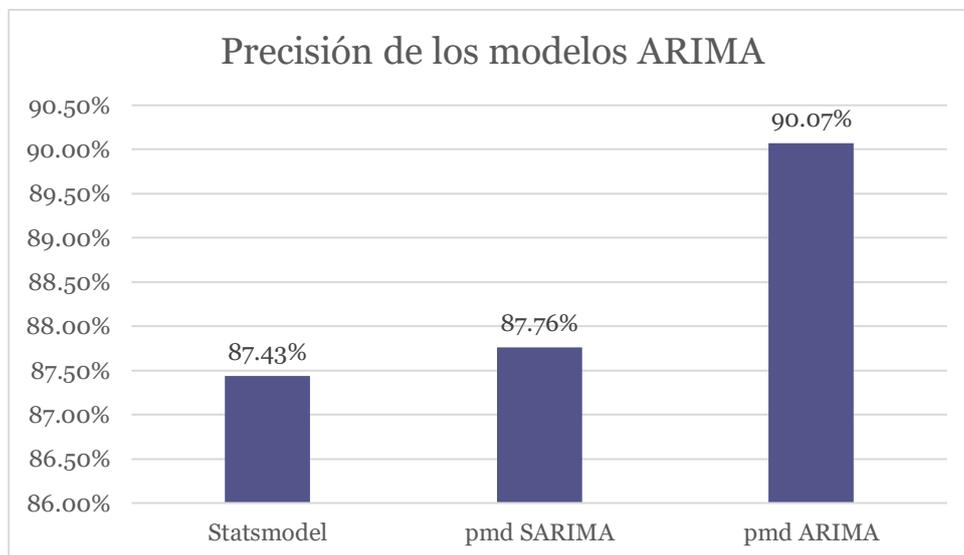
**Figura 30. Tabla de coeficientes ARIMA (1,0,1) Fuente: Elaboración Propia**



Ahora, una vez finalizada la parte más interna, pueden presentarse una comparativa de los modelos a nivel de precisión que a su vez pueda medirse de forma adecuada con los propios de los siguientes modelos LSTM. Para esto, se ha realizado un testeo con los valores de los últimos 12 meses, que se han excluido de los modelos y realizado predicciones de acuerdo con estos. De este modo pueden simularse unas predicciones equivalentes en cuanto a información conocida a las que se están haciendo de cara a futuro, y, comparándola con los valores reales que tomó la serie en esos periodos, medir la precisión de cada modelo. En la figura 31 puede apreciarse la comparativa gráfica para el modelo ARIMA de pmdarima no estacional y en la Figura 32 verse la comparativa entre las precisiones de los tres modelos.



**Figura 31. Testeo del modelo ARIMA (1,0,1) Fuente: Elaboración Propia**

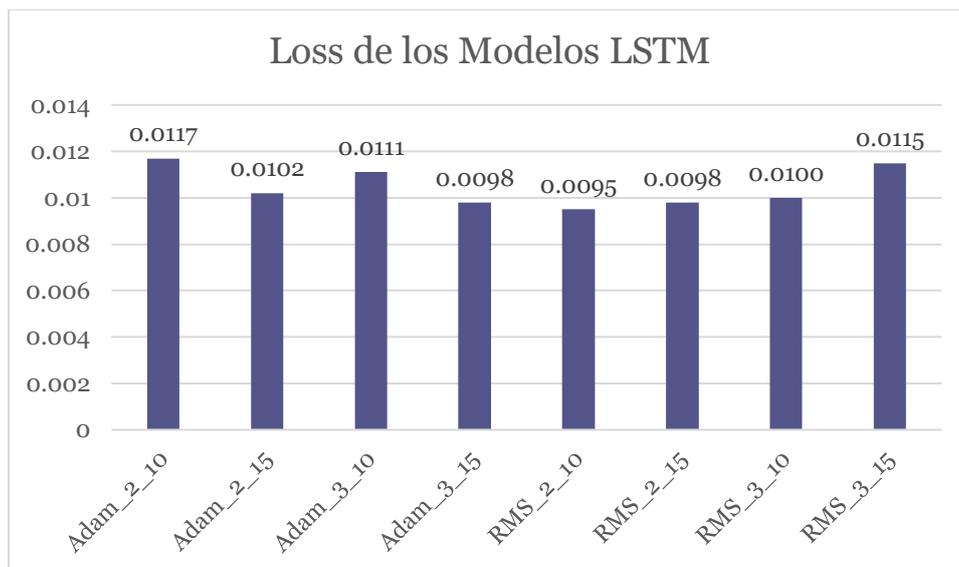


**Figura 32. Precisión de los modelos ARIMA Fuente: Elaboración Propia**

## LSTM

Ahora, ya en el apartado destinado a los modelos basados en redes neuronales, se repetirá el mismo esquema que en el anterior: primero atendiendo a un criterio interno que la propia red emplea y ofrezca como análisis propio y después a una comparativa de precisión en un test que pueda también compararse con los modelos del apartado anterior.

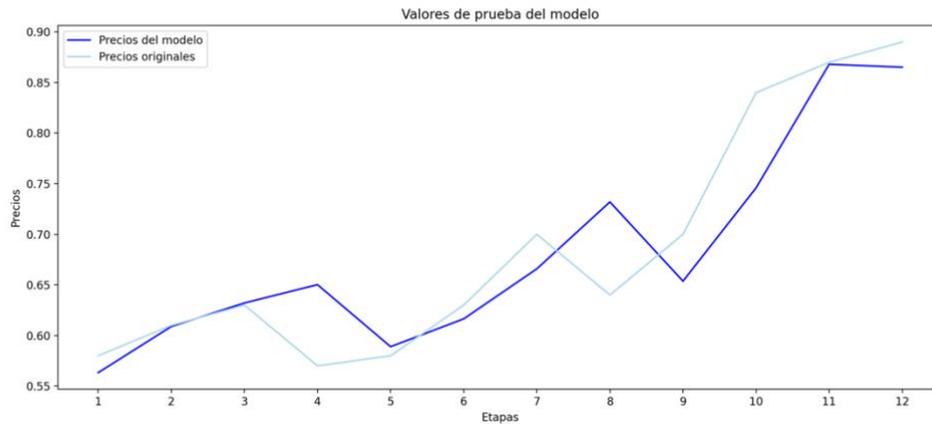
Para empezar, el criterio interno para el ámbito del control del modelo más relevante y aquel del que se va a hablar es el denominado como "loss" del modelo. Este representa la función o medida que internamente la red neuronal está intentando ajustar durante la fase de entrenamiento para mejorar el desempeño de esta. Generalmente representa un parámetro de error que la red define como aquello que quiere minimizarse para considerar un modelo mejor que otro y es, por tanto, lo que se podría llamar como el objetivo durante la primera fase de construcción de los modelos. La propia librería de keras y tensorflow permite ir mostrando como varía este elemento según avanzan los periodos, o épocas, de entrenamiento hasta que concluye. En la figura 33 puede apreciarse cuál es el loss final para cada uno de los modelos.



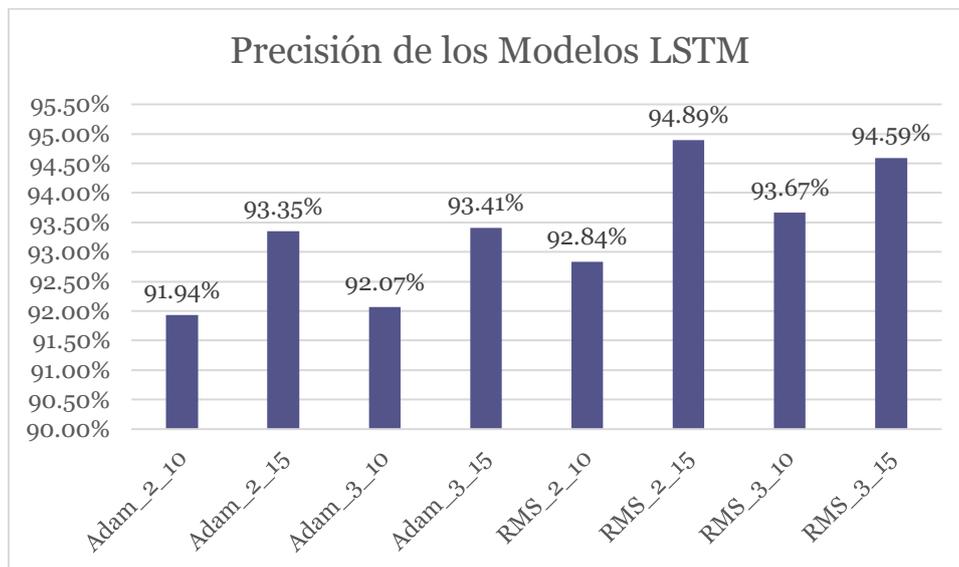
**Figura 33. Loss de los modelos ARIMA Fuente: Elaboración Propia**

La función de loss puede estipularse para que sea una concreta de entre un gran grupo dependiendo de cuales sean las intenciones de la red. En este caso, y ya que se trata de un modelo de ajuste para la predicción, se ha indicado como función el "Mean Squared Error", que hace referencia al error absoluto entre las estimaciones del modelo para cada instante y su valor real. Cabe destacar que, aunque se busca minimizar del todo, no puede considerarse como un criterio absoluto sobre cuál de los modelos es mejor.

El mejor indicativo para comparar estos modelos, y más aún en el caso de las redes neuronales LSTM es la precisión de las predicciones. Repitiendo el mismo sistema expuesto en el apartado anterior, se separaron los valores de las últimas 12 etapas para posteriormente comprarlos con las predicciones realizadas para las mismas y poder estimar así la fiabilidad de los modelos. En la figura 34 puede apreciarse la comparativa en dichas etapas entre los valores reales y las estimaciones del modelo RMS\_3\_15 y en la figura 35 los valores de precisión de todos los modelos LSTM.



**Figura 34. Precisión del modelo RMS\_3\_15 Fuente: Elaboración Propia**



**Figura 35. Precisión de los modelos LSTM Fuente: Elaboración Propia**

De esta última gráfica puede extraerse información muy interesante como la clara mejoría en relación con los modelos ARIMA, así como comportamientos entre los propios LSTM. Antes de entrar en profundidad en el siguiente apartado del capítulo, puede apreciarse, por ejemplo, como el optimizador RMS da mejores resultados en todos los modelos equivalentes de cada uno.

---

## 5.2 Análisis sobre resultados

---

Llegando ya al final de este trabajo, es el momento de repasar los resultados obtenidos por los modelos planteados y sobre cuáles son los más recomendables, así como cualquier apreciación sobre los mismos considerada como relevante para entender los datos sobre el mercado de las naranjas y cítricos.

Empezando por los modelos ARIMA, la primera conclusión es clara y concisa según todas las pruebas realizadas en el apartado anterior, el mejor modelo de entre todos ellos es el ARIMA(1,0,1) propuesto por pmdarima. De esto puede extraerse más información aparte de este dato concreto.

La gran diferencia en todos los coeficientes confirma lo que ya se expuso en el estado del arte del capítulo 2, pmdarima permite realizar la labor de ajuste para encontrar los mejores parámetros de ajuste sin apenas esfuerzo comparativo, superando con creces al modelo básico de statsmodel. Cabe recordar que, si se usase este último con los coeficientes adecuados, el resultado sería el mismo, pero se ha demostrado que la labor de búsqueda llevada por el proceso de automatización cumple de sobrada manera su función y mejora notablemente los criterios de información, la fiabilidad de los coeficientes del modelo y la precisión a la hora de hacer predicciones.

Además, otra información importante para comparar que los dos modelos de pmdarima nos proporcionan es que no existe un patrón de estacionalidad lo bastante relevante en los datos como para favorecer el modelo, ya que es el modelo no estacional el mejor de los dos optimizados. Esto no implica necesariamente que no exista un componente de este tipo en el precio de las naranjas, pero de existir se ve diluido por otros factores más relevantes y acaba por no aportar información de calidad al modelo.

Esta información respalda lo que ya se adelantó en el capítulo cuatro sobre Prophet, que al ser una herramienta basada en los distintos componentes estacionales y ser esta serie temporal mensual y haberse comprobado que no tiene componentes apreciables de ese tipo, la herramienta pierde toda su precisión y resulta inútil para estos datos concretos.

Por el lado de los modelos LSTM, hay aún más conclusiones que extraer, empezando por lo que ya se comentó sobre qué hiperparámetros de la red son más beneficiosos para la fijación de precios. Para esto se ha decidido centrarse en la precisión de las predicciones más que en el loss, ya que se estima más característico y representativo de lo que se pretende.

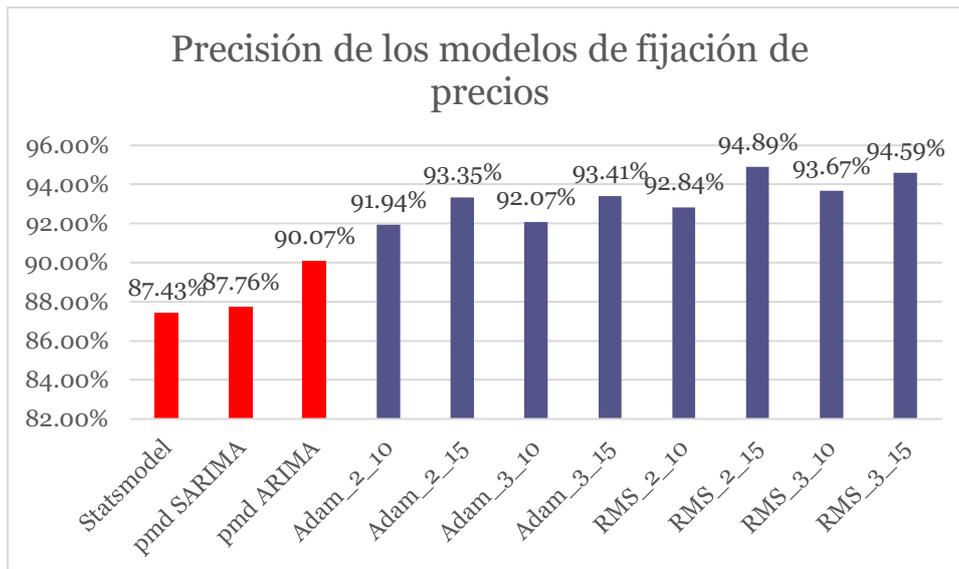


En cuanto a la estructura de los modelos, puede extraerse mucha información de la figura 35 sobre cuáles son las mejores decisiones a la hora de montar los modelos antes del entrenamiento:

- El optimizador: como se adelantó en el apartado anterior del capítulo, el optimizador RMSprop ha demostrado adaptarse mejor a esta serie temporal que la alternativa Adam, ya que en todas las combinaciones demuestra ser superior.
- Número neuronas en las capas ocultas: queda demostrado que todos los modelos con 15 neuronas por capa mejoran a sus contrapartidas de solo 10, además que no se haya demostrado ninguna influencia significativamente negativa en cuanto a la velocidad de entrenamiento de los mismos.
- El número de capas ocultas: este campo ofrece resultados inconsistentes, ya que en tres de las cuatro comparativas el aumento de capas mejora los resultados, pero muy ligera o prácticamente nada, y en el cuarto que incluye los mejores resultados, es el de 2 capas el que resulta ser superior con un 94,89 contra 94,59.

En definitiva, las conclusiones claras son que lo mejor para la configuración de un modelo LSTM para esta serie temporal es el optimizador RMSprop con 15 neuronas por capa, siendo los dos modelos correspondientes a esto los más precisos de todo el trabajo en cuanto a las predicciones.

Como consideración adicional cabe mencionar que, con intención de no estar dejándose escenarios potencialmente mejores fuera del análisis, se han hecho pequeñas pruebas aumentando aún más tanto el número de neuronas como de capas, dando resultados más lentos y con peores precisiones que los ya expuestos, especialmente si se intenta añadir una cuarta capa oculta de neuronas. Es por esto por lo que los ocho modelos expuestos en el capítulo anterior y en este son considerados como los mejores y más representativos del estudio de los datos. Para ir concluyendo con el análisis, se han unificado las representaciones de las precisiones de todos los modelos para visualizar los resultados generales de este capítulo en cuanto a la calidad de todos los modelos de fijación, estos datos aparecen aunados en la figura 36.



**Figura 36. Precisión de los modelos LSTM Fuente: Elaboración Propia**

En esta gráfica se confirma la propuesta teórica del capítulo dos sobre la mejor capacidad predictiva de las redes neuronales recurrentes (RNN) en general y las LSTM en particular sobre la aproximación tradicional de los ARIMA, además de permitir apreciar cuales son los mejores modelos dentro de su tipo y las precisiones de los modelos de cara a la fijación de precios, que es el objetivo último de este trabajo.

Para finalizar y a modo de resumen de este capítulo y del análisis de los resultados de los modelos **podemos concluir lo siguiente:**

- El proceso auto-arima de pmdarima consigue lo que se esperaba de esta eficiente herramienta y encuentra el mejor ajuste posible para el marco ARIMA.
- La serie temporal de los precios de la naranja no demuestra un componente estacional fuerte, lo que empeora la funcionalidad y la precisión de los modelos Seasonal ARIMA y Prophet, especialmente este último que resulta ser prácticamente inaplicable para estos datos.
- Los modelos RNN/LSTM demuestran tener un mejor comportamiento general que los modelos ARIMA y obtienen una mayor precisión a la hora de fijar los precios de acuerdo con las pruebas realizadas.
- El optimizador RMSprop y sus hiper parámetros propios demuestran una mejor adaptación a esta serie temporal que su alternativa Adam.
- El uso de 15 neuronas por capa oculta de las redes demuestra ser mejores a la hora de entrenar y probar los modelos.

- No se aprecia un impacto positivo significativo en aumentar el número de capas ocultas de 2 a 3, y aumentarlo a 4 perjudica fuertemente tanto al tiempo de entrenamiento como a la precisión del modelo.
- Todas las previsiones prevén una ligera bajada de precios en el mercado de la naranja para la próxima etapa.
- El mejor modelo según el ajuste de precisión y el segundo mejor según el Mean Square Error es el modelo **RMS\_2\_15**, con una predicción a un mes vista de **0,88746** €/kg de naranjas

Para finalizar este capítulo y antes de dar paso a las conclusiones del trabajo de fin de grado, cabe destacar que, aunque el proceso de estudio y aplicación práctica se haya realizado sobre una serie temporal concreta de las naranjas en los últimos 20 años, gran parte de todo el trabajo y los últimos dos capítulos podrían aplicarse a otros productos cítricos y de frutas para obtener nuevas conclusiones, y así aportar información útil a la toma de decisión que beneficie a sectores distintos sin un gran esfuerzo extra y con facilidad.

## 6. Conclusiones y Trabajo Futuro

---

### 6.1 Conclusiones

---

A lo largo de este trabajo de fin de grado se han planteado una serie de conceptos y aplicaciones prácticas con la intención de cumplir los objetivos planteados en la introducción. En este apartado se expone una breve síntesis sobre todo lo conseguido en relación con estos y sobre sí, como conclusión del trabajo, se consideran alcanzados.

Empezando por el estado del arte, se han planteado y explicado algunas de las principales herramientas informáticas que hay en el sector a la hora de trabajar con fijación de precios en series temporales. Se han presentado statsmodel y pmdarima como herramientas capaces de aplicar con gran eficacia el estándar estadístico que supone los modelos ARIMA, Prophet como una alternativa dentro de lo más tradicional y, sobre todo, la base teórica detrás de las RNN y LSTM como aplicación de la inteligencia artificial vanguardista al campo en cuestión.

Después, en el análisis del problema, se ha expuesto al lector el estado actual y de los últimos años del sector de los cítricos en España y la Comunidad Valenciana, buscando así familiarizar con este y entender la importancia que puede tener la aplicación exitosa de las tecnologías planteadas. Mostrando las grandes cifras de negocio, la predominante posición en el mercado global y el gran peso en la balanza de pagos española, queda claro que se trata de un sector relevante para el entorno, de acuerdo con lo también planteado en la motivación de este trabajo.

Ya en lo relativo al diseño y la implementación en el trabajo, se ha llevado a cabo una aplicación práctica de las tecnologías planteadas. Basándose en unos datos históricos reales de 20 años, se ha programado una implementación para todo lo presentado en el estado del arte, así como para algunas variaciones dentro de estas herramientas, dando lugar a un total de 12 modelos distintos, cada uno con su ajuste y sus diferentes predicciones para el futuro.

Estos modelos también han sido sometidos a dos principales vías de validación: un análisis teórico para compararlos en términos de datos propios del modelo (como el error del ajuste o el criterio de información de Akaike), y un análisis basado en probar la precisión de las predicciones realizadas por cada uno en relación con unos datos reservados para la prueba. Así, se ha podido medir la eficacia de los modelos y compararlos entre sí.



Para finalizar, esta comparación y análisis de los modelos explicados y aplicados ha permitido comprobar que realmente cumplían su función esperada, dando lugar a predicciones fiables, además de poder elegir el mejor entre ellos, el RMS\_2\_15. Con esto último, se da por alcanzado el objetivo principal del trabajo, habiendo obtenido un modelo testeado como válido para realizar unas estimaciones fiables del precio de las naranjas a un mes vista. Además, se considera que este modelo, así como los comparables, son capaces de aportar información útil tanto para las naranjas como otros cítricos o frutas, estando en consonancia también con la motivación que dio lugar a este trabajo, proponer herramientas informáticas basadas en la predicción capaces de contribuir a la toma de decisiones en un entorno de incertidumbre.

---

## 6.2 Relación con estudios cursados

---

Como reflexión sobre la relación de este trabajo con los estudios realizados durante el grado, son varias las asignaturas que se pueden relacionarse con lo aplicado en este, y a su vez formar parte de la motivación para el mismo.

La asignatura más relevante y relacionada con el trabajo ha sido sin lugar a duda Sistemas inteligentes. Esta ha servido de toma de contacto con las tecnologías de aprendizaje e inteligencia artificial que, y aunque, por la composición del doble grado, no se pudiera seguir por esa rama, el interés despertado al cursarla puede verse reflejado en la motivación que ha dado lugar a este trabajo, descrita en la introducción. Otras relaciones que pueden encontrarse en este trabajo con los estudios del grado son:

- **Lenguajes, tecnologías y paradigmas de la programación:** una entrada más en detalle de la informática como ingeniería, así como mostrar paradigmas de programación como el funcional que comparte Python.
- **Estructuras de datos y algoritmos:** explica más en detalle la importancia de la eficiencia y la optimización a la hora de aplicar algoritmos y estructuras, factor que se ha tenido en cuenta a la hora de plantear los modelos del trabajo.
- **Estadística:** el trabajo versa de series temporales, y es imposible hablar de estas sin ver relación en las herramientas estadísticas y de análisis de las mismas vistas en parte en esta parte de los estudios.
- **Econometría:** Aunque se trate de una asignatura correspondiente al grado de ADE, guarda relación con las series temporales y los modelos ARIMA, empleados en el trabajo, además de convalidar para el grado de Ingeniería Informática.

---

## 6.3 Trabajo Futuro

---

Por último, y como final de este proyecto, se ha considerado interesante proponer como sería más adecuado continuar con el trabajo a partir de este punto. El uso de técnicas predictivas para la toma de decisión ya se ha comentado muchas veces aquí como extremadamente útil, y puede ser muy interesante seguir con esta vía para apoyar a las empresas locas y favorecer la economía. Como posibles trabajos futuros a plantear se han considerado dos vías distintas, pero igual de interesantes.

- **Enfocado al uso:** desde el punto de vista de la usabilidad y facilidades para usuarios, puede ser muy interesante unificar todas estas herramientas en una única aplicación o aplicación web que permita a usuarios sin ningún conocimiento informático acceder a los modelos de interés. De este modo, se crearía una herramienta mucho más fácil de distribuir y utilizar, favoreciendo su extensión a distintos agentes y sectores.
- **Enfocado a la predicción:** pese a que se considera que se han obtenido unas predicciones bastante fiables, siempre se puede mejorar. La forma más interesante de hacer esto se considera que es añadiendo datos adicionales a la serie temporal como tal que permitieran extraer una información complementaria, como pueden ser stocks, datos macro o factores medioambientales. Aportando esta información exógena mediante modelos ARIMAX o nuevos formatos de entrada para los LSTM, puede mejorarse la capacidad de predicción, mejorando así aún más la información disponible.

Estas dos vías de trabajos futuros se plantean como caminos para continuar con la motivación base del trabajo y los objetivos cumplidos, ofrecer herramientas sencillas pero potentes que ayuden en la toma de decisiones y puedan mejorar la actividad y la economía del entorno.



## Referencias bibliográficas

Akaike, H. (1974). A new look at the statistical model identification. Publicado en IEEE Transactions on Automatic Control ( Volume: 19, Issue: 6, December 1974).

<https://ieeexplore.ieee.org/document/1100705>

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., Ljun, M. (1976). Time series analysis: Forecasting and control. Holden-Day, San Francisco. Editorial Wiley.

<https://books.google.es/books?hl=es&lr=&id=rNt5CgAAQBAJ&oi=fnd&pg=PR7&dq=time+series+analysis:+Forecasting+and+control&ots=DK68xOj2Vx&sig=3Xajmsw8ISILZ9gednXobM7L7Bs#v=onepage&q&f=false>

Conda. (2022). Conda documentation. <https://www.conda.io>

EcuRed. (2022). Python. <https://www.ecured.cu/Python#Historia>

Eurostat. (2022). Database.

<https://ec.europa.eu/eurostat/web/main/data/database>

Daniel, R. (6 de febrero de 2020). ¿A dónde van los márgenes que no reciben los agricultores ? El Economista.

<https://www.eleconomista.es/empresas-finanzas/agro/noticias/10339676/02/20/2/A-donde-van-los-margenes-que-no-reciben-los-agricultores.html>

Harrison, M. (2017). Illustrated Guide to Python 3: A Complete Walkthrough of Beginning Python with Unique Illustrations Showing how Python Really Works. Editorial Hairysun.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-term Memory. Neural Computation, 9(8), 1735-1780.

<https://doi.org/10.1162/neco.1997.9.8.1735>

Hopfield, J. (1982). Neural networks and physical systems with emergent collective computational abilities. *PNAS*, 79(8), 2554-2558.

<https://doi.org/10.1073/pnas.79.8.2554>

IndexMundi. (2022). Naranjas Precio Mensual - Euro por Kilogramo. Consulta 28 de Julio 2022.

<https://www.indexmundi.com/es/precios-de-mercado/?mercancia=naranjas&meses=240&moneda=eur>

Jet Brain. (2022). PyCharm. IDE de Python para desarrolladores profesionales.

<https://www.jetbrains.com/es-es/pycharm/>

Kimoto, T., & Asakawa, K. (1990). Stock market prediction system with modular neural networks. International Joint Conference on Neural Networks.

Makridakis, S. (2018). The M4 Competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, 34(4), 802-808.

<https://doi.org/10.1016/j.ijforecast.2018.06.001>

Ministerio de Agricultura, Pesca y Alimentación (MAPA). (15 de junio 2022). Desarrollo de la campaña de cítricos 2021/2022.

[https://www.mapa.gob.es/es/agricultura/temas/producciones-agricolas/informesobredesarrollodelacampana2021-22decitricosa15juniode2022\\_tcm30-623033.pdf](https://www.mapa.gob.es/es/agricultura/temas/producciones-agricolas/informesobredesarrollodelacampana2021-22decitricosa15juniode2022_tcm30-623033.pdf)



Lonja de Valencia. (2022). Informe de cotizaciones de la mesa de precios cítricos del consulado de la lonja de Valencia desde Agosto 2021 hasta Junio 2022.

[https://www.camaravalencia.com/es-ES/Documents/mesa\\_citricos.pdf](https://www.camaravalencia.com/es-ES/Documents/mesa_citricos.pdf)

Olah, C. (27 de August de 2015). Understanding LSTM Networks.

<https://colah.github.io/posts/2015-08-Understanding-LSTMs/>

Pik, J., Ghosh, S. (2021). Hands-On Financial Trading with Python, A practical guide to using Zipline and other Python libraries for backtesting trading strategies. Editorial Packt.

Pmdarima. (2022). About the project. Project page de pmdarima en alkaline-ml.

<http://alkaline-ml.com/pmdarima/about.html#about>

Portal Fruticola (n.d.) La Comunidad Valenciana se ubica como la principal región exportadora de cítricos del mundo.

<https://www.portalfruticola.com/noticias/2022/02/14/espana-la-comunidad-valenciana-se-ubica-como-principal-region-exportadora-de-mandarinas-y-naranjas-en-el-mundo/?pdf=469037>

Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning Internal Representations by Error Propagation. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, 1, 318-362.

<https://doi.org/10.1016/B978-1-4832-1446-7.50035-2>

Siarni-Namini, S., & Siarni Namin, A. (2018). Forecasting Economics and Financial Time Series: ARIMA vs. LSTM. *ArXiv*.

<https://doi.org/10.48550/arXiv.1803.06386>

Statsmodel Organization. (2022). About Statsmodels.

<https://www.statsmodels.org/stable/about.html#>

Yan, Y. (2014). Python for Finance: Apply powerful finance models and quantitative analysis with Python. Editorial Packt



## ANEXO

### OBJETIVOS DE DESARROLLO SOSTENIBLE

Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS).

<b>Objetivos de Desarrollo Sostenibles</b>	<b>Alto</b>	<b>Medio</b>	<b>Bajo</b>	<b>No Procede</b>
ODS 1. <b>Fin de la pobreza.</b>		<b>X</b>		
ODS 2. <b>Hambre cero.</b>			<b>X</b>	
ODS 3. <b>Salud y bienestar.</b>				<b>X</b>
ODS 4. <b>Educación de calidad.</b>				<b>X</b>
ODS 5. <b>Igualdad de género.</b>				<b>X</b>
ODS 6. <b>Agua limpia y saneamiento.</b>				<b>X</b>
ODS 7. <b>Energía asequible y no contaminante.</b>				<b>X</b>
ODS 8. <b>Trabajo decente y crecimiento económico.</b>	<b>X</b>			
ODS 9. <b>Industria, innovación e infraestructuras.</b>				<b>X</b>
ODS 10. <b>Reducción de las desigualdades.</b>	<b>X</b>			
ODS 11. <b>Ciudades y comunidades sostenibles.</b>				<b>X</b>
ODS 12. <b>Producción y consumo responsables.</b>	<b>X</b>			
ODS 13. <b>Acción por el clima.</b>				<b>X</b>
ODS 14. <b>Vida submarina.</b>				<b>X</b>
ODS 15. <b>Vida de ecosistemas terrestres.</b>				<b>X</b>
ODS 16. <b>Paz, justicia e instituciones sólidas.</b>				<b>X</b>
ODS 17. <b>Alianzas para lograr objetivos.</b>				<b>X</b>

## Reflexión sobre la relación del TFG/TFM con los ODS y con el/los ODS más relacionados.

La relación de este trabajo con los distintos objetivos de desarrollo sostenible puede no saltar a la vista, ya que no se han ido comentando a lo largo del mismo de forma explícita, pero forman parte fundamental del mismo desde el momento inicial mismo. Como ya se dijo en la introducción y la motivación, este trabajo buscaba, a través de la innovación tecnológica, aportar a un sector relevante del entorno en el que tiene lugar.

El motivo por el que se eligió el sector de la naranja y los cítricos es por su gran impacto en la economía, e incluso sociedad, de la de la comunidad Valenciana, para que al aportar herramientas útiles para este, no se beneficiase solo a un grupo de gente en concreto, sino que permease a todo el entorno mismo del sector. Esta motivación es la que mayor relación guarda con los ODS, ya que es el haber cumplido con esta lo que favorece a los mismos. A continuación, se expondrá brevemente cuáles son los ODS más relacionados con este trabajo, y el por qué:

- **Trabajo decente y crecimiento económico:** este trabajo ha pretendido siempre aportar información que mejore la actividad económica de los agricultores, muy castigados en los años recientes. Con esto, mejorar la calidad del trabajo de los mismos al contar con más información que permita una buena toma de decisiones será mucho más sencillo, y a través de su gran peso en la economía producir un crecimiento general de esta.

- **Reducción de las desigualdades:** este trabajo trata de aportar unas herramientas de fácil uso, pero sólida eficacia a unos agricultores que tal vez no cuenten con acceso a grandes capacidades tecnológicas. Estos agricultores se encuentran en una situación constante de desigualdad de recursos frente a competencia más industrializada que cuenta con acceso a otros mecanismos para obtener ventaja, y con este trabajo también se pretende reducir esta desigualdad entre pequeños agricultores y grandes superficies del sector.

- **Producción y consumo responsable:** con el acceso a estas herramientas, los agricultores podrán tomar mejores decisiones en cuanto a que producir, ya que contarán con cierta referencia a futuros precios. De este modo, evitarán producir aquello que se prediga tendrá un exceso de oferta, mejorando así la estructura general de producción reduciendo cosas que ya se han visto en el sector como tirar producto a causa de los bajos precios.

