



Colegio Universitario de Estudios Financieros

Máster Universitario en Ciencia de Datos

RESUMEN EJECUTIVO

MODELOS DE INTERVENCIÓN DE SERIES TEMPORALES: APLICACIÓN AL COVID-19

Alumno: Gómez Rivas, Alejandro José

Tutor: Queralt Sánchez de las Matas, Ricardo A.

Madrid, 9 de septiembre de 2022

1. RESUMEN

A principios del 2020, en China, una nueva variante de enfermedad respiratoria aguda denominada COVID-19 azotó al país por completo y también, en cuestión de pocos meses, al mundo entero. Como consecuencia de esta pandemia, se han cerrado muchas empresas, con una situación de incertidumbre para muchos datos socioeconómicos y bursátiles, por lo que un pronóstico preciso ante una posible futura pandemia o de una situación atípica, se vuelve crucial en tales situaciones.

El principal objetivo de este Trabajo de Fin de Máster es realizar un análisis e implantación de distintos modelos predictivos tradicionales en distintas series temporales para determinar cuáles de ellos tienen un pronóstico más certero del futuro y ver el impacto que ha tenido la pandemia sobre éstas.

2. INTRODUCCIÓN

El presente Trabajo de Fin de Máster tiene por objetivo lo explicado en el resumen: el análisis de distintas series temporales en la época COVID-19 y su previsión en el futuro mediante distintos modelos predictivos.

Este trabajo consiste en una actividad académica en el que se aplican conceptos que se han adquirido a lo largo de la asignatura de “Técnicas de predicción de series temporales” impartida en el Máster Universitario en Ciencia de Datos. Se ha planteado como un desafío personal, puesto que requiere poseer el conocimiento de múltiples conceptos de estadística avanzada para comprender el funcionamiento de los modelos de predicción, además de conocimientos sólidos de Matemáticas y de programación en Python.

Las fases del desarrollo del proyecto son las siguientes:

- Se seleccionan cuatro series temporales como variables objetivo de estudio y se les realiza a cada una un preprocesamiento de los datos para poder realizar correctamente los pasos posteriores.
- Se realiza un análisis EDA aplicando cálculos estadísticos y representaciones gráficas para observar la tendencia, la estacionalidad y otras características.
- Se explica cada uno de los modelos, se aplican para cada serie mediante entrenamiento-test, se comparan sus resultados con los últimos datos reales y se comprueban cuáles de ellos dan menos error y se aplican finalmente para realizar una predicción hasta finales del año vigente.

- Por último, se contrastan sus resultados y se realiza una conclusión general con lo obtenido.

A continuación, se muestra de manera gráfica la estructura:

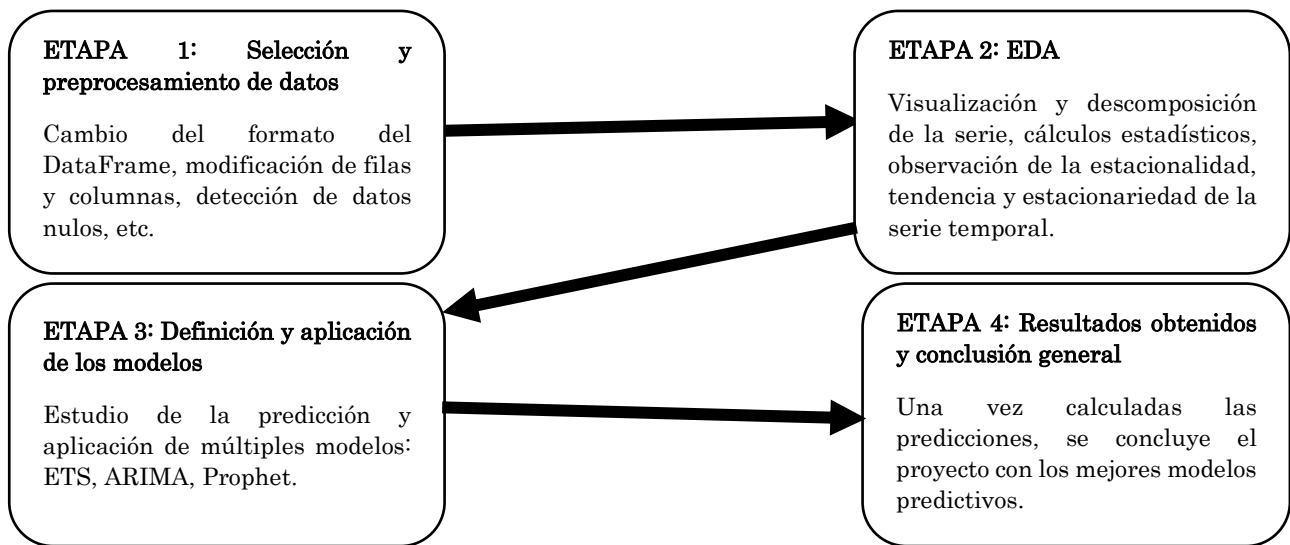


Figura 1: Desarrollo del TFM. Fuente: elaboración propia.

3. MARCO TEÓRICO

3.1. SERIES TEMPORALES

Una serie temporal es una sucesión de observaciones de una variable (p. ej. ingresos, precios, etc.) tomadas en varios instantes de tiempo. Los datos de una serie de tiempo van ligados al instante temporal en el que han sido medidos.

Las series temporales presentan normalmente una o varias características que se denominan como componentes, que ayudan a explicar el comportamiento que tienen las series en el tiempo. Dichos componentes son los siguientes:

- **Tendencia.** Este componente hace referencia al comportamiento o el movimiento de la serie temporal a largo plazo.
- **Componente cíclico.** En este se refleja los comportamientos recurrentes.
- **Componente estacional.** Consiste en movimientos de oscilación de un periodo de tiempo.
- **Componente irregular.** También llamado como “ruido”, recoge las alteraciones de la serie, pequeñas en su incidencia, y sin una pauta periódica ni tendencial reconocible.

Las series temporales se clasifican normalmente en base a su comportamiento temporal de las propiedades estadísticas, y esta es la estacionariedad. Una serie se dice que es estacionaria cuando su distribución y sus parámetros estadísticos no varían con el paso del tiempo. Se mide a menudo con la

media y la varianza, aunque hay métodos estadísticos más eficaces como la prueba de Dickey-Fuller, más desarrollado en el TFM.

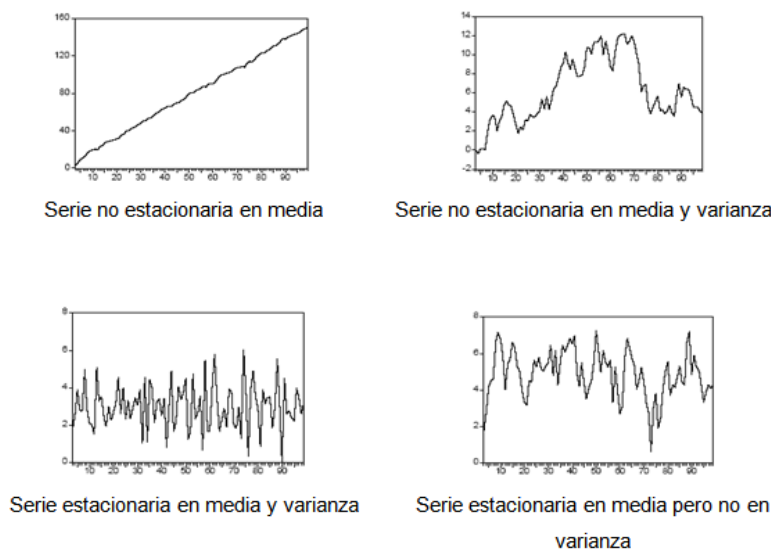


Figura 2: tipos de estacionariedad. Fuente: Estadística y ML con R, Francisco Parra

El análisis de las series engloba un conjunto de técnicas matemáticas y estadísticas que permiten extraer todas las características que se aprecian en los datos históricos y, de esta manera, poder obtener metodologías para predecir los valores futuros.

Para la realización del proyecto, se han seleccionado cuatro series temporales de índole macroeconómico y social: que se ubican dentro del campo de la economía, de la demografía y de la industria, y se han utilizado los datos de manera mensual, ya sea obteniéndolos en las páginas oficiales o bien realizando un muestreo de los mismos.

3.2. MODELOS DE PREDICCIÓN DE SERIES TEMPORALES

En el proyecto se han seleccionado tres modelos predictivos estadísticos que son tradicionales y útiles para los pronósticos de series de tiempo.

- Modelo de Suavizado Exponencial (ETS): se trata de un modelo que utiliza los datos promedios históricos para predecir su comportamiento en el futuro.
- Modelo Autorregresivo Integrado de Media Móvil (ARIMA): es un modelo que permite describir un valor como una función lineal de los datos pasados y errores debidos al azar, en el que se puede incluir un componente cíclico o estacional.
- Modelo PROPHET: es un modelo basado en un modelo aditivo donde las tendencias no lineales son ajustadas anualmente, semanalmente y diariamente para ver la estacionalidad.

Muy útil para conjuntos de datos con observados históricas muy detalladas o eventos conocidos importantes e irregulares.

4. MARCO PRÁCTICO

El procedimiento práctico que se aplica para el análisis de las series temporales es el mismo siempre: lo primero de todo es la obtención de los datos de las series temporales y se les realiza una limpieza y preprocesado de los mismos para poder aplicar fácilmente el código para visualización e interpretación, luego se hace un sencillo análisis exploratorio de las series temporales a través de sus componentes y se aplican los modelos a batir para comprobar cuál de ellas es la que más se aproxima a los datos reales. Además, para comprobar el impacto que ha tenido el COVID-19 en las series temporales, se le realiza a cada serie temporal una intervención suprimiendo dicha etapa y sustituyendo sus datos con los obtenidos en la predicción de datos pasados mediante entrenamiento-test del modelo predictivo que mejor precisión tuvo en el modelo original.

Se han seleccionado para el proyecto cuatro series temporales, las cuales se han cogido los datos que se alojan en las páginas oficiales del gobierno español. Estas series temporales son el número de desempleados, el número de turistas, la matriculación de vehículos y el número de hipotecas.

5. RESULTADOS DE LAS PREDICCIONES

5.1. SERIE TEMPORAL 1: DESEMPLEO

Para la serie temporal original, se predijo mejor con el modelo ETS con un MAPE del 12,08%, gracias a la estacionalidad marcada, la tendencia constante (a excepción del momento COVID-19) y su característica de estacionariedad. Aplicando este modelo para previsión a finales del 2022, se obtiene un pronóstico con tendencia creciente y coincidente con el de los datos pasados.

Para la serie intervenida en el momento del COVID-19, se predijo mejor también con el modelo ETS y además obteniendo un error mínimo de aproximadamente un 1%, lo cual predice con mucha exactitud. Los resultados obtenidos para finales del 2022 son bastante similares a las de la serie original.

Se puede observar la gráfica y los datos obtenidos con más detalle en el Informe final, tanto de esta serie temporal como las del resto.

5.2. SERIE TEMPORAL 2: TURISMO

En la serie temporal original se predijo mejor con el modelo ETS, ya que guarda unas características muy similares a las de la serie temporal de desempleo. Como previsión para finales del 2022, se obtiene un pronóstico con tendencia creciente y unos valores demasiado altos en comparación con los de los datos pasados.

En el caso de la serie intervenida en el momento del COVID-19, se predijo mejor también con el modelo ETS obteniendo un MAPE del 8,6%, un error bastante bajo gracias a la eliminación del efecto de la pandemia. Los resultados obtenidos para finales del 2022 son bastante más similares a las obtenidas en datos pasados.

5.3. SERIE TEMPORAL 3: MATRICULACIONES

Para la serie temporal original, se predijo mejor con el modelo ETS, pero en general los resultados obtenidos por cada uno de los modelos no fueron muy buenos, ya que obtuvieron unos errores de alrededor del 42% debido a su volátil tendencia. Aplicando este modelo para previsión a finales del 2022, se obtiene un pronóstico con tendencia creciente pero un poco incierto ya que no cumple con el patrón estacional de la serie temporal.

Para la serie intervenida en el momento del COVID-19, se predijo mejor con el modelo ARIMA con un error del 16% aproximadamente, lo cual indica que ha mejorado. Los resultados obtenidos para finales del 2022 son bastante similares a las de la serie original.

5.4. SERIE TEMPORAL 4: HIPOTECAS

En la serie temporal original, se predijo mejor con el modelo PROPHET con un MAPE del 15%, y su previsión a finales del 2022, se obtiene un pronóstico con tendencia volátil no coincidente con la de los datos pasados.

Para la serie intervenida, ARIMA es el modelo que mejor predice y los resultados obtenidos para finales del 2022 guardan similitud con la de los datos históricos.

6. CONCLUSIONES

Al haber obtenido los resultados anteriores se han llegado a obtener varias conclusiones.

La primera es que el modelo ETS es el modelo predictivo estadístico que mejores resultados han obtenido y más se ha repetido para realizar la predicción para finales del año vigente, ya que casi todas las series temporales que se han empleado cumplen con el requisito de una estacionalidad

marcada. En el caso del modelo de ARIMA, se ha comportado mejor con series temporales que cumplen con la condición de no estacionariedad, como ha sido el caso de las matriculaciones de vehículos.

La segunda deducción es que la introducción del modelo PROPHET para predecir las series temporales, han obtenido unos resultados inferiores a los que se esperaban, a excepción de los resultados obtenidos en la serie temporal original de las hipotecas. No han sido buenas predicciones por lo general porque este modelo es más preciso para modelos que tengan puntos de datos atípicos grandes, eventos irregulares o con tendencias de crecimiento no lineal que se aproximen a un límite.

Como conclusión final y general, se ha comprobado que la intervención realizada en las series temporales eliminando la parte afectada por el efecto del COVID-19 ha obtenido unos resultados más precisos que en los resultados obtenidos en las series originales. Esto significa que los modelos predictivos tradicionales, basados en la estadística, fallan cuando hay un fenómeno temporal irregular, como es el COVID-19, pero son modelos óptimos para emplearlos en series temporales con una estacionalidad o una tendencia marcada. Para series temporales que no cumplan con estos requisitos, habría que recurrir al uso de modelos predictivos basados en redes neuronales o en Machine Learning.