

**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**

Este documento resume el trabajo realizado en el proyecto. El principal objetivo de éste es la aplicación de distintos modelos de análisis y de predicción para diversas series temporales y ver el comportamiento que obtienen tras haber pasado el momento inicial del COVID-19. Los resultados obtenidos han sido favorables para algunas series de tiempo, aunque para otras han obtenido predicciones dudosas debido a su incierta tendencia o estacionalidad.

Finalmente, aunque algunos resultados fueran poco fiables, el hecho de probar y ajustar mediante métodos de predicción como es el Facebook Prophet o de Redes Neuronales pueden prometer logros para una tarea tan compleja como predecir el comportamiento que pueda tener un activo financiero o un dato macroeconómico ante tales situaciones extraordinarias y saber cómo actuar ante ellas.

En estos tiempos de gran incertidumbre, cualquier herramienta de predicciones cuenta, y las metodologías actuales deben ser testeadas y explotadas al máximo para comprobar sus resultados y para seguir desarrollando modelos aún más eficientes a partir de ellas.

1. **Introducción**

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

Este trabajo, en definitiva, se trata de una actividad académica en el que se han aplicado 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, por lo que no se trata de una investigación científica ni una divulgación sobre cómo aplicar dichos modelos para cada serie de tiempo en un momento específico.

A continuación, se puntualizan las fases del trabajo:

* Se seleccionan distintas series temporales como variables objetivo de estudio y se realiza 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 aplicarán para realizar una predicción hasta finales del año vigente.
* Por último, se contrastan sus resultados y se realiza una conclusión general con lo obtenido.

Este proyecto plantea un desafío personal puesto que se requiere el manejo de múltiples conceptos de Machine Learning para comprender el funcionamiento de los modelos de predicción, además de conocimientos sólidos de Matemáticas y de Estadística Avanzada.

A continuación, se muestra de manera gráfica la estructura:

**ETAPA 2: EDA**

Visualización y descomposición de la serie, cálculos estadísticos, observación de la estacionalidad, tendencia y estacionariedad de la serie temporal.

**ETAPA 1: Selección y preprocesamiento de datos**

Cambio del formato del DataFrame, modificación de filas y columnas, detección de datos nulos, etc.

**ETAPA 4: Resultados obtenidos y conclusión general**

Una vez calculadas las predicciones, se concluye el proyecto con los mejores modelos predictivos.

**ETAPA 3: Definición y aplicación de los modelos**

Estudio de la predicción y aplicación de múltiples modelos: ETS, ARIMA, Prophet.

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

1. **Marco teórico**
   1. **Series temporales**

Una serie temporal consiste en una secuencia de observaciones ordenados que están ordenados de manera cronológica. Los datos de una serie de tiempo van ligados al instante temporal en el que han sido medidos.

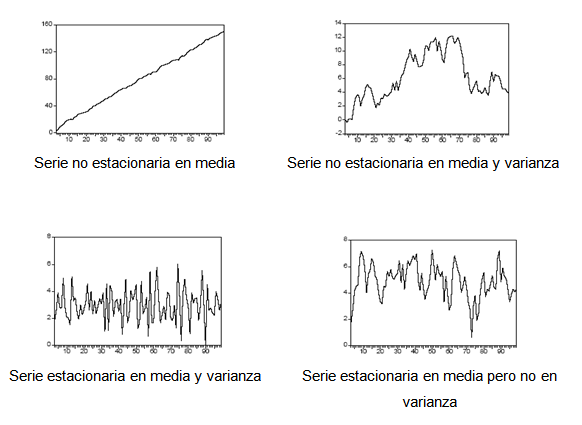
El análisis de las series engloba un conjunto de técnicas matemáticas y estadísticas que permiten extraer todas las características que se aprecian en los datos históricos y, de esta manera, poder obtener metodologías para predecir los valores futuros.

Las series temporales se encuentran involucrados en numerosos estudios de carácter científico-técnico. Algunos de estos campos en los que se encuentran presentes son:

* Economía: como el valor de los activos bursátiles, los ingresos de una empresa o los datos macroeconómicos como el PIB.
* Demografía: puede ser el crecimiento de la población, la tasa de natalidad o el desempleo.
* Medicina: algunos ejemplos son los electrocardiogramas, la evolución de una patología o el tabaquismo.
* Industria: la evolución logística de una empresa, la predicción del consumo de un bien o la pernoctación hotelera de un país.
* Meteorología: la evolución de la temperatura, pluviómetros, etc.

Para la realización del proyecto, se han seleccionado especialmente series temporales 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.

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*

* 1. **Modelos**

En el proyecto se han seleccionado tres modelos de predicciones que son muy representativos en 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. Por eso se ha aplicado para el caso de este proyecto.

1. **Aplicación**

Para la aplicación y el desarrollo de los modelos definidos anteriormente, se realiza un entrenamiento con los datos históricos y un test con los últimos meses, en la que se incluye el periodo COVID-19. En algunos de los resultados obtenidos por los modelos se dista mucho del evento extraordinario que sucedió, por lo que se interviene la serie temporal y se realiza otro entrenamiento-test con un modelo específico para sustituir los datos en el momento COVID-19 y así poder realizar una predicción más precisa para la serie temporal concreta, pero teniendo en cuenta que esa serie temporal es más impredecible a la hora de obtener resultados precisos en eventos importantes y/o irregulares.

1. **Resultados**

Pendiente de acabarlo cuando se tengan todas las series temporales predichas y concluidas.

1. **Conclusiones**

Pendiente de acabarlo cuando se tengan todas las series temporales predichas y concluidas.