



Facultad de Ciencias económicas y Empresariales

# **Análisis cuantitativo sobre la cotización de International Airlines Group (IAG) antes y después del COVID-19**

Autor: Patricia Delgado del Castillo

Tutor: Luis Ángel Calvo Pascual

Grado: E3 Analytics

Madrid | Junio 2023

## RESUMEN

La declaración de una pandemia mundial por Covid-19 en marzo de 2020 dejó un panorama desafiante en el sector turístico y, en particular, en la industria de las aerolíneas comerciales. La demanda de vuelos experimentó una fuerte caída y la cadena de suministro se vio interrumpida en su totalidad. En este contexto, el presente trabajo se propone analizar de manera cuantitativa la cotización de International Airlines Group antes y después del Covid-19, considerando el impacto significativo que la pandemia tuvo en la economía mundial y, en particular, en el sector turístico.

Entre los análisis realizados se ha llevado a cabo la identificación de las variables más relevantes para predecir la cotización de IAG, observando la variación que ocurre según se hayan escogido los datos previos o posteriores a la pandemia. Además, se han evaluado diversos modelos de predicción, destacando el uso de *GPR* en regresión, así como los enfoques de *ensemble* y *quadratic discriminant* en clasificación. Por último, se ha buscado cuantificar el impacto real de la pandemia en la cotización de IAG mediante la utilización de la técnica de *regression discontinuity*. Este análisis ha revelado un efecto negativo significativo en el precio de cierre de la compañía debido a la pandemia a diferencia del volumen, que mostró un incremento sustancial.

**Palabras clave:** aerolíneas, *mutual information*, *machine learning*, predicciones, *regression discontinuity*

## **ABSTRACT**

The declaration of a global pandemic due to COVID-19 in March 2020 posed significant challenges for the tourism sector, particularly to the commercial airline industry. The demand for flights experienced a sharp decline, and the supply chain was completely disrupted. In this context, the present study aims to quantitatively analyze the stock performance of International Airlines Group before and after the COVID-19 outbreak, considering the significant impact the pandemic had on the global economy and the tourism sector.

Among the conducted analyses, the identification of the most relevant variables for predicting the stock performance of IAG was carried out, observing variations based on whether the data before or after the pandemic was chosen. Furthermore, various prediction models were evaluated, with notable emphasis on using GPR for regression tasks, as well as ensemble and quadratic discriminant approaches for classification tasks. Lastly, the study sought to quantify the actual impact of the pandemic on IAG's stock performance using the regression discontinuity technique. This analysis revealed a significant negative effect on the company's closing price due to the pandemic, while the volume showed a substantial increase.

**Key words:** airlines, mutual information, machine learning, predictions, regression discontinuity.

# ÍNDICE

|  |           |
|--|-----------|
| CAPÍTULO 1: INTRODUCCIÓN .....   | 9         |
| <b>1.1. Contexto.....</b>  | <b>9</b>  |
| 1.1.1. Covid-19, turismo y líneas aéreas .....   | 9         |
| 1.1.2. La compañía IAG .....   | 14        |
| <b>1.2. Antecedentes.....</b>  | <b>16</b> |
| <b>1.3. Objetivos.....</b>   | <b>19</b> |
| <b>1.4. Metodología.....</b>   | <b>20</b> |
| <b>1.5. Estructura del TFG .....</b>   | <b>21</b> |
| CAPÍTULO 2: FEATURE SELECTION.....   | 22        |
| <b>2.1. Mutual information .....</b>   | <b>22</b> |
| <b>2.2. Sobre la disparidad entre los tests clásicos y <i>mutual information</i> .....</b> | <b>24</b> |
| <b>2.3. Selección de la muestra .....</b>  | <b>25</b> |
| <b>2.4. Feature selection sobre IAG .....</b>  | <b>29</b> |
| 2.4.1. Antes del Covid-19 .....  | 30        |
| 2.4.2. Después del Covid-19.....   | 31        |
| <b>2.5. Comparativa resultados .....</b>   | <b>32</b> |
| CAPÍTULO 3: MODELOS MACHINE LEARNING REGRESIVOS .....                                      | 34        |
| <b>3.1. Preparación de los datos .....</b>   | <b>34</b> |
| <b>3.2. Optimización bayesiana .....</b>   | <b>34</b> |
| 3.2.1. Antes del Covid-19 .....  | 37        |
| 3.2.2. Durante el Covid-19.....  | 43        |
| <b>3.3. Predicción de la cotización de IAG .....</b>                                       | <b>46</b> |
| 3.3.1. Antes del Covid-19 .....  | 46        |
| 3.3.2. Durante el Covid-19.....  | 48        |
| CAPÍTULO 4: MODELOS ML CLASIFICATORIO .....  | 50        |
| <b>4.1. Preparación de los datos .....</b>   | <b>50</b> |
| <b>4.2. Optimización bayesiana MATLAB.....</b>   | <b>51</b> |
| 4.2.1. Antes del Covid-19 .....  | 51        |
| 4.2.2. Después del Covid-19.....   | 56        |
| <b>4.3. Predicción de la cotización de IAG .....</b>                                       | <b>59</b> |
| 4.3.1. Antes del Covid-19 .....  | 59        |
| 4.3.2. Durante el Covid-19.....  | 60        |
| <b>4.4. Series temporales .....</b>  | <b>60</b> |
| 4.4.1. Serie temporal en MATLAB .....  | 62        |
| 4.4.2. Serie temporal en PYTHON .....  | 65        |
| CAPÍTULO 5. REGRESSION DISCONTINUITY .....   | 68        |
| <b>5.1. Sobre este modelo .....</b>  | <b>68</b> |
| <b>5.2. Análisis de resultados de IAG.....</b>   | <b>69</b> |
| <b>5.3. Comparativa de IAG con otras variables.....</b>                                    | <b>71</b> |
| CAPÍTULO 6: CONCLUSIONES .....   | 77        |
| BIBLIOGRAFÍA .....   | 81        |

## LISTADO DE ABREVIATURAS

- **IAG:** International Airlines Group
- **UNWTO:** World Tourism Organization
- **OMS:** Organización Mundial de la Salud
- **OACI:** Organización de Aviación Civil Internacional
- **PIB:** Producto interior bruto
- **ANSP:** Proveedor de servicios de navegación aérea
- **ALTA:** Asociación de Transporte Aéreo de América Latina
- **ML:** Machine Learning
- **RD:** Regression discontinuity
- **MI:** Mutual Information
- **KL:** Kullback-Leibler
- **GPR:** Gaussian Process Regression
- **RMSE:** Root-mean-square error
- **PCA:** Principal Component Analysis
- **Curva ROC:** Receiver Operating Characteristic
- **AUC:** Area Under the Curve
- **ARIMA:** Autoregressive integrated moving average

## ÍNDICE DE GRÁFICAS

|   |    |
|---|----|
| <b>Gráfico 1.</b> Variación por región de las llegadas de turistas internacionales en 2020 en comparación con 2019 .....                              | 9  |
| <b>Gráfico 2.</b> Variación por región de las llegadas de turistas internacionales en 2021 comparación con 2020 .....                                 | 10 |
| <b>Gráfico 3.</b> Variación por región de las llegadas de turistas internacionales en 2022 comparación con 2021 .....                                 | 11 |
| <b>Gráfico 4.</b> Variación por región de las llegadas de turistas internacionales en 2023 comparación con 2022 .....                                 | 11 |
| <b>Gráfico 5.</b> Comparativa de llegadas de turistas internacionales: globalmente (representado en miles) vs. España.....                            | 12 |
| <b>Gráfico 6.</b> Análisis de pasajeros transportados por aerolínea de IAG en los primeros trimestres de cada año.....                                | 16 |
| <b>Gráfico 7.</b> Dependencia de la variable objetivo respecto a tres características a través de los estadísticos F-test y MI .....                  | 25 |
| <b>Gráfico 8.</b> Resultados mutual information antes del Covid-19 .....  | 31 |
| <b>Gráfico 9.</b> Resultados mutual information después del Covid-19 .....  | 32 |
| <b>Gráfico 10.</b> Líneas de canal .....  | 37 |
| <b>Gráfico 11.</b> Comparación entre las predicciones Exponenciales GPR y valores reales de las cotizaciones de IAG .....                             | 41 |
| <b>Gráfico 12.</b> Tendencia de las predicciones Exponenciales GPR y discrepancia con los valores reales de las cotizaciones de IAG .....             | 42 |
| <b>Gráfico 13.</b> Residuos Exponential GPR .....   | 42 |
| <b>Gráfico 14.</b> Comparación entre las predicciones Cuadráticas Exponenciales GPR y valores reales de las cotizaciones de IAG .....                 | 44 |
| <b>Gráfico 15.</b> Tendencia de las predicciones Cuadráticas Exponenciales GPR y discrepancia con los valores reales de las cotizaciones de IAG ..... | 45 |
| <b>Gráfico 16.</b> Residuos Squared Exponential GPR .....   | 45 |
| <b>Gráfico 17.</b> Predicciones Exponential GPR.....  | 47 |
| <b>Gráfico 18.</b> Predicciones Squared Exponential GPR .....   | 49 |
| <b>Gráfico 19.</b> Matriz de confusión, Ensemble .....  | 54 |
| <b>Gráfico 20.</b> Curva ROC, Ensemble .....  | 55 |
| <b>Gráfico 21.</b> Matriz de confusión, Quadratic Discriminant .....  | 58 |

|  |    |
|--|----|
| <b>Gráfico 22.</b> Curva ROC, Quadratic Discriminant .....   | 58 |
| <b>Gráfico 23.</b> Predicciones Ensemble, Matriz de Confusión .....  | 59 |
| <b>Gráfico 24.</b> Predicciones Quadratic Discriminant, Matriz de Confusión .....  | 60 |
| <b>Gráfico 26.</b> Serie temporal, eliminación de la tendencia .....   | 63 |
| <b>Gráfico 27.</b> Serie temporal, eliminación de la estacionalidad .....  | 64 |
| <b>Gráfico 28.</b> Serie temporal, random fluctuation .....  | 64 |
| <b>Gráfico 29.</b> Serie temporal, predicciones .....  | 65 |
| <b>Gráfico 30.</b> Predicciones ARIMA .....  | 67 |
| <b>Gráfico 31.</b> RD utilizando el argumento "ci=95" para mostrar intervalos de confianza del 95% en el gráfico. ....   | 69 |
| <b>Gráfico 32.</b> RD utilizando el argumento "binselect="es"", que representa la elección óptima de bin (partición de los datos) según el error cuadrático medio. ....  | 70 |
| <b>Gráfico 33.</b> RD utilizando el argumento "binselect="qsmv"" para realizar una partición de los datos utilizando el enfoque de particionado QS (particionado de la cuasi-varianza) y la elección de la varianza mimética ..... | 70 |
| <b>Gráfico 34.</b> RD de IAG.....  | 72 |
| <b>Gráfico 35.</b> RD de Air France.....   | 72 |
| <b>Gráfico 36.</b> RD de Lufthansa.....  | 72 |
| <b>Gráfico 37.</b> RD del Petróleo .....   | 72 |
| <b>Gráfico 38.</b> RD de Airbus .....  | 72 |
| <b>Gráfico 39.</b> RD del IBEX 35 .....  | 72 |
| <b>Gráfico 40.</b> RD de Boeing .....  | 73 |
| <b>Gráfico 41.</b> RD del FTSE 100 .....   | 73 |

## ÍNDICE DE TABLAS

|  |    |
|--|----|
| <b>Tabla 1.</b> Variables seleccionadas para la muestra .....                          | 26 |
| <b>Tabla 2.</b> Resultados mutual information antes del Covid-19 .....                 | 30 |
| <b>Tabla 3.</b> Resultados mutual information después del Covid-19 .....               | 31 |
| <b>Tabla 4.</b> Comparativa resultados mutual information .....                        | 32 |
| <b>Tabla 5.</b> Resultados Rational Quadratic GPR.....                                 | 37 |
| <b>Tabla 6.</b> Resultados Exponential GPR.....  | 40 |
| <b>Tabla 7.</b> Resultados Squared Exponential GPR .....                               | 43 |
| <b>Tabla 8.</b> Predicciones Exponential GPR.....                                      | 47 |
| <b>Tabla 9.</b> Predicciones Squared Exponential GPR.....                              | 48 |
| <b>Tabla 10.</b> Resultados Ensemble (accuracy: 76.7%).....                            | 52 |
| <b>Tabla 11.</b> Resultados Ensemble (accuracy: 77.6%).....                            | 53 |
| <b>Tabla 12.</b> Resultados Linear Discriminant .....                                  | 56 |
| <b>Tabla 13.</b> Resultados Quadratic Discriminant .....                               | 57 |
| <b>Tabla 14.</b> Predicciones Ensemble .....   | 59 |
| <b>Tabla 15.</b> Predicciones Quadratic Discriminant .....                             | 60 |
| <b>Tabla 17.</b> Serie temporal, model fitting .....                                   | 65 |
| <b>Tabla 18.</b> Predicciones ARIMA .....  | 67 |
| <b>Tabla 19.</b> Resultados Regression Discontinuity ‘Cierre’ y ‘Volumen’ de IAG ..... | 70 |
| <b>Tabla 20.</b> Resultados Regression Discontinuity variables más relevantes .....    | 74 |



# CAPÍTULO 1: INTRODUCCIÓN

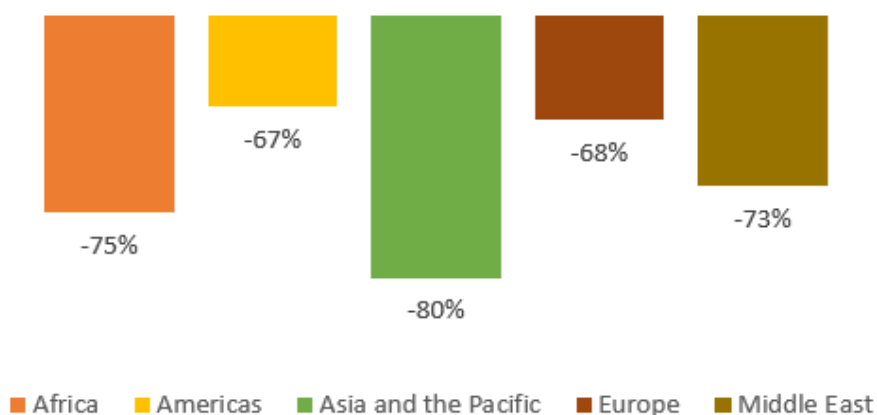
## 1.1.Contexto

### 1.1.1. Covid-19, turismo y líneas aéreas

La declaración de una pandemia mundial por Covid-19 en marzo de 2020 tuvo un impacto significativo en la economía mundial, especialmente en el sector turístico. Las medidas adoptadas por los gobiernos de todo el mundo para contener la propagación del virus, como las restricciones de movimiento y las políticas de distanciamiento social, tuvieron un **efecto drástico en este sector** [1].

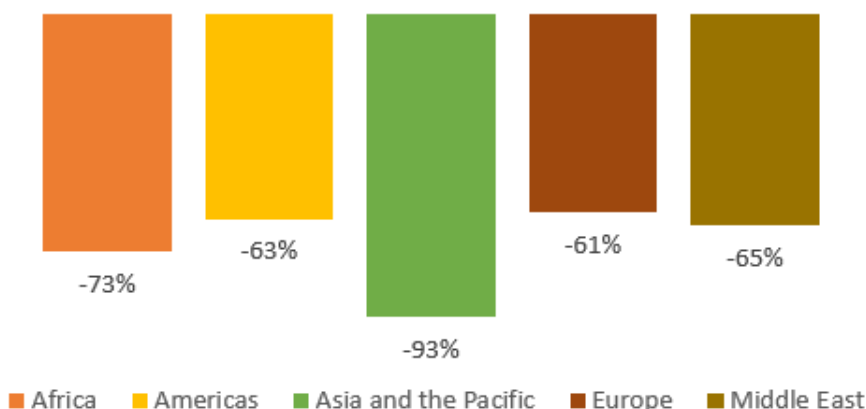
La pandemia causó estragos en el turismo puesto que las crisis de salud pública generaron temores y disminuyeron la demanda de viajes, lo que tuvo graves consecuencias económicas y sociales, especialmente en aquellos países que dependían y dependen en gran medida del turismo. En este contexto, es crucial comprender la evolución del turismo en respuesta a la pandemia en los siguientes cuatro gráficos que se muestran a continuación, en los que se puede observar la variación que experimentó el turismo en cada región con respecto al año previo:

*Gráfico 1. Variación por región de las llegadas de turistas internacionales en 2020 en comparación con 2019*



Fuente: Elaboración propia en base a los datos obtenidos de UNWTO [2].

**Gráfico 2.** Variación por región de las llegadas de turistas internacionales en 2021 comparación con 2020



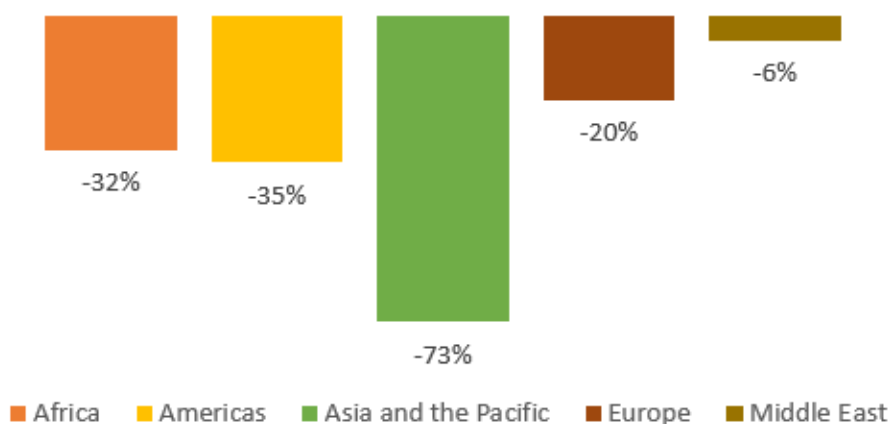
Fuente: Elaboración propia en base a los datos extraídos de UNWTO [2].

Estas gráficas revelan una marcada disminución en el turismo en todas las regiones durante los años 2020 y 2021, siendo así consistentes con el impacto global de la pandemia de Covid-19 en la industria del turismo.

En el año 2020, se observó una fuerte caída en todas las regiones, indicando un impacto significativo en la industria del turismo a nivel mundial, con restricciones de viaje, cierres de fronteras y preocupaciones de seguridad como principales factores que contribuyeron a esta disminución.

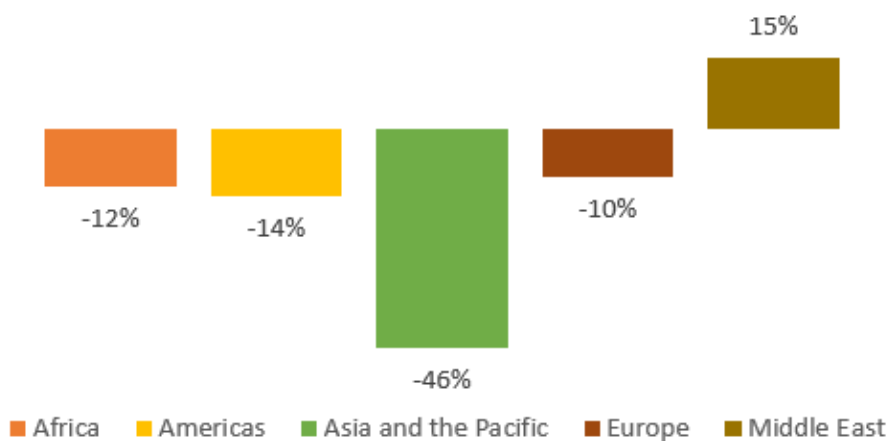
A medida que avanzó el año 2021, se observó una ligera mejora en algunos destinos turísticos, pero los porcentajes de cambio aún mostraban una disminución significativa en comparación con los años anteriores. Los porcentajes de cambio negativos se mantuvieron en el rango del -61% al -93%, constatando que la recuperación del turismo fue lenta y enfrentó desafíos adicionales como variantes del virus y nuevas olas de contagios.

**Gráfico 3.** Variación por región de las llegadas de turistas internacionales en 2022 comparación con 2021



Fuente: Elaboración propia en base a los datos extraídos de UNWTO [2].

**Gráfico 4.** Variación por región de las llegadas de turistas internacionales en 2023 comparación con 2022



Fuente: Elaboración propia en base a los datos extraídos de UNWTO [2].

En el año 2022, se observó una disminución menos pronunciada en algunos destinos turísticos, reflejando la recuperación parcial que hubo en algunas regiones, sin obviar que aún se encuentran por debajo de los niveles pre-pandemia.

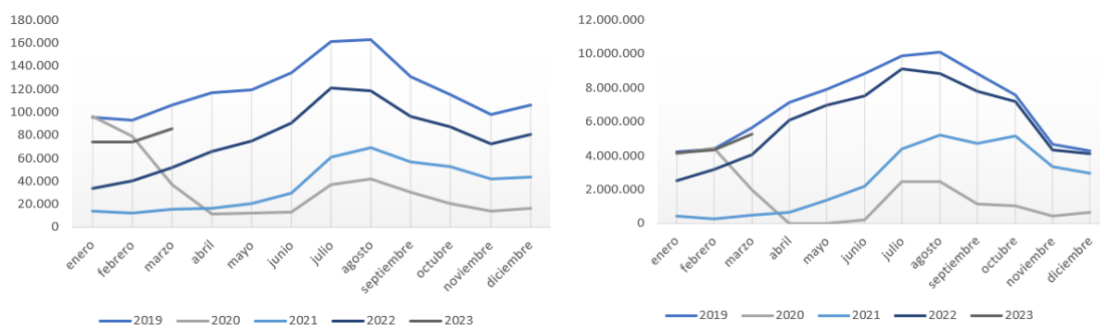
Por otro lado, el 2023 evidenció una tendencia positiva en la mayoría de las regiones, con porcentajes de cambio que fueron desde -46% hasta 15%. Aunque algunos destinos experimentaron una disminución menos significativa e incluso un crecimiento positivo, es importante destacar que la industria del turismo todavía se encuentra en proceso de recuperación y enfrenta desafíos como la confianza del turista y la adaptación a las nuevas normas y regulaciones.

Además, cabe mencionar que Asia experimentó un porcentaje de cambio más bajo en el año 2022 y 2023 debido a las altas restricciones que se mantuvieron en muchos países de la región. Estas restricciones, que incluyeron cierres de fronteras, cuarentenas obligatorias y medidas de distanciamiento social, tuvieron un impacto significativo en la actividad turística. Si bien estas medidas pudieron haber tenido éxito en contener la propagación del virus, también limitaron considerablemente la llegada de turistas extranjeros y la movilidad interna.

España fue uno de los primeros países en enfrentar altos niveles de infecciones por Covid-19. Es importante destacar que el **sector turístico es uno de los principales motores de la economía española**, contribuyendo aproximadamente al **12% de su PIB**. Durante la pandemia, esta industria experimentó una reducción significativa en su actividad, llegando a niveles mínimos [3]. Además, la Organización Mundial del Turismo afirmó que en 2021 España ocupaba la tercera posición en el ranking de los países más visitados en el mundo, después de México, con 31.7 millones de visitas. Esta es una cifra alta, si bien dista lejos de la cifra de 83 millones de turistas anuales que logró en 2019, posicionándose en segundo lugar. Además, en 2020 España fue el segundo destino con mayores ingresos, sólo por detrás de Estados Unidos, quedándose en tercer puesto en 2021 [4].

Por otro lado, las **llegadas internacionales se desplomaron un 77% en 2020 y un 63% en 2021** teniendo un impacto considerable en la facturación del sector turístico en España, lo que se reflejó en datos de empleo mínimos y una tasa de desempleo más alta [5]. El siguiente gráfico muestra la disminución significativa en las llegadas internacionales desde el año 2019 hasta el 2023.

**Gráfico 5.** Comparativa de llegadas de turistas internacionales: globalmente (representado en miles) vs. España



Fuente: Elaboración propia en base a los datos extraídos de UNWTO [2].

No se puede obviar que la pandemia del coronavirus ha sido un problema grave desde el año 2020. El 30 de enero de 2021, la OMS clasificó al Covid-19 como una “Emergencia de Salud Pública de Importancia Internacional” [6]. Las políticas de restricción de viajes implementadas por los gobiernos de todo el mundo tuvieron un impacto no sólo en la industria del turismo, sino especialmente en las compañías aéreas.

La demanda de vuelos experimentó una fuerte caída y la cadena de suministro se vio completamente interrumpida. En concreto, la pandemia tuvo un gran impacto en la industria de las aerolíneas comerciales, sin haberse recuperado las cifras pre-pandemia en 2023. Las restricciones de viaje, las órdenes gubernamentales de permanecer en casa, las cuarentenas, el distanciamiento social y el temor provocaron **en el año 2020 una disminución del 40% en el tráfico de pasajeros a nivel mundial** y una reducción del 47% al 58% en los asientos programados por las aerolíneas, según estimaciones de organismos internacionales de aviación. Además, las nuevas medidas de limpieza y seguridad generaron una mayor carga para las compañías aéreas, afectando también al horario de los vuelos [6]. Así, el volumen de pasajeros disminuyó un 60% generando un tráfico de 1800 millones de personas en 2020, frente a los 4500 millones que se produjeron en 2019 [7].

Según los informes de la OACI, se estima que la pandemia ocasionó pérdidas financieras significativas para las líneas aéreas, alcanzando los 370.000 millones de dólares. Además, los aeropuertos y los ANSP experimentaron pérdidas adicionales de aproximadamente 115.000 millones y 13.000 millones, respectivamente [8].

En cuanto al mercado de valores, ninguna crisis sanitaria anterior había afectado tan drásticamente como la pandemia de Covid-19. El sector turístico, en particular, experimentó una fuerte caída en la valoración de las empresas relacionadas en el mercado de valores como resultado de esta pandemia [9]. Existe una clara correlación negativa entre la pandemia y el mercado de valores del sector turístico.

El mercado de valores global sufrió importantes impactos desde el comienzo de la pandemia. En los primeros cinco meses de 2020, se observó una importante disminución en los principales índices bursátiles. El índice S&P 500 experimentó una caída del 34%

desde su nivel más alto hasta su nivel más bajo. Además, las bolsas de Brasil, Hong Kong, Italia y Japón también experimentaron importantes descensos en sus valores, con disminuciones del 46%, 25%, 41% y 31%, respectivamente [10]. La compañía seleccionada para esta investigación, IAG, también experimentó una fuerte caída en el precio de sus acciones que se analizará con más detalle a lo largo del trabajo. Desde entonces, la pandemia de Covid-19 ha representado la crisis más significativa a la que se ha enfrentado la industria de las aerolíneas desde la Segunda Guerra Mundial [11].

No obstante, la industria de las aerolíneas no ha estado exenta de este tipo de desafíos. En el pasado, ha enfrentado eventos similares, como el brote de SARS en 2007, los ataques de 2001 y la crisis financiera de 2008. En relación a epidemias anteriores, el brote de SARS en 2003 presentó similitudes relevantes. Durante ese periodo, la ALTA reportó una reducción del 35% en el número de pasajeros en comparación con el periodo anterior al brote. No obstante, es importante destacar que el impacto del Covid-19 fue notablemente más pronunciado en comparación con el impacto del SARS. En marzo de 2020, como resultado de las medidas gubernamentales y la interrupción de la oferta y la demanda, se experimentó una reducción significativa del 98% en el número de pasajeros [12]. En este sentido, la enfermedad del COVID-19, una enfermedad de naturaleza desconocida tuvo un impacto significativo en las actividades económicas a escala global, propagándose a más de 200 naciones en un lapso de meses [10].

### *1.1.2. La compañía IAG*

IAG es una destacada organización que engloba diversas aerolíneas a nivel global. Con una flota de 558 aeronaves y una extensa red de 256 destinos, IAG transporta aproximadamente 94 millones de pasajeros anualmente. La compañía tiene su sede en España y sus acciones se cotizan tanto en la bolsa de Londres como en los mercados bursátiles españoles. IAG, como empresa matriz, colabora y trabaja en estrecha coordinación con sus compañías operativas y gracias a su independencia de estas, ha podido desarrollar una estrategia a largo plazo que refleja los objetivos y valores del grupo, estableciendo metas de rendimiento para las compañías operativas, monitoreando su progreso y asignando eficientemente los recursos dentro del mismo [13].

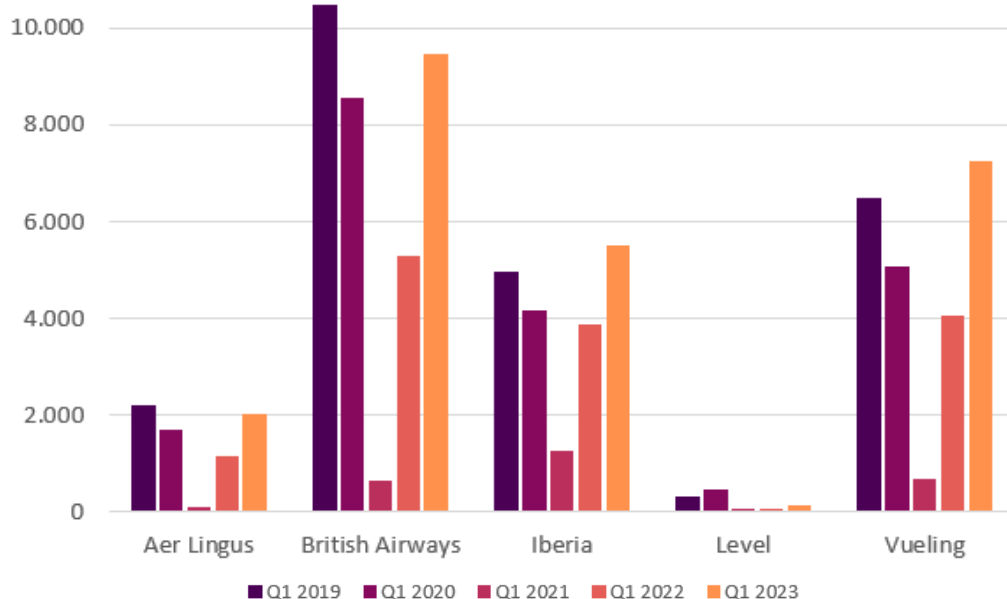
El modelo de negocio de IAG ha sido diseñado estratégicamente para maximizar las oportunidades y generar un valor óptimo. La visión de la compañía es convertirse en el

grupo líder a nivel mundial en el sector de las aerolíneas, enfocándose en la creación de valor sostenible tanto para los accionistas como para los clientes. Este modelo de negocio permite a IAG posicionarse de manera competitiva y adaptarse a los cambios constantes del entorno. En lugar de buscar soluciones universales y gracias a la estructura del grupo para aprovechar la diversidad de las aerolíneas y empresas asociadas que forman parte del mismo, estas entidades son capaces de ofrecer una propuesta única a sus clientes, cubriendo una amplia gama de ocasiones de viaje. En este sentido, se permite que cada compañía operadora alcance sus propios objetivos de rendimiento y mantenga su identidad distintiva dentro del grupo [13].

Dentro de la cartera de marcas y negocios de IAG, **se encuentran diversas compañías** que abarcan desde aerolíneas de servicio completo, como Iberia, British Airways y Aer Lingus, hasta aerolíneas de bajo coste tanto para rutas de corto y medio alcance, como Iberia Express y Vueling, como para rutas de largo alcance, como Level.

La sucesiva regulación de restricciones de movilidad debido a la pandemia de Covid-19 en los mercados clave en los que operan las compañías de IAG tuvo un impacto significativo en su valor en el mercado de valores. Esto se debió a la disminución o cancelación parcial o total de los flujos futuros de capital, ya que no fue posible operar vuelos o los vuelos no resultaron rentables. En el gráfico siguiente se evidencia la disminución del número de pasajeros por cada aerolínea del grupo en el primer trimestre de 2023 en comparación con el primer trimestre del año 2019. Se puede observar que todavía no se ha alcanzado una plena recuperación en todas las aerolíneas con respecto a los niveles previos a la pandemia, como se ha constatado anteriormente.

**Gráfico 6.** Análisis de pasajeros transportados por aerolínea de IAG en los primeros trimestres de cada año



Fuente: Elaboración propia en base a los datos extraídos de IAG [14].

## 1.2. Antecedentes

En el ámbito de los antecedentes, se destacan varios estudios. El primero, “COVID-19 and tourism sector stock price in Spain: medium-term relationship through dynamic regression models” [15], se enfocó en el impacto significativo que la pandemia global, la enfermedad del Covid-19, tuvo en el sector turístico, especialmente en España, uno de los primeros países en ser afectado por la pandemia y uno de los destinos turísticos más importantes del mundo. La evolución de la pandemia tuvo un impacto significativo en los valores de las empresas turísticas en el mercado de valores. Por tanto, cuantificar esta relación se volvió crucial para predecir el efecto de la pandemia en las acciones del sector turístico y mejorar la respuesta de los responsables de la toma de decisiones y los inversores ante la crisis.

En este estudio, se desarrolló un modelo de regresión dinámica para predecir el comportamiento de las acciones en el sector turístico español en función de la evolución de la pandemia de Covid-19 a medio plazo. A través del análisis realizado, se obtuvo la confirmación de que tanto el número de muertes como los casos relacionados eran factores predictivos relevantes para los precios de las acciones en el sector turístico.



Es importante destacar que este estudio se centró específicamente en el sector turístico en su conjunto. En contraste, el presente trabajo se enfoca exclusivamente en el análisis de la cotización de IAG y en cómo la pandemia ha afectado cuantitativamente su cotización. Además, se ha adoptado un enfoque más amplio, abarcando un horizonte temporal más extenso, no siendo de 14 meses como en el anterior estudio sino de 36. Esto ha permitido analizar el impacto de la pandemia en un periodo de tiempo más prolongado y evaluar qué variables han influido en la cotización de IAG.

Además, se emplea la innovadora técnica de RD, que permite identificar el efecto causal de eventos abruptos como la pandemia en la cotización de IAG. Este enfoque metodológico avanzado aporta un valor adicional al estudio, al proporcionar una herramienta más precisa para analizar el impacto de la pandemia en la cotización de IAG y comprender mejor su relación con las variables pertinentes.

El segundo antecedente que se consideró fue un estudio titulado “Covid-19's Impact On The Airline Industry In The US Stock Market Based on the Regression Model” [10]. En este estudio, se evaluó el efecto de Covid-19 en los precios de las acciones de las 10 principales compañías aéreas estadounidenses y en el Índice de Aerolíneas de NYSE Arca utilizando la metodología de estudio de eventos. Los resultados de este estudio indicaron que el Covid-19 tuvo un impacto significativo en la industria de las aerolíneas. Además, se examinó el impacto de las ayudas gubernamentales y la vacunación masiva, y se observó que los precios de las acciones de las aerolíneas se recuperaron lentamente, pero de manera constante. A diferencia de este trabajo, que se centra en una aerolínea específica, el estudio mencionado abarcó el análisis de varias aerolíneas. Sin embargo, este trabajo no se limita únicamente a la afirmación de que el Covid-19 tuvo un impacto en las compañías aéreas, sino que se busca cuantificar ese impacto utilizando técnicas avanzadas y también considerando modelos que ayudan a predecir la futura cotización de la acción. Esto permite obtener una visión más completa y precisa de los efectos de la pandemia en el mercado de valores.

Otro antecedente relevante fue el estudio titulado “Quantifying the impact of the COVID-19 pandemic on US airline stock prices” [16]. En este trabajo, se utilizó una frecuencia diaria de datos y el método de proyecciones locales para cuantificar las respuestas dinámicas de los precios de las acciones de aerolíneas estadounidenses ante el Covid-19.

Los resultados mostraban que los precios de las acciones de las aerolíneas experimentaron una caída inmediata de 0.1 puntos porcentuales en respuesta a un estímulo del 1% de Covid-19. Además, el efecto de dicho estímulo persiste más allá del día en que ocurre, con la mayoría de los precios de las acciones de las aerolíneas cayendo hasta 0.6 puntos porcentuales después de quince días.

A diferencia del segundo antecedente mencionado, que se centra en el impacto del Covid-19 en la industria de las aerolíneas en el mercado de valores de los Estados Unidos, este estudio busca cuantificar el impacto a través de técnicas avanzadas y se enfoca en las respuestas dinámicas de los precios de las acciones. Además, se destaca que la disminución de los precios de las acciones no se explica por un aumento en el coste variable de las aerolíneas debido a la pandemia, sino por una disminución en los viajes aéreos que afecta los ingresos, la rentabilidad y los precios de las acciones de las aerolíneas estadounidenses. Al incorporar este antecedente, se ha enriquecido la comprensión del impacto cuantitativo de la pandemia en los precios de las acciones de las aerolíneas, complementando el enfoque de nuestro estudio que se centra en una aerolínea específica en un contexto más amplio. Esto ha permitido obtener una visión más completa y precisa de los efectos de la pandemia en el mercado de valores y el sector de las aerolíneas.

El último antecedente, titulado “Impact of COVID-19 on different business models of European Airlines” [17], se centró en examinar la importancia del modelo de negocio de las aerolíneas europeas (bajo costo y servicio completo) para el rendimiento en el mercado de valores durante la pandemia. En contraste con el enfoque de este trabajo, que se centra en el impacto cuantitativo de la pandemia en la cotización de una aerolínea específica, se adoptó un enfoque más financiero y cualitativo al examinar la importancia del modelo de negocio de las aerolíneas europeas durante la pandemia.

Mediante la utilización de la metodología de estudio de eventos y una muestra de 11 aerolíneas, se encontró que, durante las fases negativas de la pandemia, se observaban desviaciones significativas de los rendimientos esperados. Además, se constató que las aerolíneas de servicio completo tuvieron un mejor desempeño en comparación con las aerolíneas de bajo costo cuando los mercados de valores europeos experimentaron el mayor impacto en febrero de 2020. En este sentido, a diferencia del presente estudio, que

se enfoca específicamente en el impacto cuantitativo de la pandemia en la cotización de una aerolínea en particular y utiliza técnicas avanzadas como la RD, el antecedente se centra en el análisis de modelos de negocio y los indicadores financieros durante la crisis. Ambos estudios abordan el impacto del Covid-19 en el sector de las aerolíneas, pero se distinguen por sus enfoques: el presente estudio adopta un enfoque más cuantitativo y predictivo, brindando una perspectiva única sobre el impacto de la pandemia en la cotización de la aerolínea analizada.

### **1.3. Objetivos**

El presente trabajo tiene como objetivo principal analizar y comprender el impacto de la pandemia de Covid-19 en la cotización de la compañía IAG, así como explorar el comportamiento de variables relevantes antes y después de este evento sin precedentes. Para lograrlo, se plantean los siguientes objetivos específicos:

1. Identificar las variables más representativas que influyen en la cotización de IAG y determinar si su relación varía en función del periodo anterior y posterior a la pandemia. Este análisis permitirá comprender cómo la crisis sanitaria ha afectado las dinámicas del mercado y si existen cambios significativos en las variables que influyen en la valoración de la compañía.
2. Evaluar y comparar diferentes modelos de *Machine Learning* para predecir la cotización y las velas de IAG. Se pretende determinar qué modelo ofrece mejores resultados y si la pandemia ha influido en la precisión de las predicciones. Esto proporcionará información acerca de la capacidad de los modelos para adaptarse a un contexto de volatilidad y cambio como el generado por la pandemia.
3. Cuantificar el impacto real de la pandemia en la cotización de IAG utilizando la técnica de *regression discontinuity*. Se pretende determinar de manera precisa y rigurosa el efecto específico de la pandemia en la valoración de la compañía, aislando otros factores externos y controlando variables relevantes.

#### 1.4. Metodología

La metodología adoptada en este trabajo se ha basado en varios pilares fundamentales. En primer lugar, se ha realizado una exhaustiva revisión de estudios e informes relevantes, los cuales han sido extraídos principalmente de ‘Google Scholar’ y ‘Web of Science’. El acceso a esta documentación ha sido de gran utilidad para obtener una base sólida de conocimientos y comprender los fundamentos teóricos en los que se sustenta este trabajo. Asimismo, ha permitido comprender las técnicas punteras que han sido empleadas en el análisis de datos, lo que ha contribuido a la adopción de un enfoque metodológico riguroso y actualizado.

En segundo lugar, se ha contado con la orientación y asesoramiento de mi tutor, Luis Ángel Calvo Pascual, quien ha proporcionado valiosas prácticas y directrices específicas para el desarrollo de este trabajo. Su experiencia y conocimiento en el campo han sido de gran ayuda para garantizar la rigurosidad y precisión en la metodología empleada.

Por último, se ha empleado una combinación de programas y lenguajes de programación, como Excel, Python y MATLAB, para llevar a cabo la visualización, manipulación y análisis de los datos recopilados. La mayor parte del código se ha desarrollado utilizando la plataforma de Google Colaboratory, que permite ejecutar y compartir código en lenguaje Python. En particular, se utilizó la biblioteca *yfinance* en Python para la obtención de datos financieros necesarios para el análisis de IAG, lo que incluyó precios de cotización, volúmenes de transacción y otros indicadores clave. Además, se utilizó este programa para emplear la medida de *mutual information* y realizar el análisis de *regression discontinuity*.

En el contexto de la optimización bayesiana y las predicciones, se empleó Matlab, una herramienta ampliamente utilizada en análisis numérico y modelado. Se realizaron análisis de datos y se aplicaron técnicas de optimización bayesiana para ajustar y validar modelos predictivos, lo que permitió obtener pronósticos cuantitativos sobre la cotización de IAG en función de las variables consideradas.

## 1.5. Estructura del TFG

El presente trabajo se estructura en varios capítulos que abordan diferentes aspectos relacionados con el análisis de la cotización de la compañía IAG en el contexto de la pandemia del Covid-19. Es importante destacar que cada capítulo trata de abordar y responder a uno de los objetivos establecidos en la sección anterior. En el presente capítulo, se brinda el contexto necesario para comprender la importancia del estudio realizado en el trabajo. Se abordan dos aspectos fundamentales: la relación entre Covid-19, el turismo y las líneas aéreas, y se introduce la compañía IAG además de señalar los objetivos.

En el Capítulo 2, se ha llevado a cabo un estudio de selección de características utilizando la medida de *mutual information*, en contraste con los métodos clásicos. Esta selección de características se ha realizado tanto antes como después de la pandemia, con el objetivo de identificar las variables que influenciaron de manera significativa la cotización de IAG en cada período.

El Capítulo 3 se centra en el desarrollo de modelos de aprendizaje automático de regresión. Los datos necesarios para este análisis fueron extraídos de *yahoo finance* y se empleó la técnica de optimización bayesiana para realizar predicciones utilizando el entorno de programación MATLAB. El Capítulo 4 es muy similar con la salvedad de que se realizó una transformación de todas las variables a categorías con el fin de abordar un enfoque clasificatorio, centrándose en la predicción de las velas de IAG.

En el Capítulo 5, se exploró la técnica de *regression discontinuity* para evaluar el impacto real de la pandemia de Covid-19 en la cotización de IAG.

En el último capítulo de este trabajo se han presentado las consideraciones finales, donde se discuten las conclusiones derivadas de los análisis realizados en los capítulos anteriores y se ofrecen recomendaciones para futuras investigaciones en esta área.

## CAPÍTULO 2: FEATURE SELECTION

Los modelos de aprendizaje automático demuestran un alto rendimiento cuando se entrenan con conjuntos de datos de entrenamiento apropiados. Al utilizar ejemplos de conjuntos de datos de *scikit-learn*, se garantiza que los datos de muestra están correctamente tratados y son efectivos para la aplicación de modelos de aprendizaje automático.

Sin embargo, en la práctica, es posible encontrar conjuntos de datos de baja calidad que no sean adecuados para ningún modelo. Por lo tanto, al utilizar modelos de aprendizaje automático para predecir el precio de las acciones de IAG, es crucial realizar un paso previo en el que se seleccionen las variables relevantes de los datos y se descarten aquellas que no aporten información significativa sobre la variable objetivo.

Por esta razón, no es de extrañar que numerosos profesionales especializados en aprendizaje automático reconozcan que la selección apropiada de características es uno de los aspectos fundamentales, que requiere tiempo para lograr una predicción efectiva de la variable objetivo. Es esencial invertir tiempo en analizar y comprender cómo abordar este proceso de manera óptima.

### 2.1. Mutual information

Esta medida, que se encuentra dentro del ámbito de la teoría de la probabilidad y la teoría de la información, evalúa la relación de dependencia existente entre las variables que se quieran analizar. Específicamente, cuantifica la “cantidad de información” adquirida sobre una variable aleatoria al observar la otra variable aleatoria. Esta medida se expresa en unidades como *shannons (bits)*, *nats* o *hartleys*. El concepto de MI se encuentra estrechamente relacionado con la entropía de una variable aleatoria, un concepto fundamental en la teoría de la información que evalúa la “cantidad de información” esperada contenida en una variable aleatoria [18].

Además, esta medida va más allá de las variables aleatorias de valor real y dependencia lineal, por las cuales se ve limitado el coeficiente de correlación. MI es una medida más general que determina la diferencia entre la distribución conjunta de un par de variables

y el producto de las distribuciones marginales de esas variables. MI se calcula como el valor esperado de la información mutua puntual.

MI se determina entre dos variables y cuantifica la disminución de la incertidumbre en una dado un valor conocido de la otra. Esta medida evalúa las reducciones en la entropía bajo la condición del valor objetivo. Andrew Zhu [19] considera que la forma más sencilla de explicar este concepto es la siguiente:

$$MI(\text{feature}; \text{target}) = \text{Entropy}(\text{feature}) - \text{Entropy}(\text{feature}|\text{target})$$

A medida que el valor aumenta, la relación entre esta característica y el objetivo se vuelve más estrecha, lo que indica que deberíamos incluir esta característica en el conjunto de datos de entrenamiento. Por otro lado, si la puntuación de MI es 0 o muy baja, indica una conexión débil entre esta característica y el objetivo. La forma más precisa de calcular esta medida es a partir de la siguiente ecuación:

$$I(X; Y) = H(X) - H(X|Y)$$

La primera parte de la ecuación representa la MI. En la segunda parte,  $H(X)$  representa la entropía de la variable  $X$ , y  $H(X|Y)$  es la entropía condicional de  $X$  dado  $Y$ . Además, es importante recalcar que es una medida de dependencia o “dependencia mutua” entre dos variables aleatorias. Como tal, la medida es simétrica, lo que significa que  $I(X; Y) = I(Y; X)$ .

Por otro lado, es conveniente mencionar la divergencia de Kullback-Leibler, también conocida como KL, que es una medida que cuantifica la discrepancia o diferencia entre dos distribuciones de probabilidad. MI también puede ser expresada como la divergencia de KL entre la distribución de probabilidad conjunta de dos variables y el producto de las probabilidades marginales de cada variable [20]. Esto se plantea formalmente de la siguiente manera:

$$I(X; Y) = KL(p(X, Y) || p(X) * p(Y))$$

Como bien señaló Christopher [21], en casos donde las variables no sean independientes, es posible obtener una estimación sobre su nivel de independencia al examinar la

divergencia de KL entre la distribución conjunta de las variables y el producto de sus distribuciones marginales, cálculo que se conoce, como bien se ha descrito previamente, como MI.

## 2.2. Sobre la disparidad entre los tests clásicos y *mutual information*

Antes de realizar el análisis correspondiente para determinar las variables que serán incluidas en las predicciones, es importante justificar el uso de MI en lugar de los *tests* clásicos como el *f-test* o el *chi cuadrado*.

Y es que ambas medidas forman parte del concepto general y, sin embargo, existe una diferencia determinante que lleva a MI a ser el mejor método para seleccionar variables. Para poder explicarlo, se utilizará el ejemplo descrito en *scikit-learn* [22]. En el mismo, se parte de una variable objetivo que depende de tres características,  $x_1$ ,  $x_2$ ,  $x_3$  distribuidas uniformemente sobre  $[0, 1]$ , de la manera siguiente:

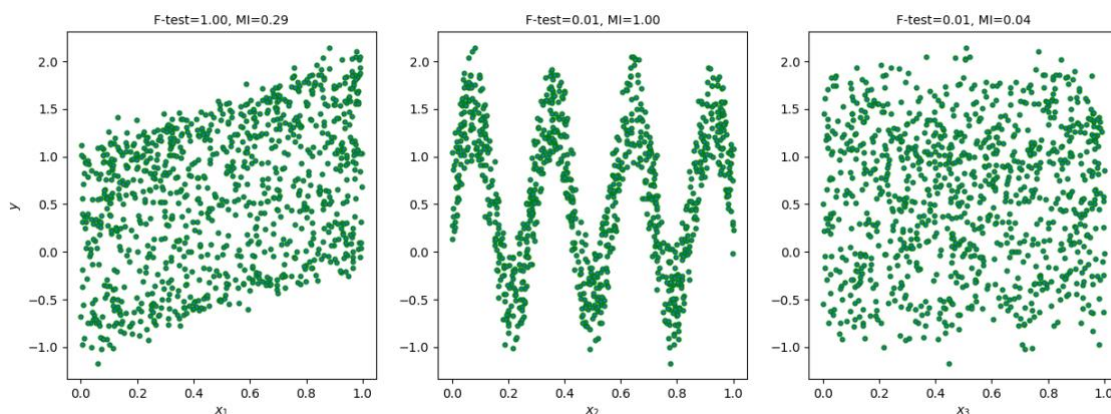
Ambas medidas forman parte del concepto general de *feature selection*. Sin embargo, existe una diferencia significativa que posiciona a la medida de MI como el mejor método a la hora de seleccionar variables. Para ilustrar esta diferencia, se ha utilizado el ejemplo descrito en *scikit-learn*. En este ejemplo, se considera una variable objetivo que depende de tres características, ' $x_1$ ', ' $x_2$ ' y ' $x_3$ ', distribuidas uniformemente en el intervalo  $[0, 1]$ .

$$y = x_1 + \sin(7 * \pi * x_2) + 0.1 * N(0,1)$$

A partir de lo expuesto anteriormente, se puede inferir que la tercera característica es completamente irrelevante en relación a la variable objetivo. Además, al trazar la dependencia de la variable objetivo, ' $y$ ', respecto de las diferentes características y los valores normalizados de los estadísticos univariantes como los *f-tests* y MI, se obtiene el siguiente resultado:



**Gráfico 7.** Dependencia de la variable objetivo respecto a tres características a través de los estadísticos *F-test* y *MI*



Fuente: Elaboración propia en base a los datos extraídos de *scikit-learn*

Como se puede observar en los tres gráficos representados, mientras que la *f-test* solo es capaz de capturar la dependencia lineal, clasificando a 'x<sub>1</sub>' como la característica más discriminante, **MI tiene la capacidad de capturar cualquier tipo de dependencia entre variables** y clasifica a 'x<sub>2</sub>' como la característica más discriminante, siendo lo más concordante en este ejemplo. Por otro lado, ambos métodos coinciden en que 'x<sub>3</sub>' es una característica irrelevante.

### 2.3. Selección de la muestra

Tras haber expuesto el contexto que prevaleció en la industria de las aerolíneas durante la pandemia del Covid-19, con un enfoque específico en el impacto sufrido por la cotización de IAG, conviene adentrarse en el análisis de la misma. No obstante, antes de examinar los procedimientos implementados y los resultados obtenidos, resulta crucial proporcionar una descripción detallada de la muestra seleccionada como punto de partida.

Es por ello que, con el objetivo de llevar a cabo un análisis cuantitativo sobre la evolución de la cotización de IAG, se han recopilado los datos de esta compañía a través de la plataforma de *yahoo finance* mediante la utilización de la herramienta de programación Python. Estos datos servirán como base para realizar los diversos análisis propuestos en este trabajo.

Como se ha explicado anteriormente, MI es una medida que permite determinar qué variables han influido de manera significativa en la variable objetivo deseada, en este caso, el precio de cierre de las acciones de IAG. Por lo tanto, una vez que obtenidos los

datos correspondientes a la variable objetivo, se ha llevado a cabo una investigación exhaustiva para identificar las variables relevantes en la predicción de la cotización de IAG. Las variables seleccionadas para que formaran parte de la muestra han sido las siguientes:

*Tabla 1. Variables seleccionadas para la muestra*

| <b>Ticker de Yahoo Finance</b> | <b>Descripción de la variable</b>            |
|--------------------------------|--|
| <b>CL=F</b>                    | Petróleo                                     |
| <b>^FTSE</b>                   | Índice del mercado de valores de Reino Unido |
| <b>^IBEX</b>                   | Índice del mercado de valores de España      |
| <b>LHA.DE</b>                  | Lufthansa                                    |
| <b>AF.PA</b>                   | Air France                                   |
| <b>DAL</b>                     | Delta Air Lines                              |
| <b>UAL</b>                     | United Airlines                              |
| <b>^VIX</b>                    | Indicador de volatilidad                     |
| <b>XAL</b>                     | Índice de compañías aéreas de Estados Unidos |
| <b>BA</b>                      | Boeing                                       |
| <b>AIR.PA</b>                  | Airbus                                       |

Fuente: Elaboración propia en base a los datos extraídos de *yahoo finance*

Es relevante mencionar que se ha tomado en consideración tanto el precio de cierre (“Close”) como el volumen (“Volume”) de cada una de estas variables. La inclusión de estas dos medidas es fundamental debido a su relevancia en el análisis financiero.

El precio de cierre es una variable clave que representa el último precio al que se ha negociado un activo financiero en un período de tiempo determinado. Proporciona información valiosa sobre la tendencia y la dirección del mercado, ya que refleja el consenso general de los participantes del mercado en cuanto al valor del activo en un momento dado. El análisis del precio de cierre permite identificar patrones y tendencias que pueden ayudar en la toma de decisiones de inversión.

El volumen, por otro lado, indica la cantidad de acciones o contratos negociados durante un período de tiempo determinado. Es una medida de la actividad y la liquidez del

mercado. El análisis del volumen puede revelar información sobre la intensidad de las transacciones y la participación de los inversores en un determinado activo. Cambios significativos en el volumen pueden indicar cambios en el interés del mercado y proporcionar señales de posibles movimientos de precios.

Al incluir tanto el precio de cierre como el volumen en el análisis, se busca capturar tanto la dirección del mercado como la actividad de negociación. Estas variables proporcionan información complementaria y en conjunto permiten obtener una visión más completa y precisa del comportamiento de los activos financieros.

Por otro lado, la elección de las variables descritas en la tabla se ha basado en diferentes razones. En primer lugar, se consideraron los precios del combustible, que han alcanzado niveles no vistos desde 2014. Este factor impacta de manera significativa en las aerolíneas que carecen de una cobertura adecuada. El coste del queroseno, el combustible utilizado por las aeronaves es uno de los principales gastos de las compañías aéreas. Su cotización experimentó un aumento del 10% en un solo día en marzo de 2022, superando los niveles registrados en 2013 y triplicándose en tan solo dos años [23].

Históricamente, el precio del queroseno para aviones ha sido un factor crucial en la rentabilidad de las aerolíneas, ya que constituye la principal o segunda fuente de costos para estas compañías. Durante la última década, el costo del queroseno ha representado entre el 15% y el 35% de sus gastos operativos. Esta dependencia de las fluctuaciones en el precio del combustible ha sido una consideración clave para evaluar el rendimiento económico de las aerolíneas en general [23].

Además, como bien se ha adelantado en la descripción de IAG, se ha tenido en cuenta que las acciones de esta compañía cotizan en la bolsa de Londres y en los mercados bursátiles españoles. Por esta razón, se ha considerado pertinente incluir los índices bursátiles relevantes en ambas ubicaciones geográficas. Los movimientos generales del mercado pueden ejercer una influencia significativa en la cotización de las acciones, especialmente en presencia de eventos económicos o financieros relevantes, como la pandemia global o la situación actual en Ucrania [24]. Por lo tanto, se ha tenido en cuenta la posible relación entre los índices bursátiles y la evolución del precio de las acciones de IAG.

Por otro lado, se ha considerado importante tener en cuenta variables relacionadas con la industria de la aviación, dado que las condiciones y tendencias de este sector pueden impactar en el precio de las acciones de IAG. La inclusión de otras aerolíneas relevantes se ha justificado por varios motivos. En primer lugar, la industria de la aviación es altamente competitiva y está sujeta a diversos factores macroeconómicos y geopolíticos que pueden afectar su desempeño. Las aerolíneas están expuestas a riesgos específicos de la industria, como la demanda de pasajeros, la competencia en precios, los costos operativos, la capacidad de gestión de flotas y las regulaciones gubernamentales [25]. La evolución de otras compañías aéreas relevantes puede proporcionar una perspectiva más amplia sobre la dinámica del sector y ayudar a identificar patrones o tendencias comunes que podrían influir en la cotización de las acciones de IAG. Asimismo, las operaciones de las aerolíneas están interconectadas en cierta medida, ya que las alianzas estratégicas, los acuerdos de código compartido y la colaboración en la gestión de rutas y vuelos son comunes en la industria. Los eventos o decisiones que afecten a otras aerolíneas importantes pueden tener repercusiones indirectas en IAG, lo que justifica considerar sus precios de cierre como variables relevantes [26].

Se puede observar también la inclusión del índice de volatilidad<sup>1</sup>. La volatilidad en los mercados financieros puede tener un impacto significativo en la cotización de las acciones de IAG y refleja la incertidumbre y el riesgo percibidos por los inversores. El índice de volatilidad, también conocido como “índice del miedo” o “índice del temor”, mide la expectativa del mercado en cuanto a la volatilidad futura. La inclusión de este índice como variable relevante proporciona información adicional sobre el entorno general del mercado y la confianza de los inversores. En momentos de alta volatilidad, los precios de las acciones tienden a fluctuar de manera más pronunciada, lo que puede influir en la cotización de IAG. Además, eventos económicos, políticos o de salud pública que generen incertidumbre o inestabilidad a nivel global pueden reflejarse en este índice y afectar las expectativas de los inversores sobre el desempeño de las acciones de IAG [27].

---

<sup>1</sup> El mencionado índice fue el primer referente en medir las expectativas del mercado en relación a la volatilidad. No obstante, es importante destacar que este índice tiene una naturaleza prospectiva, lo que implica que solo refleja la volatilidad implícita del S&P 500 en los 30 días siguientes [29].

La inclusión del índice de aerolíneas como variable relevante se fundamenta en su capacidad para ofrecer una perspectiva integral del desempeño general del sector aéreo en Estados Unidos proporcionando así información valiosa sobre las tendencias y dinámicas del sector en su conjunto. El índice de aerolíneas de Estados Unidos actúa como un indicador amplio que refleja la salud general de la industria y puede capturar los efectos colectivos de estos factores en las compañías aéreas.

Por último, la inclusión de las compañías Boeing y Airbus en el análisis se justifica por su relevancia en la industria de las aerolíneas. Ambas empresas son líderes en la fabricación de aviones comerciales y tienen un impacto significativo en el desempeño del sector [25].

Boeing es reconocida mundialmente por su experiencia en la fabricación de aviones comerciales y equipos de defensa. El desempeño de Boeing puede tener repercusiones directas en las aerolíneas, ya que muchas compañías dependen de sus aviones en sus flotas. Los eventos relacionados con Boeing, como nuevos lanzamientos de aviones, problemas de seguridad o retrasos en la entrega de pedidos, pueden afectar la confianza de los inversores y tener un impacto en la cotización de las acciones de las aerolíneas, incluida IAG. Lo mismo ocurre con Airbus, una multinacional europea que compite directamente con Boeing en el mercado de aviones comerciales y militares. Al igual que Boeing, el desempeño de Airbus también puede influir en el sector de las aerolíneas [28].

La inclusión de estas variables proporciona una visión más amplia y completa del entorno en el que opera IAG. El seguimiento del desempeño de Boeing y Airbus permite evaluar la salud y las perspectivas de la industria de las aerolíneas en general, y comprender de qué manera las acciones de estas empresas pueden influir en la cotización de IAG.

#### **2.4. Feature selection sobre IAG**

Una vez presentada la medida de MI y las variables consideradas en el análisis, se procedió a **dividir la muestra en dos conjuntos**: uno compuesto por los datos anteriores a la pandemia y otro por los datos posteriores<sup>2</sup>. Esta división permitió realizar la selección

---

<sup>2</sup> El código de la práctica está disponible en el siguiente enlace:  
[https://colab.research.google.com/drive/1oe0Qe\\_r-DuPRSbf3pLaM3rjd\\_pyWvqp1](https://colab.research.google.com/drive/1oe0Qe_r-DuPRSbf3pLaM3rjd_pyWvqp1)

de variables en ambos contextos, con el objetivo de observar posibles variaciones en los resultados. Dicha estrategia se implementó para evaluar la estabilidad y consistencia de las variables seleccionadas en diferentes períodos, brindando una perspectiva más completa y robusta sobre la influencia de las variables en el precio de cierre de las acciones de IAG.

#### 2.4.1. Antes del Covid-19

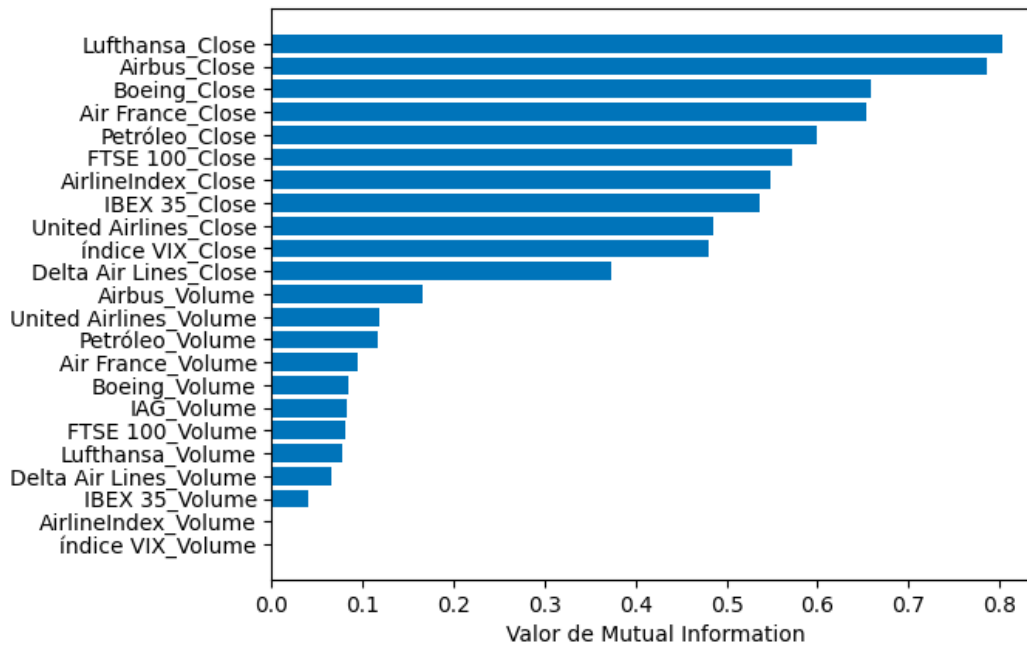
La muestra utilizada en la primera parte del análisis abarcó el período comprendido desde junio de 2017 hasta enero de 2020. Durante este intervalo de tiempo, se llevó a cabo MI obteniendo los siguientes resultados:

**Tabla 2.** Resultados mutual information antes del Covid-19

| Variables ordenadas por MI (de mayor a menor): |   |
|--|---|
| Lufthansa_Close: 0.8029464102585737            | United Airlines_Volume: 0.11878538164161068 |
| Airbus_Close: 0.7868697590453197               | Petróleo_Volume: 0.11582156215136363        |
| Boeing_Close: 0.6582725978834199               | Air France_Volume: 0.09522111842397374      |
| Air France_Close: 0.6535753445788131           | Boeing_Volume: 0.0837663766679353           |
| Petróleo_Close: 0.5989211134819805             | IAG_Volume: 0.08303080641018568             |
| FTSE 100_Close: 0.5722101550306888             | FTSE 100_Volume: 0.08133919368090492        |
| AirlineIndex_Close: 0.5482078190953863         | Lufthansa_Volume: 0.07716907539039086       |
| IBEX 35_Close: 0.5357671064981817              | Delta Air Lines_Volume: 0.06560098899310596 |
| United Airlines_Close: 0.48600430070399003     | IBEX 35_Volume: 0.03991547973675358         |
| índice VIX_Close: 0.4796820386435825           | AirlineIndex_Volume: 1.3322676295501878e-15 |
| Delta Air Lines_Close: 0.37243222664110043     | Índice VIX_Volume: 1.3322676295501878e-15   |
| Airbus_Volume: 0.16644251302627078             |   |

Fuente: Elaboración propia

**Gráfico 8. Resultados mutual information antes del Covid-19**



Fuente: Elaboración propia

#### 2.4.2. Después del Covid-19

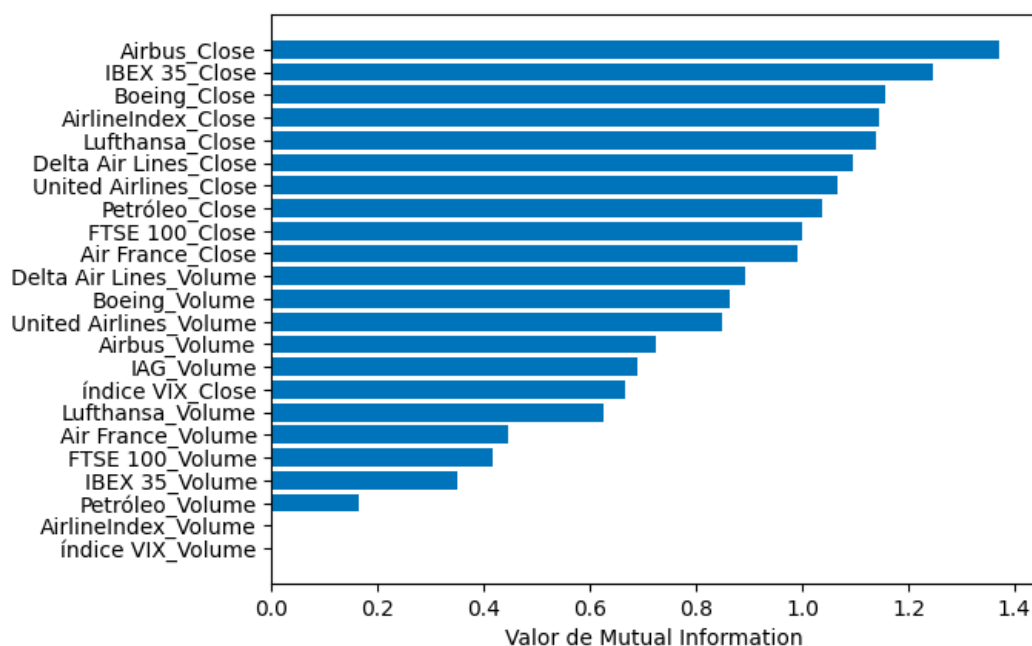
La muestra escogida para esta parte abarcó desde febrero de 2020 hasta mayo de 2023. Los resultados que se obtuvieron fueron los siguientes:

**Tabla 3. Resultados mutual information después del Covid-19**

| Variables ordenadas por MI (de mayor a menor): |   |
|--|---|
| Airbus_Close: 1.3709226108739803               | United Airlines_Volume: 0.848228517905522 |
| IBEX 35_Close: 1.2466648526774349              | Airbus_Volume: 0.7248582482240251         |
| Boeing_Close: 1.1564105125417172               | IAG_Volume: 0.6902135075917029            |
| AirlineIndex_Close: 1.1450409855072285         | índice VIX_Close: 0.6664673007803312      |
| Lufthansa_Close: 1.1386990897345552            | Lufthansa_Volume: 0.6257952228995558      |
| Delta Air Lines_Close: 1.0970404480654878      | Air France_Volume: 0.4469855908968241     |
| United Airlines_Close: 1.0678337164933573      | FTSE 100_Volume: 0.41679939345310446      |
| Petróleo_Close: 1.038591064665745              | IBEX 35_Volume: 0.3502224957273521        |
| FTSE 100_Close: 1.0001857874507394             | Petróleo_Volume: 0.16422253170428425      |
| Air France_Close: 0.9904841385925751           | AirlineIndex_Volume: 0.0                  |
| Delta Air Lines_Volume: 0.892664441486489      | índice VIX_Volume: 0.0                    |
| Boeing_Volume: 0.8627812312364584              |   |

Fuente: Elaboración propia

**Gráfico 9. Resultados mutual information después del Covid-19**



Fuente: Elaboración propia

## 2.5. Comparativa resultados

Después de mostrar los resultados obtenidos, se seleccionaron en ambos casos las catorce primeras variables de cada muestra. Esta elección se basó en un umbral de MI superior a 0.1 al analizar los datos previos a la pandemia. Al comparar los resultados obtenidos después de la pandemia, se observó un aumento en la MI, lo cual llevó a seleccionar el mismo número de variables que en el conjunto de datos anterior.

Para analizar los posibles cambios entre las variables más significativas, se presenta a continuación una tabla que muestra las ocho variables más representativas de cada conjunto:

**Tabla 4. Comparativa resultados mutual information**

| Antes del Covid-19                     | Después del Covid-19                      |
|--|---|
| Lufthansa_Close: 0.8029464102585737    | Airbus_Close: 1.3709226108739803          |
| Airbus_Close: 0.7868697590453197       | IBEX 35_Close: 1.2466648526774349         |
| Boeing_Close: 0.6582725978834199       | Boeing_Close: 1.1564105125417172          |
| Air France_Close: 0.6535753445788131   | AirlineIndex_Close: 1.1450409855072285    |
| Petróleo_Close: 0.5989211134819805     | Lufthansa_Close: 1.1386990897345552       |
| FTSE 100_Close: 0.5722101550306888     | Delta Air Lines_Close: 1.0970404480654878 |
| AirlineIndex_Close: 0.5482078190953863 | United Airlines_Close: 1.0678337164933573 |
| IBEX 35_Close: 0.5357671064981817      | Petróleo_Close: 1.038591064665745         |

Fuente: Elaboración propia



Los resultados obtenidos muestran **cambios significativos en relación a la pandemia**. Durante el período anterior, las aerolíneas tenían una influencia destacada, mientras que los índices de mercado no eran tan relevantes. Sin embargo, en el período posterior, se observa que el IBEX 35 se posicionaba como la segunda variable más relevante. La importancia del IBEX 35 en este contexto puede reflejar la confianza y las expectativas de los inversores en la recuperación económica y el crecimiento del sector de la aviación en España.

Es importante también destacar la aparición de nuevas variables en los datos de después de la pandemia, como “Delta Air Lines\_Close” y “United Airlines\_Close”, lo cual indica la importancia de tener en cuenta los efectos particulares de las aerolíneas individuales en el análisis de los precios de las acciones.

Por otro lado, resulta interesante observar que, en el período posterior, el petróleo, a pesar de presentar una MI mayor, descendió en su posición con respecto al período anterior a la pandemia. Este fenómeno puede generar cierta confusión, ya que los precios actuales del combustible son más elevados, lo que podría sugerir una influencia más significativa en la cotización de IAG. Sin embargo, es importante destacar que la tendencia alcista del queroseno se produjo después de que la pandemia provocara una caída histórica en el precio del petróleo. En este contexto, durante la presentación de los resultados trimestrales, Lufthansa reconoció que se ve “necesariamente afectado por los precios significativamente más altos de la energía” y destacó especialmente el aumento de la brecha de precios entre el petróleo y el queroseno [30].

A lo largo de los años, las aerolíneas han utilizado diferentes estrategias para cubrir sus gastos de combustible, como la compra de una cantidad determinada de queroseno a precios preestablecidos mediante instrumentos financieros como *swaps*, futuros y opciones de compra. Los datos más recientes indican que, en el caso de las aerolíneas europeas que divulgan sus operaciones de cobertura, tienen una proporción significativa de sus costos de combustible cubiertos a corto plazo. Este hecho puede explicar por qué el petróleo no tiene tanta relevancia en la cotización de las acciones de las aerolíneas en la actualidad [23].

## CAPÍTULO 3: MODELOS MACHINE LEARNING REGRESIVOS

En este capítulo, se presenta la metodología utilizada para desarrollar y aplicar modelos de aprendizaje automático con el objetivo de predecir el precio de cierre de las acciones de la compañía IAG. El análisis se divide en dos partes: uno previo a la pandemia de Covid-19 y otro posterior. Para llevar a cabo este estudio, se han empleado dos herramientas principales. Por un lado, Python como herramienta principal para la extracción y procesamiento de los datos. Por otro lado, MATLAB para la implementación de los modelos de aprendizaje automático mediante optimización bayesiana y para calcular las predicciones de la cotización de IAG. Los datos utilizados se han exportado en formato Excel, lo que ha permitido una manipulación y análisis más eficiente.

### 3.1. Preparación de los datos

En primer lugar, se ha procedido a la extracción de los datos correspondientes al período previo a la pandemia, que abarca desde junio de 2017 hasta enero de 2020. Estos datos representan un escenario estable y no están influenciados por los eventos relacionados con la pandemia. Asimismo, se han exportado los datos presentes y posteriores a la pandemia, desde febrero de 2020 hasta mayo de 2023. Esta división ha permitido evaluar la capacidad de los modelos para adaptarse a diferentes escenarios y capturar posibles cambios en la dinámica del mercado.

Las variables que componen cada una de las muestras, antes y después de la pandemia, han sido descritas en el capítulo anterior.

Es importante destacar que los últimos ocho registros de datos corresponden a las predicciones generadas mediante herramientas de modelado en MATLAB. Estos datos se han separado del conjunto de datos principal y se han utilizado posteriormente para evaluar la precisión de los modelos. De esta manera, se ha garantizado que los resultados obtenidos sean imparciales y no se hayan visto afectados por la información de predicción utilizada en el proceso de construcción de los modelos.

### 3.2. Optimización bayesiana

MATLAB es un entorno de programación y análisis numérico ampliamente utilizado en la investigación y desarrollo de modelos de aprendizaje automático. En este capítulo, se

ha utilizado MATLAB para realizar la optimización bayesiana de los modelos de regresión con el objetivo de conseguir el mejor modelo ML para, posteriormente, predecir el precio de cierre de las acciones de la compañía IAG.

La optimización bayesiana se utiliza en el campo del aprendizaje automático para mejorar el rendimiento de los modelos al encontrar la combinación óptima de hiperparámetros que afectan su desempeño [31]. En este estudio, **se ha aplicado la optimización bayesiana a través de MATLAB** para determinar la configuración más adecuada de los hiperparámetros en cada modelo evaluado. Para evaluar los resultados, se han utilizado métricas como el coeficiente de determinación ( $R^2$ ) y la raíz del error cuadrático medio (RMSE).

La optimización bayesiana se basa en el Teorema de Bayes, y utiliza un modelo sustituto que representa de forma probabilística la función objetivo. Este modelo busca los hiperparámetros específicos que maximizan o minimizan el rendimiento del modelo de predicción. A través de un procedimiento iterativo, se obtiene el valor que representa el mejor rendimiento del modelo de predicción [32].

El proceso de optimización ha comenzado importando los datos de entrenamiento desde un archivo de Excel exportado previamente de Python. Se ha utilizado un enfoque de validación cruzada con un valor de 10, lo que implica dividir el conjunto de datos en 10 subconjuntos y realizar el entrenamiento y la evaluación del modelo en varias combinaciones de estos subconjuntos. Es importante destacar que la variable objetivo utilizada en el entrenamiento de los modelos es el precio de cierre de las acciones de IAG (“IAG\_Close”). Esta variable se ha seleccionado tanto para los datos que se corresponden con el período anterior a la pandemia como para el período posterior, lo que permite, como bien se ha mencionado previamente, evaluar la capacidad de los modelos para adaptarse a diferentes escenarios y capturar las variaciones en el precio de las acciones.

Una vez importados los datos, se han seleccionado todos los modelos disponibles en MATLAB para entrenarlos y evaluar su rendimiento<sup>3</sup>. Al entrenar los modelos,

---

<sup>3</sup> Esto incluye una variedad de algoritmos de regresión, como regresión lineal, regresión polinómica, regresión de vectores de soporte (SVR), redes neuronales, entre otros.

MATLAB realiza la optimización bayesiana, ajustando los hiperparámetros de cada modelo y seleccionando aquel que produzca el menor RMSE.

La optimización bayesiana en MATLAB utiliza técnicas avanzadas para explorar y explotar el espacio de búsqueda de hiperparámetros de manera eficiente. Al finalizar el proceso, **se obtiene el modelo con el menor RMSE**, lo que indica que es el modelo que mejor se ajusta a los datos y tiene mayor capacidad predictiva. Es importante mencionar que el uso de MATLAB en el proceso de optimización bayesiana ofrece ventajas como la automatización del proceso de búsqueda de hiperparámetros, la comparación de múltiples modelos y la generación de resultados estadísticos y gráficos para evaluar el rendimiento de los modelos.

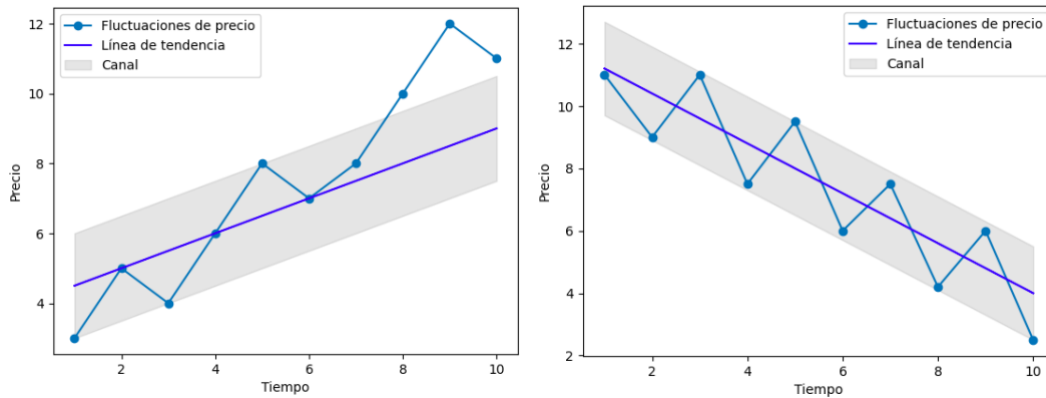
En este sentido, el uso de MATLAB en este estudio ha proporcionado una plataforma sólida para la optimización bayesiana de modelos de regresión, permitiendo encontrar el modelo óptimo que minimiza el RMSE en la predicción del precio de cierre de las acciones de IAG. Este enfoque ha mejorado la capacidad de predicción y ha ofrecido resultados precisos y confiables para el análisis financiero y la toma de decisiones en el mercado de valores.

Es importante destacar que **los modelos de ML son modelos regresivos**, lo que implica que suelen predecir los precios de cotización de un activo hacia abajo. En otras palabras, estos modelos no siempre son capaces de predecir de manera precisa el precio futuro de un activo. Para abordar esta limitación, muchos analistas financieros emplean otras técnicas adicionales en sus análisis.

Una práctica común es colocar la serie temporal de precios entre canales para tener una referencia visual de los posibles movimientos futuros del precio. En este sentido, la línea de canal o retorno es una forma comúnmente utilizada de la línea de tendencia. Según Pring [33], se refiere a una línea recta que conecta dos puntos altos significativos en el caso de tendencias alcistas, o dos puntos bajos significativos en el caso de tendencias bajistas. Estos canales resultan útiles porque, como se puede observar en el *Gráfico 10*, una vez identificados, permiten a los analistas predecir las fluctuaciones de precios dentro de la tendencia principal. Además, pueden proporcionar señales anticipadas de posibles cambios de tendencia al observar cómo la tendencia pierde fuerza gradualmente al

acercarse a la línea del canal. De manera similar, si la línea del canal es superada, como se puede observar en la figura, indica que la tendencia se ha vuelto más pronunciada.

**Gráfico 10. Líneas de canal**



Fuente: Elaboración propia

Para evitar errores, se ha considerado un enfoque cauteloso al realizar las predicciones con modelos de ML, seleccionando así un período de sólo 8 días como conjunto de test.

### 3.2.1. Antes del Covid-19

Como se ha detallado en el anterior capítulo, se ha seleccionado un total de 14 variables de este conjunto de datos para hacer optimización bayesiana. El modelo seleccionado por MATLAB como aquel con menor RMSE ha sido el *Rational Quadratic GPR*<sup>4</sup>, que presenta las siguientes características:

**Tabla 5. Resultados Rational Quadratic GPR**

#### Training Results

|                        |               |
|------------------------|---------------|
| RMSE (Validation)      | 0.12846       |
| R-Squared (Validation) | 0.96          |
| MSE (Validation)       | 0.016501      |
| MAE (Validation)       | 0.097362      |
| Prediction speed       | ~5500 obs/sec |
| Training time          | 53.053 sec    |

<sup>4</sup> Un modelo de regresión gaussiana realiza predicciones que incorporan conocimiento previo (*kernels*) y proporciona medidas de incertidumbre sobre las predicciones. Ver [34].

---

**Model Hyperparameters****Feature Selection: 14/14 individual features selected****PCA: Disabled**

Fuente: Elaboración propia

Como se puede observar en la tabla, se ha identificado que el mejor modelo para los datos antes de la pandemia (“datos\_before\_covid\_train”) ha sido el *Rational Quadratic GPR*. Este modelo ha demostrado tener un RMSE de 0.12846, lo que indica que en promedio las predicciones difieren del valor real en aproximadamente 0.12846 unidades de la variable objetivo.

Además, se ha obtenido un  $R^2$  de 0.96, lo que sugiere que el modelo puede explicar aproximadamente el 96% de la variabilidad de la variable objetivo. Un  $R^2$  alto indica una buena capacidad del modelo para ajustarse a los datos y capturar las relaciones subyacentes.

Se puede observar en las características del modelo unas métricas de evaluación del mismo en el conjunto de validación. Estas serían el MSE (Error Cuadrático Medio) de 0.016501 y un MAE (Error Absoluto Medio) de 0.097362. Estas métricas son útiles para evaluar la calidad de las predicciones del modelo. Un MSE bajo y un MAE bajo indican que las predicciones del modelo son cercanas a los valores reales.

Como se ha mencionado previamente, el objetivo de la optimización bayesiana es encontrar un punto que minimice una función objetivo utilizando el error de clasificación de validación cruzada [35]. La evaluación de la función en sí puede implicar un procedimiento arduo, y esto se aborda mediante el uso de una técnica convencional de reducción de dimensionalidad: el PCA. Al reducir el espacio de características a un número menor de Componentes Principales, se reduce la complejidad computacional, lo que disminuye el tiempo de evaluación [36]. Se ha comprobado que PCA reduce el tiempo de entrenamiento y mejora el rendimiento de clasificación en conjuntos de datos muy grandes [37].

En el contexto específico de este estudio, se exploró la aplicación del PCA después de obtener el mejor modelo, en este caso un *Rational Quadratic GPR*. Se consideró que aplicar el PCA podría potencialmente mejorar aún más los resultados obtenidos, al reducir la dimensionalidad de los datos y eliminar información redundante. Sin embargo, tras realizar la implementación del PCA en el modelo, se observó que no se obtuvo una mejora significativa en los resultados. Esto indica que, en este caso particular, la aplicación del PCA no resultó beneficioso para mejorar el rendimiento del modelo y no contribuyó a una mayor precisión en las predicciones.

A pesar de haber realizado MI en el capítulo anterior, MATLAB ofrece la posibilidad de representar gráficamente las variables más relevantes para la variable objetivo (“IAG\_Close”) utilizando el estadístico *F-test*. Esta representación gráfica permite visualizar de manera clara y concisa las variables que tienen una mayor influencia en la variable objetivo. Cada variable se representa en el eje ‘y’, mientras que en el eje ‘x’ se muestra el valor del estadístico *F-test* correspondiente a esa variable.

La interpretación de este gráfico permite identificar las variables que presentan una mayor relación con la variable objetivo. Aquellas variables que tienen un valor más alto en el estadístico *F-test* indican una mayor relevancia y poder predictivo en la determinación del precio de cierre de las acciones de IAG.

Tras analizar los resultados, se observó que existían tres variables que tenían una influencia significativamente mayor en la variable objetivo en comparación con las demás. Además, se identificaron dos variables adicionales que también mostraban una influencia considerable en la variable objetivo. Estas variables destacadas fueron seleccionadas y consideradas como las más relevantes en el proceso de predicción de la variable objetivo.

En este sentido, se repitió el proceso de optimización bayesiana utilizando exclusivamente estas cinco variables destacadas. El modelo resultante fue un *Exponential GPR* que, como se puede ver en la *Tabla 6* mostró un RMSE aún más reducido en comparación con el modelo anterior, lo cual indica una mayor precisión en la predicción. Además, el  $R^2$  del modelo mejoró, lo que indica una mejor capacidad para explicar la variabilidad de los datos.

**Tabla 6. Resultados Exponential GPR**

**Training Results**

|                        |               |
|------------------------|---------------|
| RMSE (Validation)      | 0.11976       |
| R-Squared (Validation) | 0.98          |
| MSE (Validation)       | 0.014342      |
| MAE (Validation)       | 0.085219      |
| Prediction speed       | ~6900 obs/sec |
| Training time          | 26.551 sec    |
| Model size (Compact)   | ~32 kB        |

**Model Hyperparameters**

**Feature Selection: Top 5/14 features selected using F-Test**

**PCA: Disabled**

Fuente: Elaboración propia

Considerando estos resultados, el **modelo Exponential GPR, utilizando únicamente las cinco primeras variables**, se consideró como **el mejor modelo para la predicción** del precio de cierre de las acciones de IAG. Este modelo fue exportado y utilizado para realizar las predicciones futuras del precio de IAG.

Además, cabe mencionar que, en el análisis realizado en MATLAB, se generaron diversas gráficas que permiten visualizar los resultados de la optimización bayesiana aplicada al modelo *Exponential GPR*. La primera figura, que se corresponde con el *Gráfico 11*, muestra en el eje ‘y’ el valor de las cotizaciones de IAG y en el eje ‘x’ el número de registros. En esta representación, los puntos amarillos corresponden a las predicciones generadas por el modelo, mientras que los puntos azules representan los valores reales. Esta visualización permite comparar la precisión del modelo al observar la proximidad de las predicciones (puntos amarillos) a los valores reales (puntos azules).

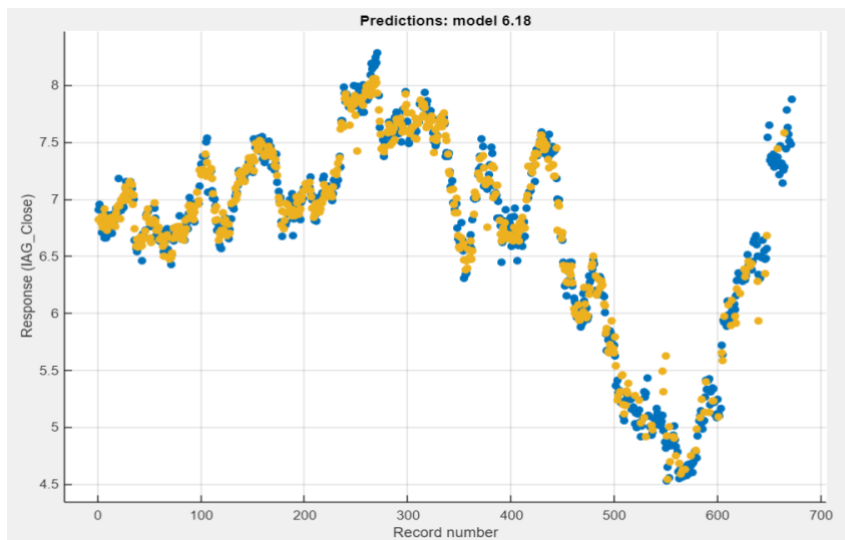
En la segunda figura, en el eje ‘y’ se encuentra la respuesta predicha (*predicted response*), mientras que en el eje ‘x’ se sitúa la respuesta real (*true response*). Se puede observar una línea diagonal negra, la cual representa el caso ideal en el que las predicciones coinciden exactamente con los valores reales. Los puntos azules, dispuestos alrededor de la línea diagonal, muestran la discrepancia entre las predicciones y los valores reales. En la parte



superior de la diagonal, se agrupan los valores que se encuentran por encima de la predicción, mientras que en la parte inferior se sitúan aquellos que están por debajo. Esta visualización proporciona información sobre la calidad de las predicciones del modelo y la tendencia de subestimar los valores reales.

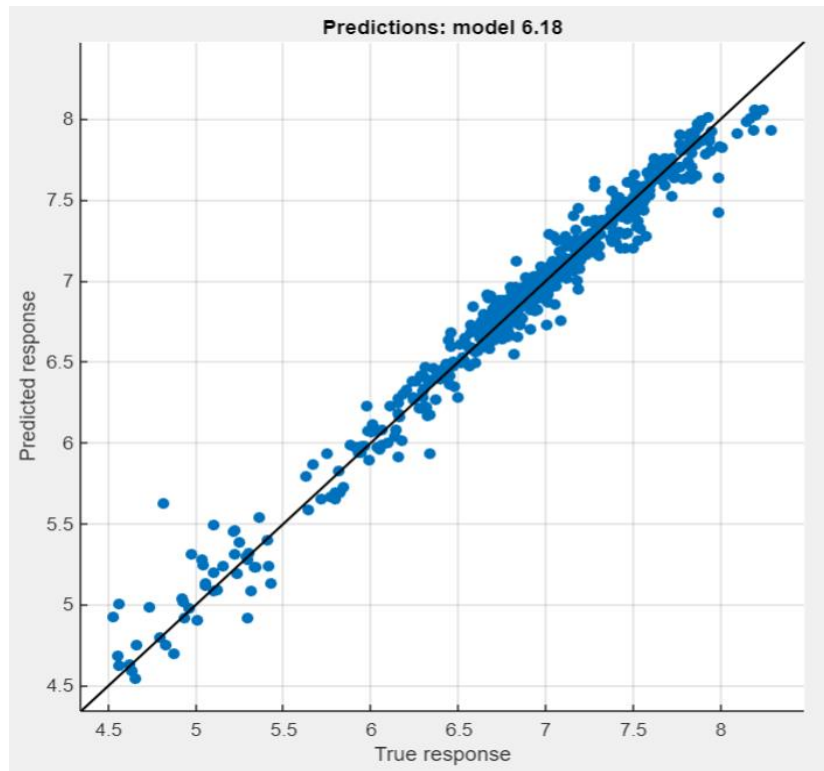
Por último, la tercera figura muestra en el eje 'y' los residuos de la variable objetivo y en el eje 'x' los valores reales (*true response*). En esta representación, los puntos dispersos alrededor de una línea horizontal negra indican la diferencia entre las predicciones del modelo y los valores reales. En las zonas más bajas, los puntos se encuentran más separados de la línea horizontal, lo que sugiere una mayor discrepancia entre las predicciones y los valores reales. En contraste, en las zonas más altas, los puntos se agrupan más cerca de la línea horizontal, indicando una mayor concordancia entre las predicciones y los valores reales. Esta visualización permite evaluar la tendencia y distribución de los residuos del modelo.

**Gráfico 11.** Comparación entre las predicciones Exponenciales GPR y valores reales de las cotizaciones de IAG



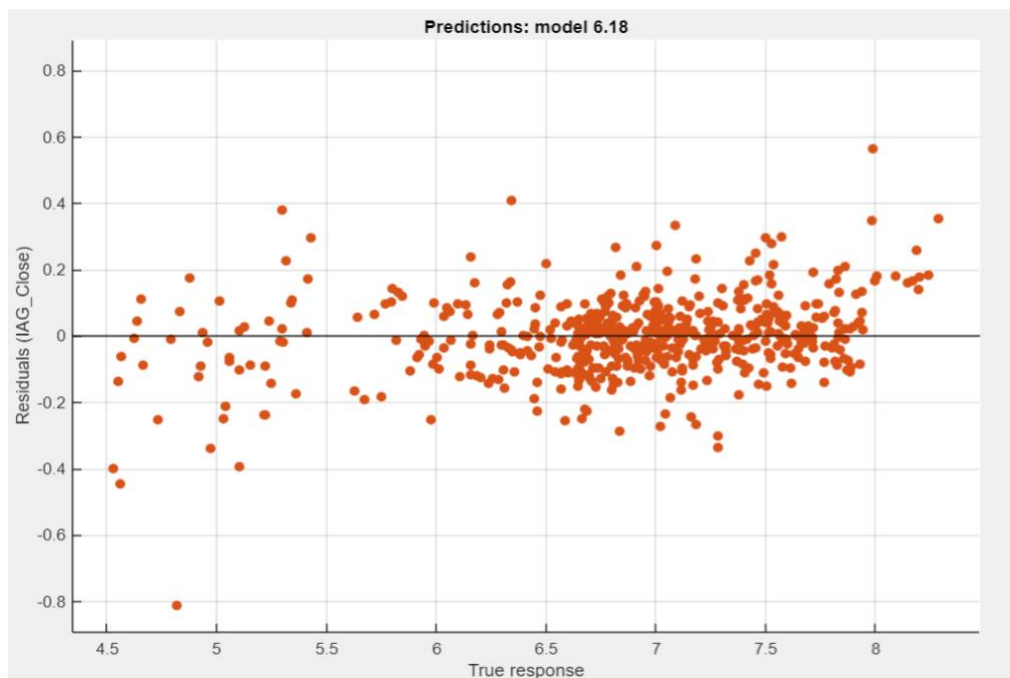
Fuente: Elaboración propia

**Gráfico 12.** Tendencia de las predicciones Exponenciales GPR y discrepancia con los valores reales de las cotizaciones de IAG



Fuente: Elaboración propia

**Gráfico 13.** Residuos Exponential GPR



Fuente: Elaboración propia

### 3.2.2. Durante el Covid-19

Durante el proceso de análisis y modelado de los datos posteriores al Covid-19, se ha seguido un procedimiento similar al mencionado anteriormente. Sin embargo, se ha realizado una adaptación específica para este conjunto de datos debido a ciertas consideraciones. Inicialmente, se seleccionaron un total de 14 variables utilizando el método de MI. Sin embargo, se identificó que la variable correspondiente a “AirlineIndex\_Close” presentaba un alto número de valores faltantes (NA) y no funcionaba de manera óptima en el proceso de optimización bayesiana. Por lo tanto, se tomó la decisión de eliminar esta variable del conjunto de datos. Esta variable fue también eliminada en los siguientes capítulos por la misma razón. Posteriormente, se procedió a aplicar la optimización bayesiana en MATLAB para determinar el modelo con el menor RMSE. El modelo seleccionado por MATLAB como el más adecuado fue un **Squared Exponential GPR**:

*Tabla 7. Resultados Squared Exponential GPR*

#### Training Results

|                        |               |
|------------------------|---------------|
| RMSE (Validation)      | 0.13052       |
| R-Squared (Validation) | 0.97          |
| MSE (Validation)       | 0.017035      |
| MAE (Validation)       | 0.063365      |
| Prediction speed       | ~7200 obs/sec |
| Training time          | 40.897 sec    |
| Model size (Compact)   | ~96 kB        |

#### Model Hyperparameters

**Feature Selection: 13/13 individuals features selected**

**PCA: Disabled**

Fuente: Elaboración propia

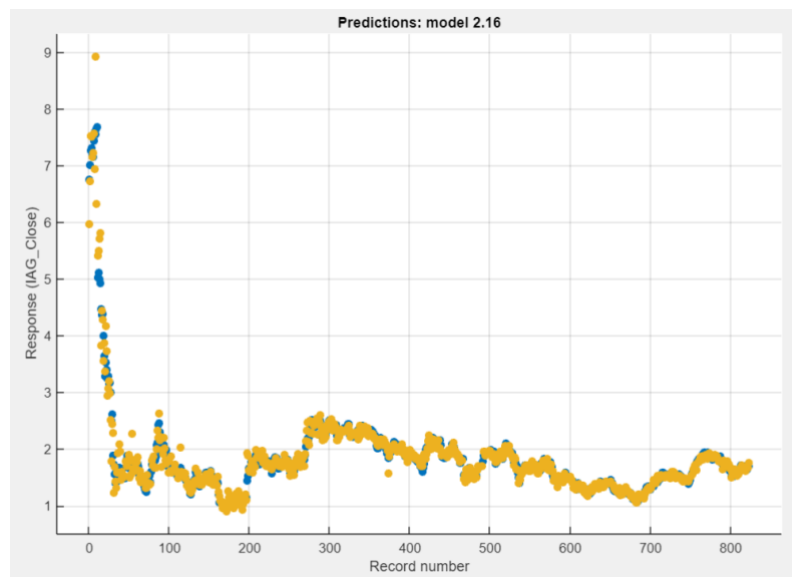
Este modelo ha demostrado tener un RMSE (Error Cuadrático Medio) de 0.13052 y un  $R^2$  de 0.97, lo que sugiere que el modelo puede explicar aproximadamente el 97% de la variabilidad de la variable objetivo.

En cuanto a las métricas de evaluación del mismo en el conjunto de validación el MSE fue de 0.017035 y el MAE de 0.063365. De nuevo, que estas medidas sean bajas indican que las predicciones del modelo son cercanas a los valores reales.

Con respecto al PCA, se exploró también esta técnica después de obtener el mejor modelo. No obstante, tras realizar la implementación del PCA en el modelo, se observó que no se obtuvo una mejora significativa en los resultados. Esto indica que, en este caso particular, la aplicación del PCA no resultó beneficioso para mejorar el rendimiento del modelo y no contribuyó a una mayor precisión en las predicciones.

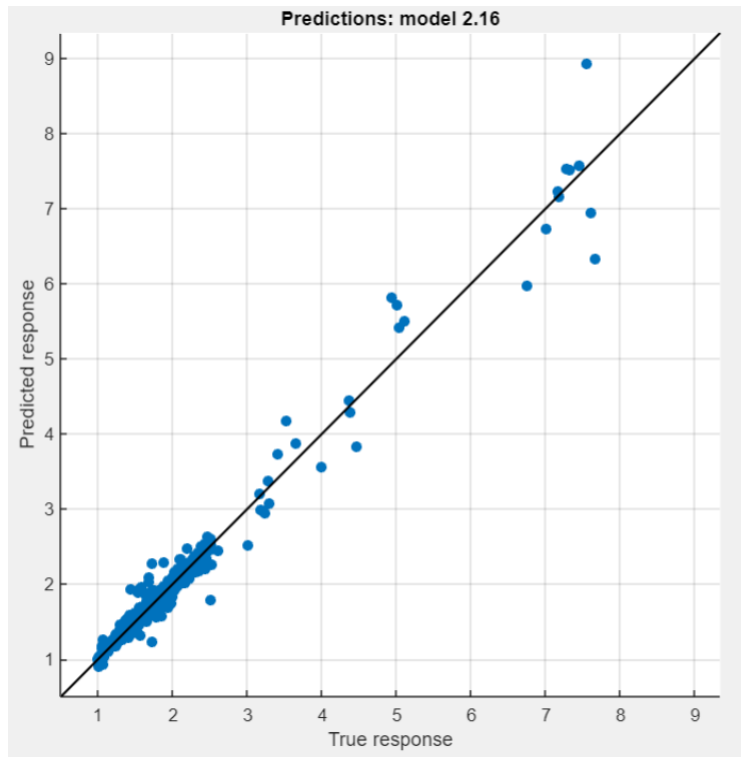
Se muestran continuación las mismas gráficas extraídas en el caso anterior, pero para el *Squared Exponential GPR* sobre los datos después de la pandemia.

**Gráfico 14.** Comparación entre las predicciones Cuadráticas Exponenciales GPR y valores reales de las cotizaciones de IAG



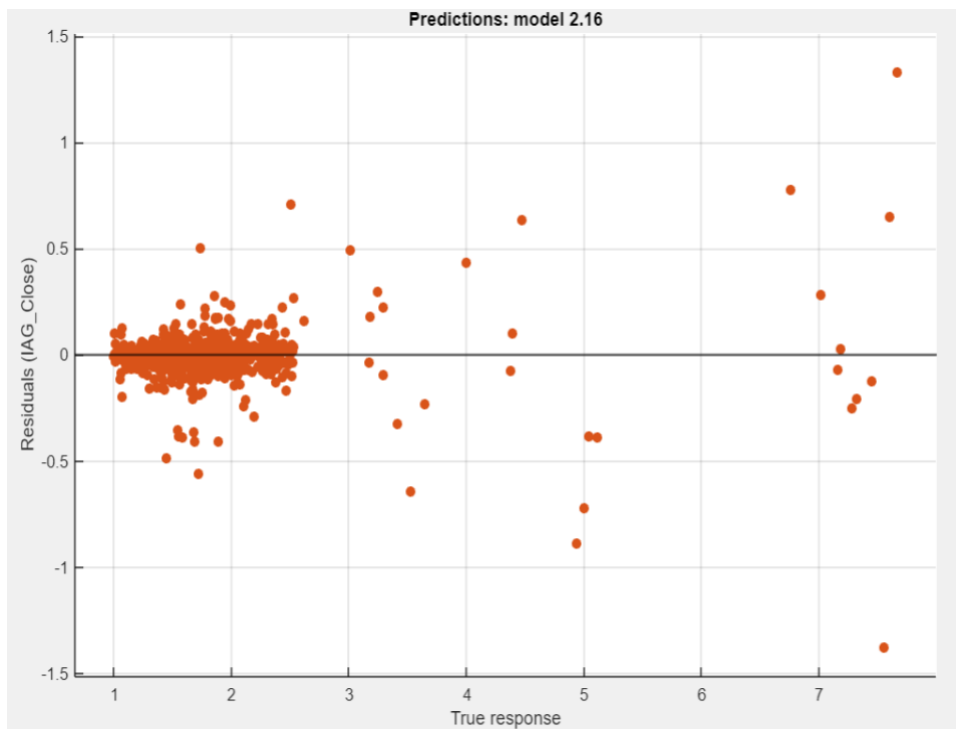
Fuente: Elaboración propia

**Gráfico 15.** Tendencia de las predicciones Cuadráticas Exponenciales GPR y discrepancia con los valores reales de las cotizaciones de IAG



Fuente: Elaboración propia

**Gráfico 16.** Residuos Squared Exponential GPR



Fuente: Elaboración propia

Como se ha descrito para el conjunto de datos anterior, a pesar de haber realizado MI en el capítulo anterior, MATLAB ofrece la capacidad de representar gráficamente las variables más relevantes para la variable objetivo (“IAG\_Close”) utilizando el estadístico F-test.

Una diferencia notable que se observó en la gráfica del estadístico *F-test* para los datos posteriores al Covid-19 en comparación con los datos previos es que no existía una marcada diferencia entre las variables más significativas y las menos significativas. Esto sugería que, en este conjunto de datos, al menos las 7 primeras variables podrían tener un impacto similar en la variable objetivo.

Por esta razón, se procedió a repetir el proceso de optimización bayesiana utilizando diferentes conjuntos de variables predictoras. Se llevaron a cabo iteraciones utilizando conjuntos de 5, 6, 7 y 8 variables, buscando encontrar un modelo con un RMSE mejor que el seleccionado previamente. Sin embargo, a pesar de realizar estas pruebas adicionales, no se logró obtener un modelo con un RMSE inferior al modelo previamente seleccionado. Esto indicaba que el modelo inicialmente elegido, el modelo basado en *Squared Exponential GPR*, seguía siendo el mejor modelo en términos de precisión y capacidad predictiva para los datos posteriores al Covid-19.

### **3.3. Predicción de la cotización de IAG**

Esta sección se basa en la predicción de la cotización de IAG utilizando los modelos seleccionados previamente sobre unos conjuntos de datos separados, que constan de las últimas 8 observaciones en ambos conjuntos de datos.

Estas observaciones fueron apartadas específicamente para fines de prueba y evaluación de la capacidad predictiva de los modelos. De esta manera, los resultados nos brindarán una idea de cómo los modelos se desempeñan en la capacidad de prever el comportamiento futuro de los precios de las acciones de IAG.

#### *3.3.1. Antes del Covid-19*

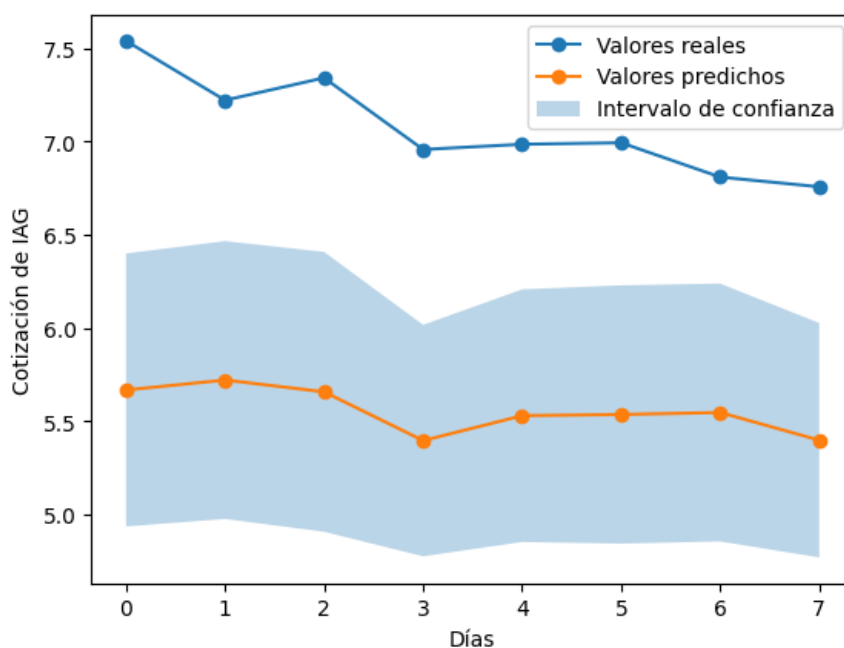
Las predicciones obtenidas a través del *Exponential GPR* fueron las siguientes:

**Tabla 8.** Predicciones Exponential GPR

| Valores reales | Predicciones (y) | A      | B      |        |
|----------------|------------------|--------|--------|--------|
| 7.5380         | 5.6664           | 0.3733 | 4.9347 | 6.3982 |
| 7.2200         | 5.7199           | 0.3800 | 4.9750 | 6.4647 |
| 7.3400         | 5.6566           | 0.3827 | 4.9064 | 6.4067 |
| 6.9560         | 5.3952           | 0.3165 | 4.7748 | 6.0155 |
| 6.9840         | 5.5287           | 0.3454 | 4.8516 | 6.2057 |
| 6.9920         | 5.5352           | 0.3532 | 4.8429 | 6.2275 |
| 6.8080         | 5.5457           | 0.3526 | 4.8547 | 6.2368 |
| 6.7560         | 5.3972           | 0.3206 | 4.7688 | 6.0256 |

Fuente: Elaboración propia

**Gráfico 17.** Predicciones Exponential GPR



Fuente: Elaboración propia

En el gráfico, se puede observar la comparativa entre los valores predichos por el modelo *Exponential GPR* y los valores reales de la cotización de IAG. Los valores predichos se representan mediante la línea naranja, mientras que los valores reales se representan mediante la línea azul.

Se puede notar que **los valores predichos generalmente siguen la tendencia de los valores reales**, lo que indica que el modelo está capturando en cierta medida la relación

subyacente entre las variables de entrada y la variable objetivo. Sin embargo, también se puede observar que **existen algunas discrepancias** entre los valores predichos y los valores reales.

Es importante tener en cuenta que los modelos GPR no pueden predecir con precisión todos los cambios bruscos en el precio de cotización de IAG, lo cual se evidencia en las diferencias entre los valores predichos y los valores reales en la serie temporal.

El intervalo de confianza, representado por el sombreado azul en el gráfico, indica el rango estimado en el cual se espera que se encuentren los valores reales con cierto nivel de confianza. En este caso, el intervalo se muestra como un rango entre dos valores para cada predicción.

### 3.3.2. Durante el Covid-19

En este otro caso, las predicciones obtenidas a través del *Squared Exponential GPR* fueron las siguientes:

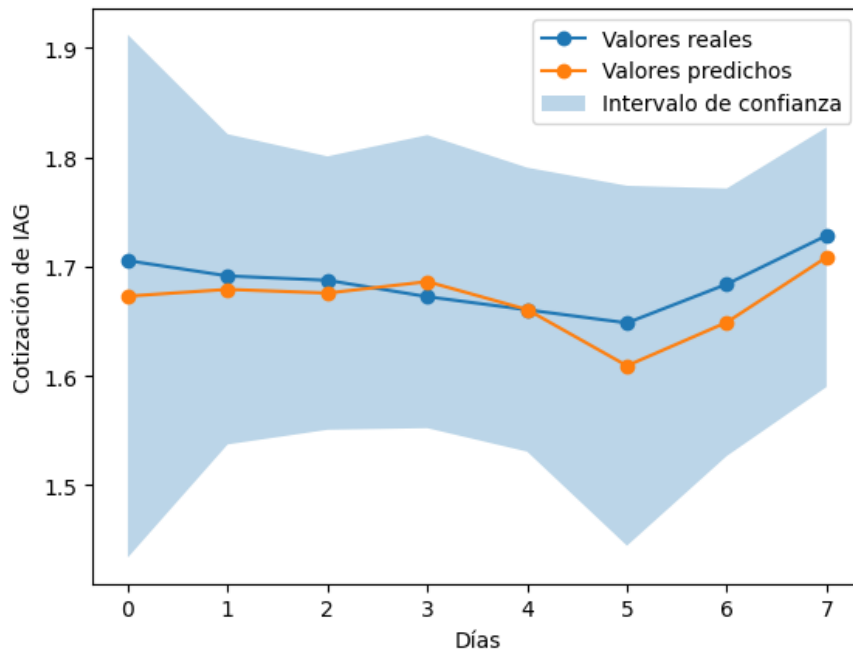
**Tabla 9.** Predicciones Squared Exponential GPR

| Valores reales | Predicciones ( $\hat{y}$ ) | A      | B      |        |
|----------------|----------------------------|--------|--------|--------|
| 1.7055         | 1.6730                     | 0.1220 | 1.4338 | 1.9122 |
| 1.6915         | 1.6792                     | 0.0724 | 1.5373 | 1.8212 |
| 1.6875         | 1.6758                     | 0.0638 | 1.5507 | 1.8009 |
| 1.6725         | 1.6863                     | 0.0684 | 1.5522 | 1.8204 |
| 1.6605         | 1.6607                     | 0.0663 | 1.5307 | 1.7906 |
| 1.6485         | 1.6093                     | 0.0840 | 1.4446 | 1.7740 |
| 1.6840         | 1.6491                     | 0.0625 | 1.5267 | 1.7715 |
| 1.7285         | 1.7087                     | 0.0606 | 1.5899 | 1.8274 |

Fuente: Elaboración propia



**Gráfico 18.** Predicciones Squared Exponential GPR



Fuente: Elaboración propia

En el caso de los valores predichos después del periodo de la pandemia, se puede observar que existe una tendencia general de seguimiento de la variación de los valores reales de la cotización de IAG. Por otro lado, y en contraste con el caso anterior, **los valores predichos sí se encuentran en su mayoría cerca de los valores reales** existiendo discrepancias mínimas. Estas discrepancias pueden deberse a diversos factores, como la volatilidad y la incertidumbre en los datos después del periodo de la pandemia, así como la presencia de otros factores no considerados en el modelo.

## CAPÍTULO 4: MODELOS ML CLASIFICATORIO

En este capítulo, se aborda la aplicación de modelos ML clasificadorio para predecir las tendencias de mercado de la compañía IAG. Siguiendo la metodología establecida en los capítulos anteriores, se utilizarán herramientas como Python y MATLAB para el procesamiento de datos y la implementación de los modelos. Como se mencionó en el capítulo anterior, la predicción del comportamiento de los precios de las acciones de IAG ha sido un objetivo fundamental en este estudio. Sin embargo, en esta etapa se busca ir más allá y explorar la capacidad de los modelos de aprendizaje automático para clasificar las tendencias de mercado, es decir, **determinar si el precio de las acciones de IAG subirá o bajará en función de diferentes variables predictoras.**

### 4.1. Preparación de los datos

Siguiendo la metodología establecida en el capítulo anterior, se ha procedido a la extracción de los datos correspondientes a dos períodos: antes y después de la pandemia de COVID-19.

Las variables utilizadas en cada extracción de datos también han sido descritas en detalle en el capítulo de *feature selection*. Sin embargo, en este caso específico de modelos de aprendizaje automático clasificadorio, se ha realizado una adaptación adicional. En un principio, se procedió a transformar la variable objetivo (“IAG\_Close”) a una categoría binaria, con el objetivo de aplicar técnicas de clasificación y realizar un análisis inicial. Sin embargo, tras aplicar el enfoque de optimización bayesiana a través de MATLAB, se obtuvieron modelos con una capacidad de predicción limitada, alcanzando solo un 50% de precisión. Ante esta situación, se decidió ampliar el enfoque y transformar todas las variables seleccionadas en el proceso de *feature selection* a categorías binarias. Esta estrategia se implementó con el propósito de evaluar si la inclusión de más variables en el modelo mejoraría la capacidad de predicción y proporcionaría resultados más sólidos y fiables.

Cabe destacar que, para transformar las variables continuas a categóricas, se asignó el valor 1 cuando la variable correspondiente mostraba un incremento en su valor con respecto al día anterior, y se asignó el valor 0 cuando la variable mostraba una disminución. Esta transformación fue la que permitió obtener una representación

categorica de las variables, abriendo así la posibilidad de utilizar técnicas de clasificación en el análisis de tendencias de mercado.

Además, es importante mencionar que al igual que en el capítulo anterior, los últimos ocho registros de datos corresponden a las predicciones generadas mediante herramientas de modelado en MATLAB. Estos datos se han separado del conjunto de datos principal y se utilizaron posteriormente para evaluar la precisión de los modelos desarrollados. Esta separación garantiza que los resultados obtenidos sean imparciales y no se vean afectados por la información de predicción utilizada en el proceso de construcción de los modelos.

## 4.2. Optimización bayesiana MATLAB

El proceso de optimización en este caso es similar al abordado en el capítulo anterior, que implica la extracción de datos y la aplicación de un enfoque de validación cruzada. Sin embargo, en esta ocasión, la variable objetivo utilizada en el entrenamiento de los modelos es la variable categorica del precio de cierre de las acciones de IAG (“IAG\_Close\_Categorica”). Se ha seleccionado esta variable tanto para el conjunto de datos previo a la pandemia como para el conjunto posterior, lo que permite evaluar la capacidad de los modelos para adaptarse a diferentes escenarios y capturar las variaciones en la dirección del precio de las acciones.

Al igual que en el problema de regresión, una vez importados los datos, se han seleccionado todos los modelos disponibles en MATLAB para entrenarlos y evaluar su rendimiento en un problema de clasificación. Sin embargo, a diferencia del enfoque anterior que buscaba minimizar el RMSE, en este caso, durante el entrenamiento de los modelos, **se utiliza la optimización bayesiana para ajustar los parámetros y seleccionar aquel modelo que logre la mayor precisión (*accuracy*)** en la validación cruzada. Dado que estamos abordando un problema de clasificación, el objetivo principal se centra en maximizar la precisión del modelo en la correcta clasificación de las categorías.

### 4.2.1. Antes del Covid-19

Se ha procedido a importar todos los datos del conjunto de entrenamiento, que incluye las mejores variables seleccionadas, en MATLAB. Durante el proceso de entrenamiento, se ha identificado que el modelo con el mejor rendimiento en términos de precisión

(*accuracy*) ha sido un *Ensemble*<sup>5</sup>, el cual ha logrado una precisión del 76.7%. Esta métrica de precisión representa la proporción de clasificaciones correctas realizadas por el modelo en relación con el total de observaciones en el conjunto de validación.

**Tabla 10.** Resultados Ensemble (*accuracy*: 76.7%)

**Training Results**

|                         |              |
|-------------------------|--------------|
| Accuracy (Validation)   | 76.7%        |
| Total cost (Validation) | 157          |
| Prediction speed        | ~470 obs/sec |
| Training time           | 16.2 sec     |
| Model size (Compact)    | ~162 kB      |

**Model Hyperparameters**

**Feature Selection: 13/13 individuals features selected**

**PCA: Disabled**

Fuente: Elaboración propia

Además de la precisión, se muestran otras métricas relevantes del modelo clasificador. El costo total (*total cost*) obtenido en la validación ha sido de 157, lo que indica el costo asociado a las clasificaciones incorrectas realizadas por el modelo. Un valor más bajo de costo total indica un menor impacto económico de las clasificaciones erróneas. Otra métrica considerada es la velocidad de predicción (*prediction speed*), la cual se ha registrado en 470 observaciones por segundo. Esta medida refleja la eficiencia del modelo en términos de velocidad de procesamiento y capacidad para realizar predicciones rápidamente.

Previo a la generación de predicciones utilizando el mejor modelo entrenado con todas las variables disponibles, se ha llevado a cabo un proceso similar al realizado en el capítulo anterior. En este caso, se ha aplicado el enfoque de selección de características utilizando pruebas F clásicas y análisis de *chi cuadrado* para evaluar la importancia de cada variable en relación con la variable objetivo. Como resultado de este análisis, se

---

<sup>5</sup> Este modelo, en lugar de depender exclusivamente de un único árbol de decisión y confiar en haber tomado las decisiones correctas en cada división, permite considerar una muestra de árboles de decisión. A través de estos métodos, se calcula qué características utilizar o qué preguntas plantear en cada división, y se crea un predictor final basado en los resultados agregados de los árboles de decisión muestreados. De esta manera, se busca mejorar la precisión y la estabilidad de las predicciones al integrar la información de múltiples árboles en un modelo conjunto. Ver [38].

identificaron tres variables que tenían un impacto significativo en la predicción de la variable objetivo. Además, se observó que otras tres variables también tenían cierta influencia en la predicción, aunque de manera menos pronunciada.

Por esta razón, se optó por realizar diferentes experimentos de optimización bayesiana utilizando diferentes conjuntos de variables, incluyendo tres, cinco y seis variables. No obstante, no se logró mejorar la precisión del modelo clasificadorio previamente mostrado en ninguna de las combinaciones evaluadas.

Como último paso, y antes de seleccionar el *Ensemble* como el modelo definitivo para generar las predicciones, se decidió realizar un análisis adicional utilizando la técnica de PCA. Como se mencionó en el capítulo anterior, PCA es una herramienta que se utiliza para reducir la dimensionalidad de un conjunto de variables, manteniendo la mayor parte de la variabilidad de los datos. En este caso, se aplicó PCA utilizando todas las variables disponibles en el conjunto de datos.

El objetivo de aplicar PCA fue evaluar si la reducción de dimensionalidad mediante la combinación lineal de variables podría mejorar el rendimiento del modelo clasificadorio. En esta ocasión, al realizar el análisis con todas las variables y PCA, el modelo resultante también fue un *Ensemble*, pero con una precisión ligeramente mejor, alcanzando un 77.6%.

**Tabla 11.** Resultados Ensemble (accuracy: 77.6%)

**Training Results**

|                         |              |
|-------------------------|--------------|
| Accuracy (Vlidation)    | 77.6%        |
| Total cost (Validation) | 151          |
| Prediction speed        | ~580 obs/sec |
| Training time           | 12.764 sec   |
| Model size (Compact)    | ~158 kB      |

**Model Hyperparameters**

**Feature Selection: 13/13 individuals features selected**

**PCA: Explaining at least 95% variance**

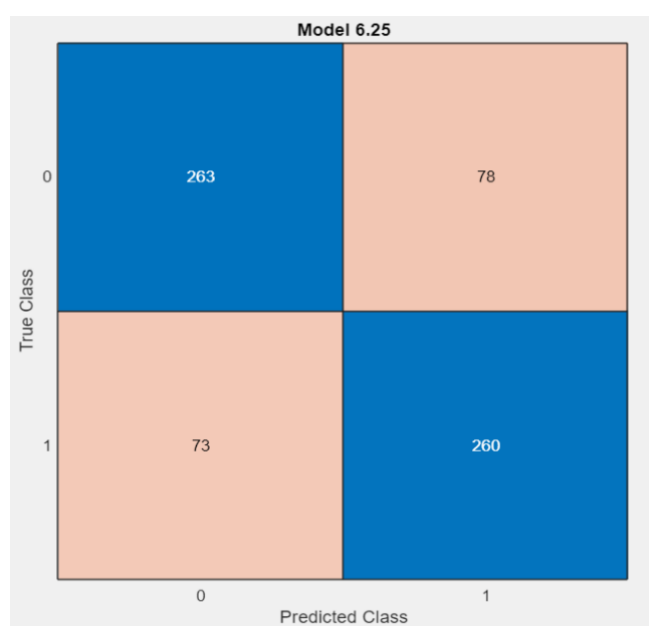
Fuente: Elaboración propia

Finalmente, ha sido este modelo el elegido para generar las predicciones de la dirección de los precios de las acciones de IAG, ya que ha demostrado tener la mayor precisión entre todas las combinaciones evaluadas, considerando tanto el uso de todas las variables como la aplicación de PCA.

Además, en el análisis realizado para el modelo clasificatorio utilizando MATLAB, se generan diferentes gráficas que permiten evaluar los resultados de la optimización bayesiana aplicada al modelo *Ensemble*. Una de estas gráficas es la matriz de confusión<sup>6</sup>, que proporciona información sobre la clasificación realizada por el modelo.

Como se puede ver en el *Gráfico 19*, la matriz de confusión consta de cuatro cuadrantes. En el cuadrante superior izquierdo, se muestra el número de observaciones correctamente clasificadas como negativas (263). En el cuadrante superior derecho, se encuentra el número de observaciones clasificadas incorrectamente como positivas (falsos positivos, que serían 78). En el cuadrante inferior izquierdo, se muestra el número de observaciones clasificadas incorrectamente como negativas (falsos negativos, que serían 73). Por último, en el cuadrante inferior derecho, se encuentra el número de observaciones correctamente clasificadas como positivas (260). El modelo ha logrado clasificar correctamente la mayoría de las observaciones, pero también ha cometido algunos errores.

**Gráfico 19.** Matriz de confusión, Ensemble



Fuente: Elaboración propia

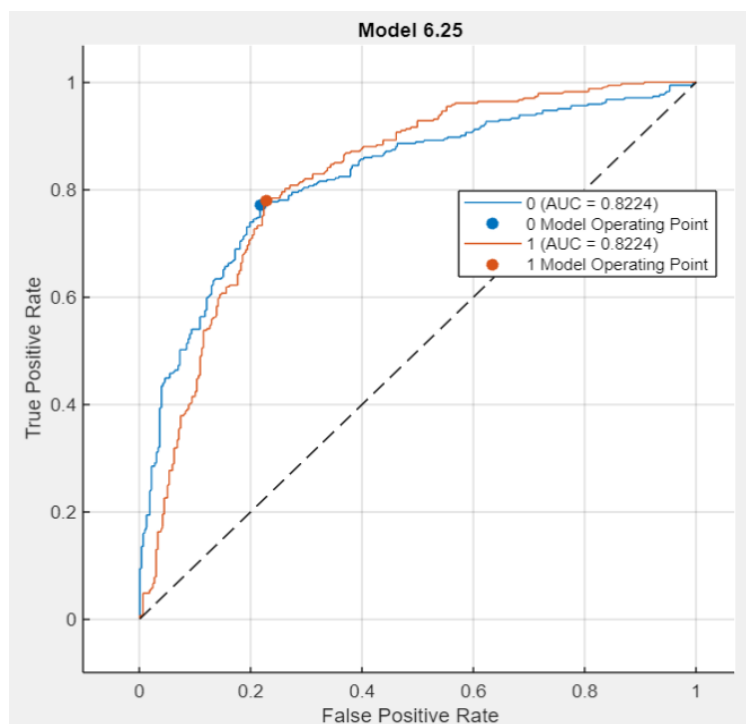
<sup>6</sup> Ver [39].

Además de la matriz de confusión, la siguiente figura representa la curva ROC. Esta representación gráfica muestra la relación entre la tasa de verdaderos positivos (TPR) y la tasa de falsos positivos (FPR) para diferentes umbrales de clasificación. En este caso, se generó una curva ROC para los valores clasificados como 0 y los valores clasificados como 1. En el gráfico de la curva ROC, el eje 'x' representa la tasa de falsos positivos (FPR) y el eje 'y' representa la tasa de verdaderos positivos (TPR). Cada punto en la curva representa un umbral de clasificación diferente.

El valor AUC es una métrica que se utiliza para evaluar el rendimiento del modelo. Cuanto más cercano sea el valor AUC a 1, mejor será la capacidad del modelo para distinguir entre las clases positiva y negativa. En este caso, se obtuvo un **AUC de 0.8224** tanto para los valores clasificados como 0 como para los valores clasificados como 1, lo que indica un buen rendimiento en la distinción entre las clases positiva y negativa.

Además, la presencia de un punto en el medio de la curva ROC indica que los datos están balanceados en términos de las clases positiva y negativa. Esto significa que hay un equilibrio entre los casos clasificados correctamente como positivos y los casos clasificados correctamente como negativos.

*Gráfico 20. Curva ROC, Ensemble*



Fuente: Elaboración propia

#### 4.2.2. Después del Covid-19

Para los datos correspondientes al período posterior a la pandemia, se importaron todos los datos del conjunto de entrenamiento con las variables seleccionadas en el capítulo segundo transformadas a categórica. Entre los modelos evaluados, se observó que el mejor desempeño lo obtuvo un modelo *Linear Discriminant*<sup>7</sup> con una precisión del 81.4%.

**Tabla 12. Resultados Linear Discriminant**

##### Training Results

|                         |                |
|-------------------------|----------------|
| Accuracy (Vlidation)    | 81.4%          |
| Total cost (Validation) | 153            |
| Prediction speed        | ~10000 obs/sec |
| Training time           | 1.8915 sec     |
| Model size (Compact)    | ~8 kB          |

##### Model Hyperparameters

**Feature Selection: 13/13 individuals features selected**

**PCA: Disabled**

Fuente: Elaboración propia

En este caso, el costo total (*total cost*) obtenido en la validación fue de 153, lo que indica el costo asociado a las clasificaciones incorrectas realizadas por el modelo.

De nuevo para este conjunto de datos, previo a la generación de predicciones, se llevó a cabo *feature selection* aplicando la medida de *chi cuadrado* para evaluar la importancia de cada variable en relación con la variable objetivo. En este caso, las variables mostraron un patrón ascendente en términos de importancia, creando una especie de ‘escalera inversa’. A medida que se avanzaba en la selección de variables, su impacto en la variable objetivo iba aumentando de manera progresiva, lo que dificultó determinar qué conjunto de variables era más adecuado para predecir las velas de IAG.

En este contexto, se tomó la decisión de realizar exclusivamente optimización bayesiana utilizando una selección de 5 variables, sin lograr mejorar la precisión del modelo clasificatorio presentado anteriormente.

---

<sup>7</sup> Ver [40].



Debido a este resultado, se optó por llevar a cabo un análisis adicional utilizando la técnica de PCA, escogiendo todas las variables disponibles en el conjunto de datos. Siguiendo el mismo enfoque empleado previamente, el modelo resultante obtenido a través de este análisis fue un *Quadratic Discriminant*<sup>8</sup>, que mostró una precisión del 81.5%.

**Tabla 13.** Resultados *Quadratic Discriminant*

**Training Results**

|                         |               |
|-------------------------|---------------|
| Accuracy (Validation)   | 81.5%         |
| Total cost (Validation) | 152           |
| Prediction speed        | ~3500 obs/sec |
| Training time           | 4.8454 sec    |
| Model size (Compact)    | ~10 kB        |

**Model Hyperparameters**

**Feature Selection: 13/13 individuals features selected**

**PCA: Explaining at least 95% variance**

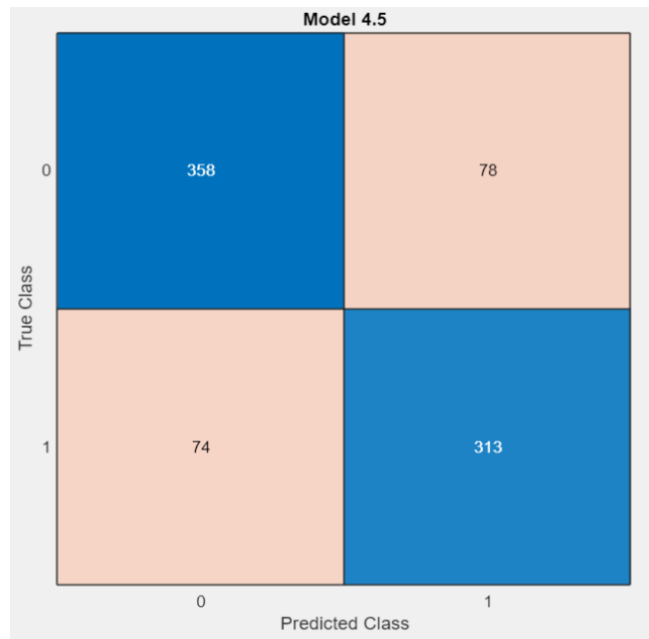
Fuente: Elaboración propia

Finalmente, este fue el modelo seleccionado como el modelo final para generar las predicciones de las velas de IAG al haber demostrado tener la mayor precisión entre todas las combinaciones evaluadas.

Para el *Quadratic Discriminant* la matriz de confusión obtenida ha sido la que se muestra en el *Gráfico 21*. Al igual que para el caso anterior, el modelo ha logrado clasificar correctamente la mayoría de las observaciones, a pesar de haber cometido algunos errores.

<sup>8</sup> Ver [40].

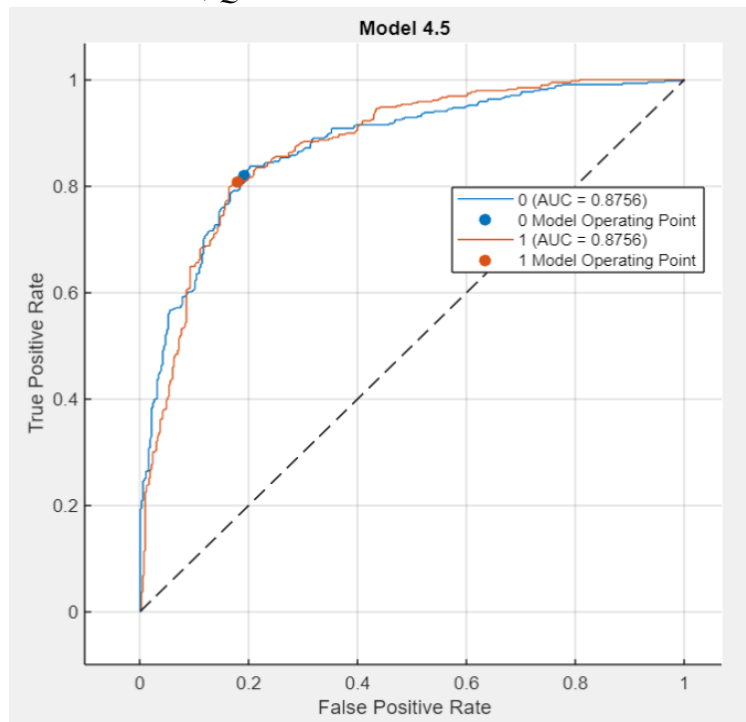
**Gráfico 21.** Matriz de confusión, *Quadratic Discriminant*



Fuente: Elaboración propia

Por otro lado, en la curva ROC se obtuvo un **AUC de 0.8756**, superior al caso anterior que había sido de 0.8224, indicando un buen rendimiento en la distinción entre las clases positiva y negativa. Además, vuelve a existir un equilibrio entre los casos clasificados correctamente como positivos y los casos clasificados correctamente como negativos.

**Gráfico 22.** Curva ROC, *Quadratic Discriminant*



Fuente: Elaboración propia

### 4.3. Predicción de la cotización de IAG

En esta sección, para realizar la predicción de las velas de IAG, se escogieron los modelos seleccionados en el apartado anterior y se aplicaron a los conjuntos de datos *test*, que constaban de las últimas 8 observaciones en ambos casos.

Estas observaciones han sido reservadas específicamente con el propósito de evaluar la capacidad predictiva de los modelos en la predicción del comportamiento futuro de las velas de IAG. Al realizar estas predicciones, se esperó obtener información sobre cómo el modelo se desempeñaba en la tarea de prever si la vela será de color rojo o verde, indicando así el movimiento ascendente o descendente del precio de las acciones de IAG en un periodo determinado.

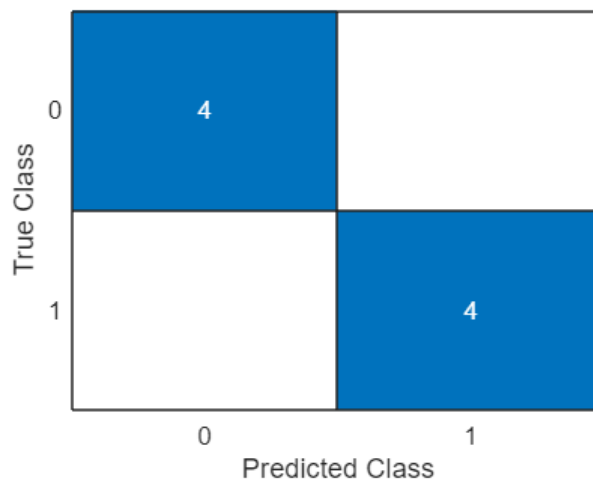
#### 4.3.1. Antes del Covid-19

Las predicciones obtenidas a través del *Ensemble* fueron las siguientes:

*Tabla 14. Predicciones Ensemble*

| Valores reales | Predicciones ( $\hat{y}$ ) |
|----------------|----------------------------|
| 1              | 1                          |
| 0              | 0                          |
| 1              | 1                          |
| 0              | 0                          |
| 1              | 1                          |
| 1              | 1                          |
| 0              | 0                          |
| 0              | 0                          |

*Gráfico 23. Predicciones Ensemble, Matriz de Confusión*



Fuentes: Elaboraciones propias

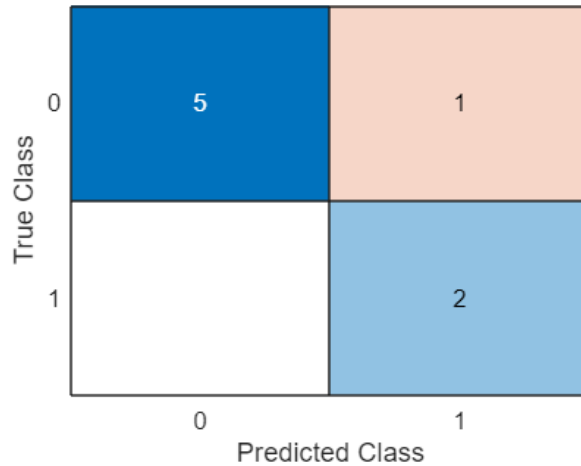
### 4.3.2. Durante el Covid-19

Las predicciones obtenidas a través del *Quadratic Discriminant* fueron las siguientes:

**Tabla 15.** Predicciones *Quadratic Discriminant*

| Valores reales | Predicciones (y) |
|----------------|------------------|
| 0              | 1                |
| 0              | 0                |
| 0              | 0                |
| 0              | 0                |
| 0              | 0                |
| 0              | 0                |
| 1              | 1                |
| 1              | 1                |

**Gráfico 24.** Predicciones *Quadratic Discriminant*, Matriz de Confusión



Fuentes: Elaboraciones propias

### 4.4. Series temporales

En el campo de la economía, las series temporales desempeñan un papel fundamental. Se utilizan cuando se estudian dos variables relacionadas: el tiempo y una variable objetivo que depende de esta última. La forma general de una serie temporal se puede describir mediante la siguiente ecuación:

$$y(t) = \text{tendencia} + \text{estacionalidad} + \text{random fluctuation}$$

En esta ecuación, 'y(t)' representa la variable objetivo en un momento específico 't'. La tendencia se refiere a la dirección general o patrón de crecimiento o declive a lo largo del tiempo. La estacionalidad se relaciona con patrones recurrentes o periódicos que se repiten en intervalos regulares, como las estaciones del año o los días de la semana. La *random fluctuation*, es decir, las fluctuaciones aleatorias, también conocidas como ruido o variabilidad no explicada, representan las desviaciones no sistemáticas de la serie temporal.

En este análisis, **consiste en eliminar la tendencia y la estacionalidad para poder analizar más detalladamente la *random fluctuation***. Al eliminar la tendencia y la estacionalidad, se puede obtener una visión más clara de las variaciones irregulares o aleatorias que pueden contener información valiosa sobre el comportamiento de la variable objetivo en un momento dado. Esto puede ayudar a identificar patrones ocultos, anomalías o cambios estructurales en los datos y mejorar la capacidad predictiva de los modelos utilizados en el análisis de series temporales.

Para eliminar la tendencia en una serie temporal, se utiliza el operador diferencia. Este operador, aplicado a una serie temporal 'y(t)', se define como  $(1-z^{-1})$ , donde 'z' representa el desplazamiento en el tiempo.

La aplicación del operador diferencia transforma la serie 'y(t)' en la diferencia entre dos valores consecutivos:

$$y(t) = y(t) - y(t - 1)$$

Esta transformación tiene como resultado la eliminación de la tendencia lineal presente en la serie temporal. Si la serie sigue una línea recta, al aplicar el operador diferencia, se aproximará a una constante. Por ejemplo, si tenemos una serie temporal que sigue una recta con una pendiente de 1, al aplicar el operador diferencia, se obtendrá una constante. Esto se puede ilustrar mediante la expresión  $t = t - (t - 1)$ , donde 't' representa los valores de la serie. De esta manera, se transforma una serie que sigue una línea recta en una serie constante.

En el caso de que la serie temporal siga una parábola de grado 2 y los datos se ajusten a esta forma, se utiliza el operador diferencia elevado al cuadrado, es decir,  $(1-z^{-2})$ . Al aplicar este operador, se logra transformar la serie temporal de una parábola en una línea recta.

En resumen, el operador diferencia es una herramienta utilizada para eliminar la tendencia en una serie temporal. Su aplicación permite aproximar la serie a una forma estacionaria, como una constante, lo que facilita el análisis de las fluctuaciones aleatorias presentes en los datos.

Para eliminar la estacionalidad en una serie temporal, es necesario identificar los patrones repetitivos que ocurren en un periodo determinado. Por ejemplo, en el caso de las ventas de roscones en Navidad, es común observar un aumento en las ventas durante ese periodo y una disminución el resto del año.

En este caso, se utiliza nuevamente el operador diferencia, pero en lugar de utilizar un desplazamiento de una unidad de tiempo, se utiliza un desplazamiento de 's' unidades de tiempo, donde 's' representa el periodo en el cual se repite la gran tendencia estacional. La expresión del operador diferencia en este caso sería  $(1-z^{-s})$ , quedando la serie temporal de la siguiente forma:

$$y(t) = y(t) - y(t - s)$$

Una vez eliminados tanto la tendencia como la estacionalidad, se puede estudiar la componente de fluctuación aleatoria o *random fluctuation* en la serie temporal.

Una de las formas comunes de aproximarse a la fluctuación aleatoria es mediante un **modelo autorregresivo**. Este modelo utiliza los propios valores de la serie temporal, pero considerando temporalidades más bajas. Esto implica que cada valor en la serie se modela en función de los valores anteriores de la misma serie, con un enfoque en las relaciones de dependencia temporal.

Al aplicar un modelo autorregresivo, se busca capturar las variaciones aleatorias y no sistemáticas presentes en los datos, permitiendo un análisis más detallado de la componente de fluctuación aleatoria en la serie temporal.

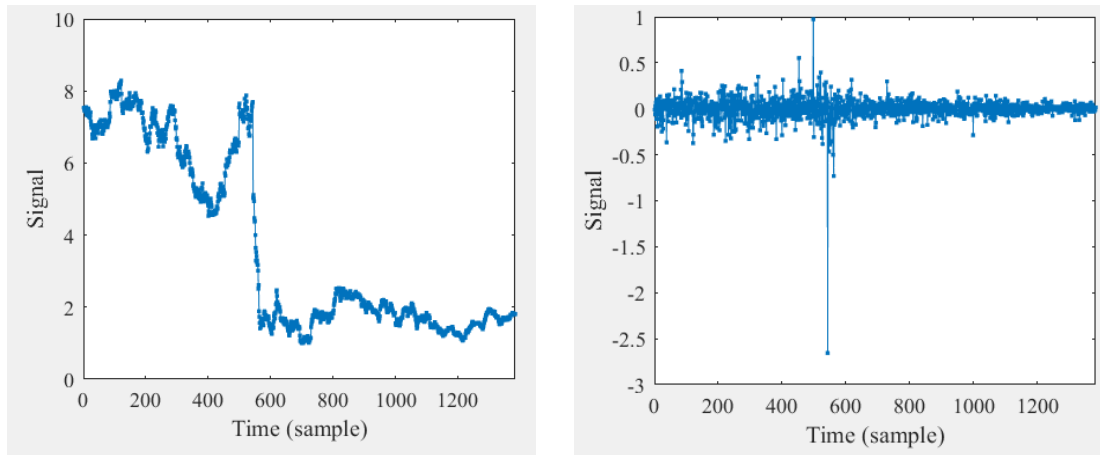
#### *4.4.1. Serie temporal en MATLAB*

Para predecir la cotización de IAG utilizando series temporales, se ha utilizado el lenguaje de programación MATLAB. Se ha importado una muestra de datos que incluye las fechas junto con los precios de cierre de IAG desde el 1 de enero de 2018 hasta el 31 de mayo de 2023.

El primer paso consistió en **eliminar la tendencia de la serie temporal**, con el objetivo de obtener una serie estacionaria. Esto se logró aplicando técnicas de transformación

adecuadas, introduciendo un operador diferencia de 1, que permitió eliminar la componente lineal de tendencia en los datos.

**Gráfico 25.** Serie temporal, eliminación de la tendencia

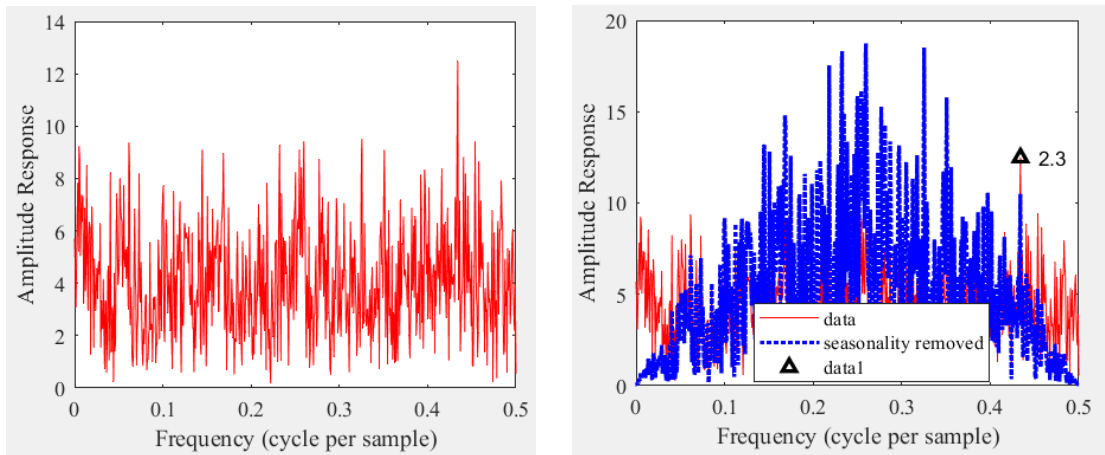


Fuente: Elaboración propia

El siguiente paso consistió en abordar la **eliminación de la estacionalidad** presente en los datos. Al analizar la gráfica, se constató que no se observaba una distribución regular de barras más altas en intervalos específicos, lo que indicaba la ausencia de un patrón estacional evidente en este modelo. Sin embargo, se identificó la presencia de una barra más alta al final de la línea temporal, lo cual generó la necesidad de investigar y abordar esta particularidad. Con el propósito de eliminar dicha barra y evaluar su impacto en la serie de datos, se realizó una prueba mediante la eliminación de la barra alta que se observa en el *Gráfico 27* en la figura de la izquierda.

Al ejecutar el procedimiento, se encontró que la eliminación de la barra requería la eliminación de dos muestras consecutivas. Como resultado de esta eliminación, se generó la figura que se muestra en la derecha. No obstante, al analizar esta nueva figura, se constató que la modificación realizada no logró mejorar la estacionalidad de los datos, ya que se observaron incluso más barras altas en comparación con la situación anterior.

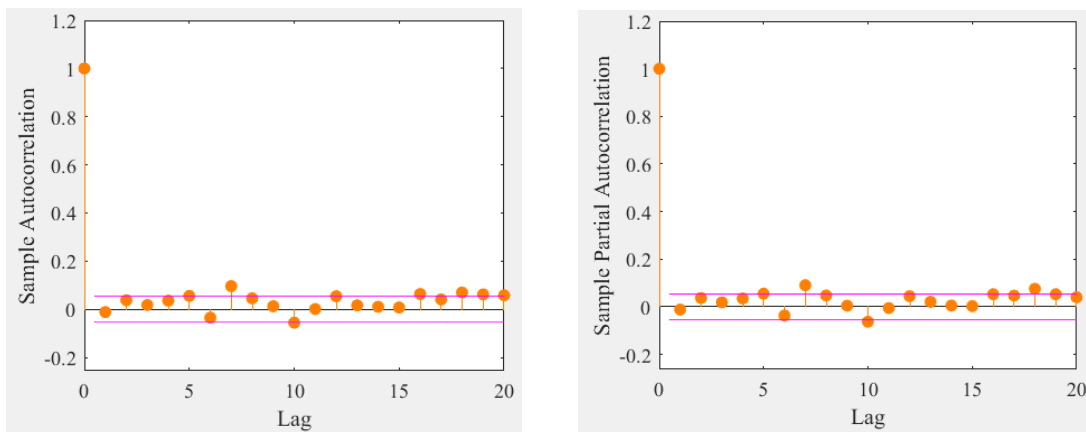
**Gráfico 26.** Serie temporal, eliminación de la estacionalidad



Fuente: Elaboración propia

Una vez eliminada la tendencia y la estacionalidad, quedaría la *random fluctuation*, la cual aproximamos utilizando un modelo ARMA. Al observar el gráfico de autocorrelación, se tuvo que identificar los puntos que se encontraban fuera del intervalo de confianza marcado en rosa que, en este caso particular, eran los puntos 1 y 8. Lo mismo ocurrió en el gráfico de autocorrelación parcial.

**Gráfico 27.** Serie temporal, random fluctuation



Fuente: Elaboración propia



Después de ajustar todos los parámetros, el ajuste del modelo quedó de la siguiente manera:

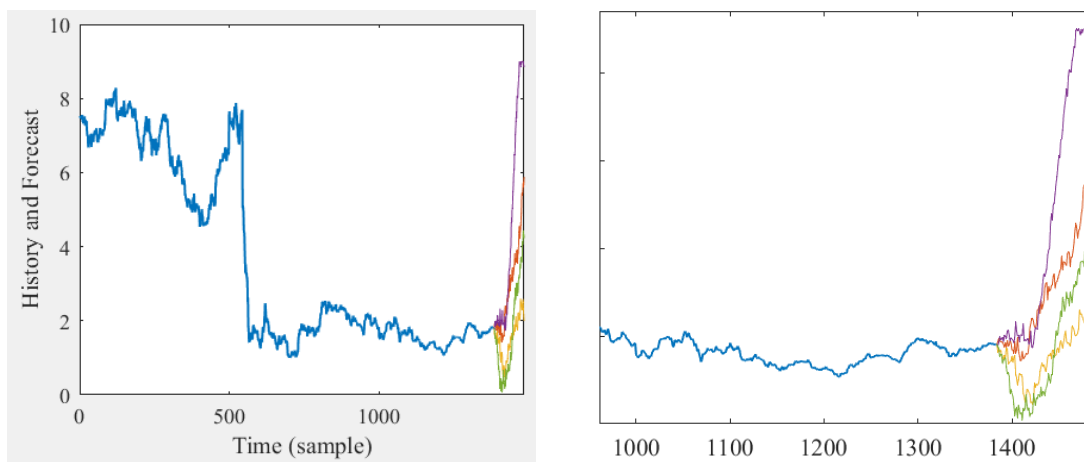
**Tabla 16.** *Serie temporal, model fitting*

| <b>Model Fitting</b> |     |
|----------------------|-----|
| <b>ARLags</b>        | 1 8 |
| <b>MALags</b>        | 1 8 |
| <b>D</b>             | 1   |
| <b>Seasonalit</b>    | 0   |

Fuente: Elaboración propia

Una vez finalizado este proceso, se generaron las predicciones utilizando cuatro modelos de series temporales y utilizando 100 muestras de ejemplo, con un rango mínimo de 0 y máximo de 9. El resultado obtenido fue el siguiente:

**Gráfico 28.** *Serie temporal, predicciones*



Fuente: Elaboración propia

#### 4.4.2. *Serie temporal en PYTHON*

Se ha analizado un **modelo ARIMA** en Python para predecir la cotización de IAG debido a su mayor precisión en comparación con Matlab. En MATLAB, como se ha demostrado en el caso anterior, el proceso es más manual, ya que se pueden seleccionar las muestras de ejemplo, el número de modelos y el rango en el cual se generarán las predicciones. En contraste, en Python, el proceso es más automatizado y se basa en métodos estándar de modelado de series temporales, lo cual brinda resultados más precisos y reproducibles.

El modelo ARIMA de Python proporciona una mayor especificidad en cuanto a la configuración de los parámetros. Para este estudio, se ha optado por estudiar las predicciones en un periodo posterior a la pandemia, abarcando desde el 4 de abril de 2020 hasta el 31 de mayo de 2023.

El proceso de modelado en Python siguió varios pasos de manera similar al caso anterior<sup>9</sup>. Primero, se importó la función ‘auto\_arima’ de la biblioteca ‘pmdarima.arima’, la cual sirve para encontrar el mejor modelo ARIMA para los datos.

Se siguieron los mismos pasos que en Matlab, comenzando por eliminar la estacionalidad utilizando la prueba ADF (*Augmented Dickey-Fuller*). Esta prueba determina si hay estacionalidad en los datos y para este conjunto de datos indicó que no existía.

A continuación, se fijaron los valores de los parámetros p, d y q del modelo ARIMA. Se utilizó la función ‘auto\_arima’ para calcular diferentes modelos ARIMA y evaluar su rendimiento mediante la medición del error. El modelo ARIMA seleccionado como el mejor fue: **SARIMAX(2, 1, 1)x(2, 1, 0, 5)**.

Es importante destacar que estos pasos se realizaron de manera automática utilizando las funcionalidades de la biblioteca ‘pmdarima.arima’. Esto permitió obtener el modelo ARIMA óptimo sin necesidad de realizar una selección manual de parámetros.

Por último, una vez ajustado el modelo ARIMA, se utilizó la función ‘predict’ del modelo ARIMA ajustado, para calcular las predicciones del periodo correspondiente al conjunto de prueba. Estas predicciones representan los valores estimados de la cotización de IAG en la serie temporal.

Las predicciones obtenidas se muestran en la tabla siguiente, donde cada fila corresponde a un índice específico en el conjunto de prueba, seguido de la predicción calculada para ese punto en la serie temporal.

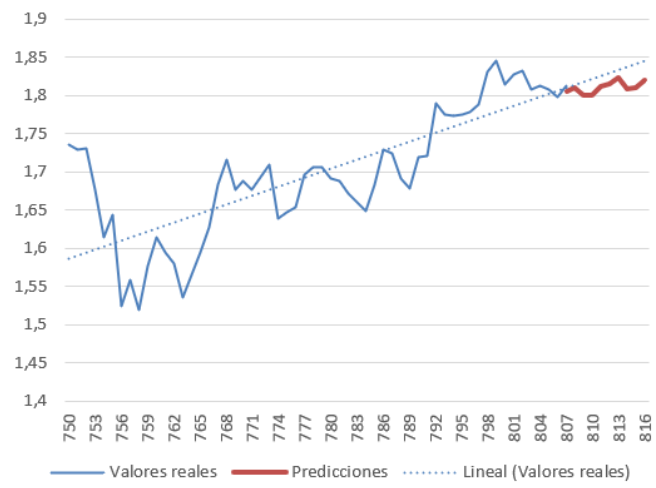
---

<sup>9</sup> Código disponible en:  
[https://colab.research.google.com/drive/1ACrJ82wF8ILdQ66DFLnBWppC7vq\\_wzH0](https://colab.research.google.com/drive/1ACrJ82wF8ILdQ66DFLnBWppC7vq_wzH0)

**Tabla 17. Predicciones ARIMA**

|     | <b>Predicciones</b> |
|-----|---------------------|
| 807 | 1,80621             |
| 808 | 1,81051             |
| 809 | 1,80107             |
| 810 | 1,80082             |
| 811 | 1,81126             |
| 812 | 1,81594             |
| 813 | 1,82341             |
| 814 | 1,80850             |

**Gráfico 29. Predicciones ARIMA**



Fuentes: Elaboraciones propias

## CAPÍTULO 5. REGRESSION DISCONTINUITY

### 5.1. Sobre este modelo

Un objetivo importante en las ciencias sociales es comprender el efecto causal de un tratamiento en los resultados de interés. Existen preguntas que interesan a la sociedad como comprender el impacto de las políticas fiscales en el crecimiento económico, el efecto de las políticas monetarias en la inflación, la relación entre el comercio internacional y el empleo, o incluso los efectos de las regulaciones del sector financiero en la estabilidad del sistema bancario. El análisis de tales efectos es relativamente sencillo cuando el tratamiento de interés se asigna al azar, ya que esto garantiza la comparabilidad de las unidades asignadas a las condiciones de tratamiento y control. Sin embargo, debido a su propia naturaleza, muchas intervenciones de interés para los científicos sociales no pueden asignarse al azar por razones éticas o prácticas [41].

En ausencia de asignación aleatoria, los diseños de investigación que permiten el estudio riguroso de intervenciones no experimentales son particularmente prometedores. Uno de estos diseños es el modelo de RD, que ha surgido como una de las estrategias no experimentales más creíbles para el análisis de efectos causales. En el diseño RD, todas las unidades tienen una puntuación, y se asigna un tratamiento a aquellas unidades cuyo valor de la puntuación supera un umbral conocido, y no se asigna a aquellas unidades cuyo valor de la puntuación está por debajo del umbral. La característica clave del diseño es que la probabilidad de recibir el tratamiento cambia abruptamente en el umbral conocido. Si las unidades no pueden “ordenarse” perfectamente alrededor de este umbral, el cambio discontinuo en esta probabilidad se puede utilizar para aprender sobre el efecto causal local del tratamiento en un resultado de interés, ya que las unidades con puntuaciones ligeramente por debajo del umbral se pueden usar como grupo de comparación para las unidades con puntuaciones ligeramente por encima de él [41].

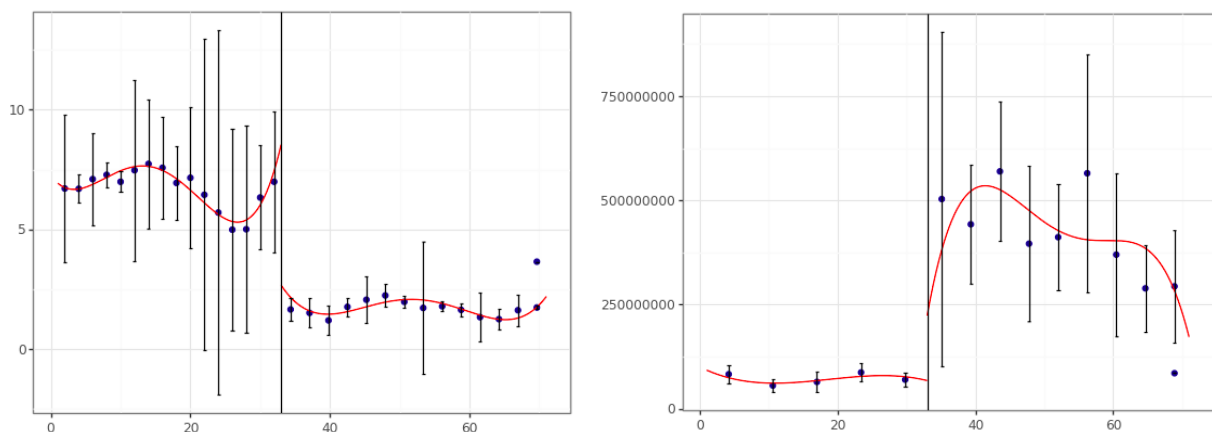
Según Cattaneo [41], el primer paso para utilizar el diseño RD en la práctica es aprender a reconocerlo. Hay **tres componentes fundamentales en el diseño RD**: una puntuación, un umbral y un tratamiento. Sin estas tres características básicas definitorias, no se puede emplear la metodología RD. Por lo tanto, un análisis RD no siempre es aplicable a los datos, a diferencia de otros métodos no experimentales como los basados en ajustes de regresión o enfoques más sofisticados de selección sobre observables, que siempre se

pueden utilizar para describir la relación condicional entre los resultados y los tratamientos. La diferencia radica en que el RD es un diseño de investigación, no una estrategia de estimación. Para estudiar efectos causales con un diseño RD, la puntuación, el tratamiento y el umbral deben existir y estar bien definidos, y la relación entre ellos debe cumplir condiciones particulares que sean objetivas y verificables. La característica definitoria clave de cualquier diseño RD canónico es que la probabilidad de asignación del tratamiento en función de la puntuación cambia de manera discontinua en el umbral, una condición que se puede probar directamente.

## 5.2. Análisis de resultados de IAG

Para llevar a cabo el análisis del impacto del Covid-19 en el precio de cierre y volumen de IAG, se han extraído los datos correspondientes a estos indicadores a lo largo de los meses<sup>10</sup>. Al realizar una transformación de la variable de fecha a números, se ha observado que el mes de febrero ha sido asignado automáticamente al número 33 en la secuencia. Este valor numérico, que corresponde al mes de interés, se ha utilizado como argumento en los diferentes procedimientos para generar las figuras resultantes que se presentan a continuación.

**Gráfico 30.** RD utilizando el argumento "ci=95" para mostrar intervalos de confianza del 95% en el gráfico.

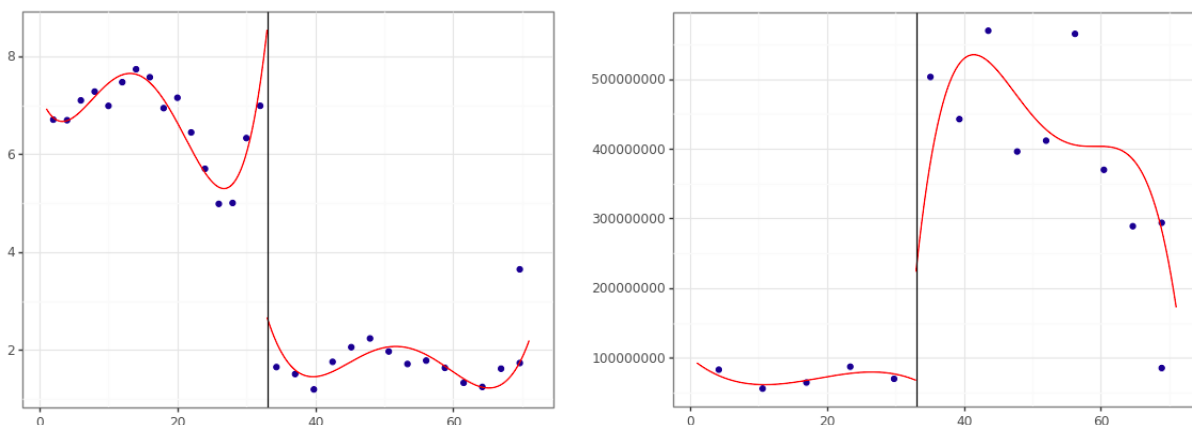


Fuente: Elaboración propia

<sup>10</sup> Código disponible en:

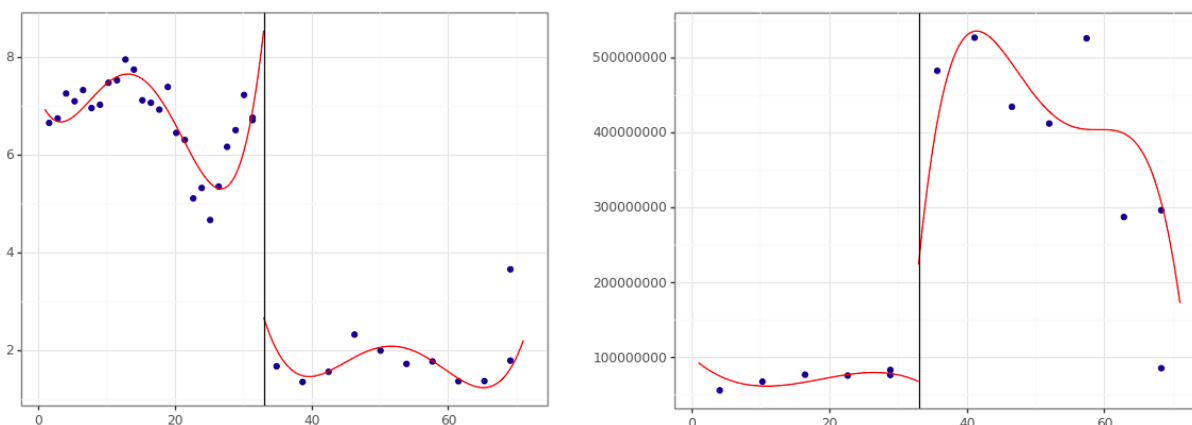
<https://colab.research.google.com/drive/1QNEwGpooVmhWVzM3bDI76QNU8Go2FXxl>

**Gráfico 31.** RD utilizando el argumento "binselect="es"", que representa la elección óptima de bin (partición de los datos) según el error cuadrático medio.



Fuente: Elaboración propia

**Gráfico 32.** RD utilizando el argumento "binselect="qsmv"" para realizar una partición de los datos utilizando el enfoque de particionado QS (particionado de la cuasi-varianza) y la elección de la varianza mimética



Fuente: Elaboración propia

Los resultados empíricos respecto al impacto del Covid-19 en el precio y volumen de IAG se resumen de la siguiente manera:

**Tabla 18.** Resultados Regression Discontinuity 'Cierre' y 'Volumen' de IAG

| <b>Precio de cierre</b> |         |       |         |           |                  |
|-------------------------|---------|-------|---------|-----------|------------------|
| Method                  | Coef.   | S.E.  | t-stat  | P >  t    | 95% CI           |
| Conventional            | - 4.781 | 1.011 | - 4.731 | 2.238e-06 | [-6.761, -2.8]   |
| Robust                  | -       | -     | - 4.227 | 2.368e-05 | [-7.299, -2.675] |
| <b>Volumen</b>          |         |       |         |           |                  |
| Method                  | Coef.   | S.E.  | t-stat  | P >  t    | 95% CI           |

|              |                      |               |       |           |                                    |
|--------------|----------------------|---------------|-------|-----------|------------------------------------|
| Conventional | <b>111740540.975</b> | 146459208.812 | 0.763 | 4.455e-01 | [-175314233.501,<br>398795315.45]  |
| Robust       | -                    | -             | 0.447 | 6.551e-01 | [-251612802.346,<br>400182978.018] |

Fuente: Elaboración propia

Los resultados obtenidos a través del análisis de RD revelaron hallazgos interesantes, como se ha observado en las figuras y en el coeficiente que se muestra en la tabla. Es evidente que **la pandemia del Covid-19 tuvo un impacto negativo en los precios de IAG**, lo cual es consistente con las expectativas dadas las circunstancias excepcionales y los desafíos que enfrentó la industria de la aviación durante ese período.

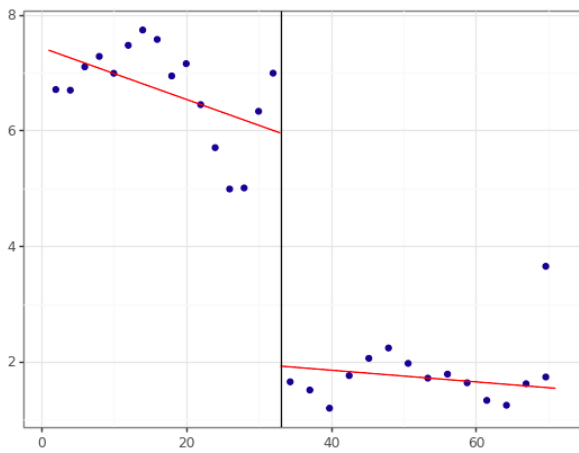
Sin embargo, es notable destacar que el análisis también reveló un patrón inverso en el volumen de transacciones de IAG. Aunque el precio experimentó una disminución significativa, **el volumen de transacciones mostró un incremento sustancial**. Este hallazgo puede resultar intrigante y puede estar relacionado con diversos factores. Una posible explicación podría ser que la volatilidad y la incertidumbre generadas por la pandemia crearon oportunidades para los inversores y especuladores en el mercado de valores. Los movimientos bruscos en los precios podrían haber atraído a aquellos que buscaban aprovechar la fluctuación y obtener ganancias a corto plazo. Esto podría explicar el aumento en el volumen de transacciones, ya que más inversores participaron activamente en la compra y venta de acciones de IAG durante ese período.

### 5.3. Comparativa de IAG con otras variables

En esta sección, se lleva a cabo un análisis comparativo entre las variables más relevantes con el objetivo de estudiar el impacto cuantitativo que la pandemia tuvo en el mercado bursátil. Se analizaron de nuevo los precios de cierre de las variables más relevantes antes y después de la pandemia, utilizando la metodología de RD. A través de la generación de gráficos, se podrán visualizar los resultados obtenidos, así como los coeficientes estimados que proporcionará el análisis de RD.

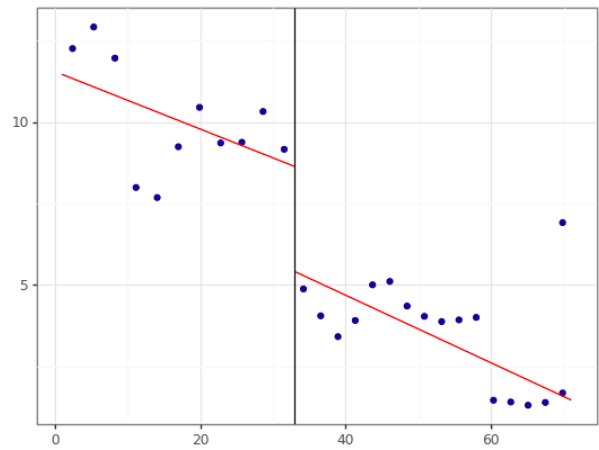
Para este análisis, se consideraron únicamente los precios de cierre de las variables más importantes que se mantuvieron disponibles tanto en el periodo anterior como en el periodo posterior a la pandemia. Se utilizó el método “binselect=es” con un polinomio de orden 1 ( $p=1$ ) para representar las tendencias mediante rectas.

**Gráfico 33. RD de IAG**



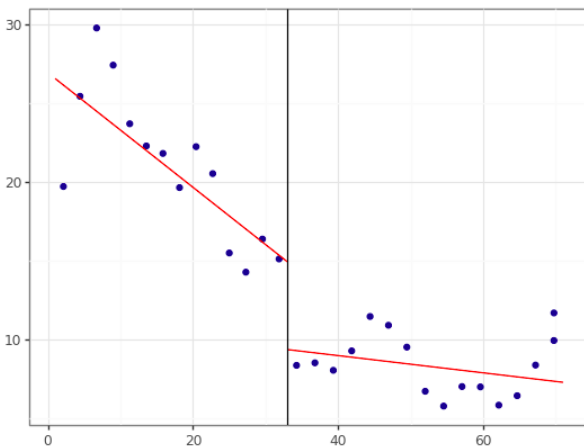
Fuente: Elaboración propia. Divisa en euros

**Gráfico 34. RD de Air France**



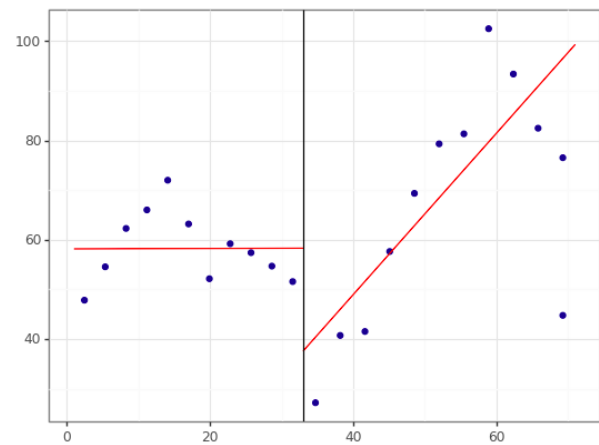
Fuente: Elaboración propia. Divisa en euros

**Gráfico 35. RD de Lufthansa**



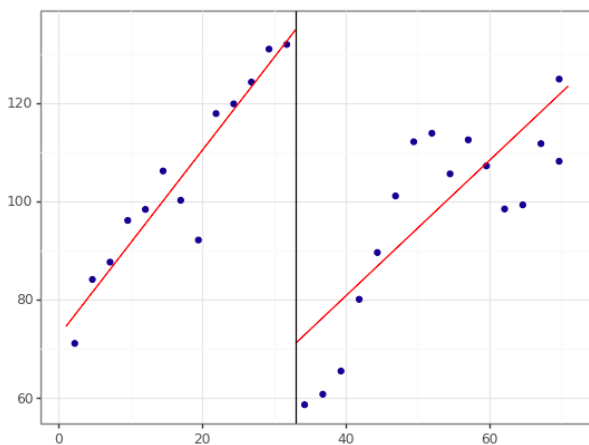
Fuente: Elaboración propia. Divisa en euros

**Gráfico 36. RD del Petróleo**



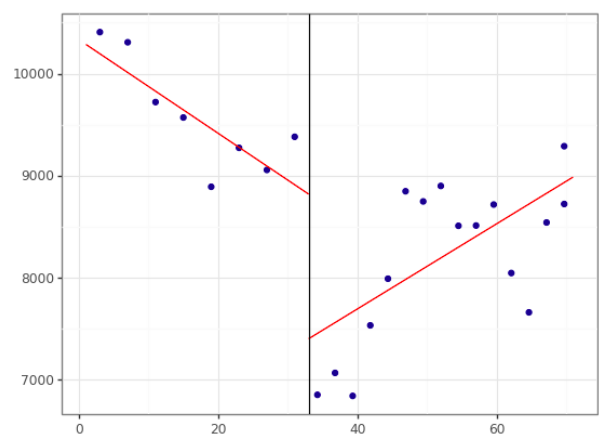
Fuente: Elaboración propia. Divisa en dólares

**Gráfico 37. RD de Airbus**



Fuente: Elaboración propia. Divisa en euros

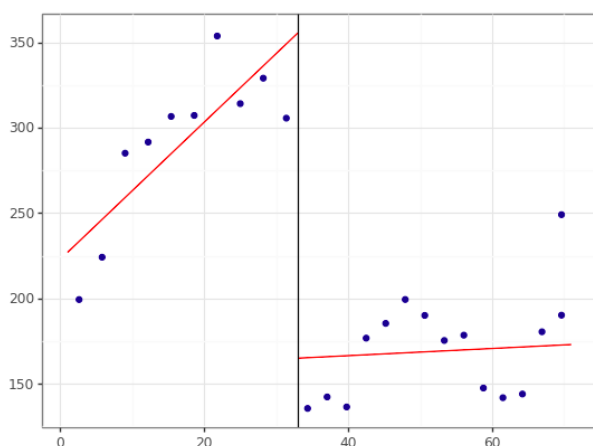
**Gráfico 38. RD del IBEX 35**



Fuente: Elaboración propia. Divisa en euros

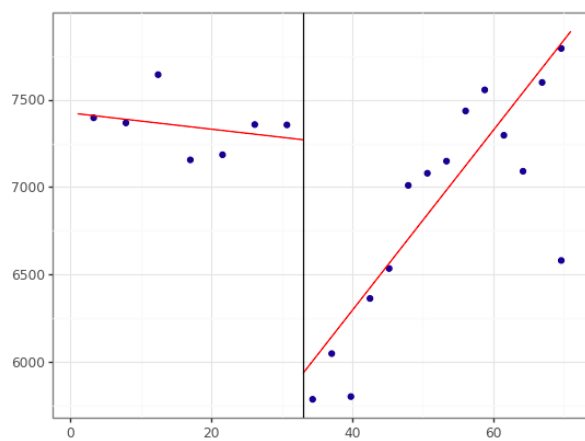


**Gráfico 39. RD de Boeing**



Fuente: Elaboración propia. Divisa en euros

**Gráfico 40. RD del FTSE 100**



Fuente: Elaboración propia. Divisa en libras

Tras analizar las figuras generadas, se identificaron patrones significativos en el comportamiento de las acciones de las diferentes variables. En particular, en cuanto a las aerolíneas, se observó un impacto negativo en el precio de las acciones de Lufthansa, con una marcada caída en sus valores. En contraste, IAG experimentó una disminución más moderada en sus precios. Es importante destacar que la escala de los valores de Lufthansa es mayor en comparación con IAG, lo cual influye en la apreciación de la magnitud de los cambios. En el caso de Air France, también se observó un impacto negativo en el precio de sus acciones después de la pandemia, similar a lo ocurrido con IAG. Sin embargo, la disminución en el precio de las acciones de Air France fue más significativa en comparación con IAG. De nuevo, es relevante señalar que, previo a la pandemia, el precio de las acciones de Air France también era superior al de IAG.

En el caso de Airbus y Boeing, dos destacadas compañías aeroespaciales, se evidenció una tendencia de recuperación gradual en el precio de sus acciones después de la pandemia. Airbus mostró una recuperación ascendente, con un incremento progresivo en sus valores a diferencia de Boeing que se mantuvo relativamente estable en sus precios. Sin embargo, es importante destacar que, después del umbral, que hace referencia a febrero de 2020, el valor de las acciones de Boeing era considerablemente más elevado que el de Airbus, lo cual influyó en la magnitud de la subida posterior a la pandemia.

En el sector del petróleo, se identificó un impacto inicial seguido de un aumento significativo en los precios de las acciones. Esto puede sugerir una recuperación posterior a los desafíos iniciales causados por la pandemia.

En el mercado bursátil español, representado por el índice IBEX 35, se observó una disminución previa a la pandemia y una posterior recuperación, lo que indica una fluctuación en el mercado durante el período analizado. Por otro lado, el índice bursátil del Reino Unido, representado por el FTSE 100, destacó por una recuperación significativa después de la pandemia, superando incluso los niveles previos a la pandemia.

Estos resultados **evidenciaron el impacto heterogéneo de la pandemia en los diversos sectores y mercados**, lo cual refleja la complejidad y dinamismo del entorno económico durante este período. Además, para obtener una mejor comprensión del impacto cuantitativo que la pandemia tuvo en la cotización de estos sectores, se calculó, como se ha mostrado en el apartado anterior, el coeficiente de la RD. Este coeficiente proporciona una medida precisa de la influencia de la pandemia en las variables analizadas y se muestra detalladamente en la tabla a continuación.

*Tabla 19. Resultados Regression Discontinuity variables más relevantes*

| <b>IAG</b>       |                 |        |         |           |                   |
|------------------|-----------------|--------|---------|-----------|-------------------|
| Method           | Coef.           | S.E.   | t-stat  | P >  t    | 95% CI            |
| Conventional     | - <b>4.781</b>  | 1.011  | - 4.731 | 2.238e-06 | [-6.761, -2.8]    |
| Robust           | -               | -      | - 4.227 | 2.368e-05 | [-7.299, -2.675]  |
| <b>Lufthansa</b> |                 |        |         |           |                   |
| Method           | Coef.           | S.E.   | t-stat  | P >  t    | 95% CI            |
| Conventional     | - <b>5.318</b>  | 2.005  | - 2.652 | 7.991e-03 | [-9.247, -1.388]  |
| Robust           | -               | -      | - 2.383 | 1.719e-02 | [-10.575, -1.029] |
| <b>Airbus</b>    |                 |        |         |           |                   |
| Method           | Coef.           | S.E.   | t-stat  | P >  t    | 95% CI            |
| Conventional     | - <b>52.167</b> | 22.765 | - 2.292 | 2.193e-02 | [-96.786, -7.549] |
| Robust           | -               | -      | - 1.822 | 6.845e-02 | [-95.969, 3.5]    |
| <b>Boeing</b>    |                 |        |         |           |                   |
| Method           | Coef.           | S.E.   | t-stat  | P >  t    | 95% CI            |

|                   |                   |         |         |           |                      |
|-------------------|-------------------|---------|---------|-----------|----------------------|
| Conventional      | - <b>90.101</b>   | 56.082  | - 1.607 | 1.081e-01 | [-200.019, 19.818]   |
| Robust            | -                 | -       | - 1.21  | 2.261e-01 | [-200.082, 47.303]   |
| <b>Air France</b> |                   |         |         |           |                      |
| Method            | Coef.             | S.E.    | t-stat  | P >  t    | 95% CI               |
| Conventional      | - <b>2.36</b>     | 1.932   | - 1.222 | 2.218e-01 | [-6.146, 1.426]      |
| Robust            | -                 | -       | - 0.882 | 3.777e-01 | [-6.245, 2.368]      |
| <b>Petróleo</b>   |                   |         |         |           |                      |
| Method            | Coef.             | S.E.    | t-stat  | P >  t    | 95% CI               |
| Conventional      | - <b>14.812</b>   | 10.203  | - 1.452 | 1.466e-01 | [-34.81, 5.186]      |
| Robust            | -                 | -       | - 1.13  | 2.585e-01 | [-35.203, 9.455]     |
| <b>IBEX 35</b>    |                   |         |         |           |                      |
| Method            | Coef.             | S.E.    | t-stat  | P >  t    | 95% CI               |
| Conventional      | - <b>1710.683</b> | 878.76  | -1.947  | 5.157e-02 | [-3433.02, 11.655]   |
| Robust            | -                 | -       | -1.611  | 1.072e-01 | [-3581.667, 350.318] |
| <b>FTSE 100</b>   |                   |         |         |           |                      |
| Method            | Coef.             | S.E.    | t-stat  | P >  t    | 95% CI               |
| Conventional      | - <b>1164.232</b> | 352.916 | - 3.299 | 9.707e-04 | [-1855.936, -472.53] |
| Robust            | -                 | -       | - 2.704 | 6.848e-03 | [-1809.69, -288.741] |

Fuente: Elaboración propia

Tras observar las figuras y las tablas con los coeficientes, es evidente que **la pandemia tuvo un impacto significativo y negativo en los precios de las acciones de todas las variables analizadas**. Los coeficientes negativos obtenidos indican una disminución en las cotizaciones después del umbral, que en este caso corresponde a febrero de 2020.

Entre las aerolíneas, se observa que tanto IAG como Lufthansa experimentaron una disminución en el precio de sus acciones. Sin embargo, se destaca que Lufthansa sufrió una caída más pronunciada, lo que podría indicar una mayor vulnerabilidad de la empresa en el contexto de la crisis sanitaria. En cuanto a Air France, se observa una disminución en la cotización similar a la de IAG, aunque de menor magnitud. Esto podría indicar una menor exposición de la compañía a los efectos de la crisis.

En el sector aeroespacial, Airbus también registró una disminución significativa en su cotización, aunque su recuperación después de la pandemia fue menos marcada en

comparación con otras variables. Por otro lado, Boeing sufrió un impacto negativo más notable, con una disminución considerable en el precio de sus acciones.

En el caso del petróleo, se aprecia una caída en la cotización, reflejando la reducción de la demanda global y las tensiones en el mercado energético durante la pandemia.

Por otro lado, en los índices bursátiles, tanto el IBEX 35 español como el FTSE 100 británico experimentaron una disminución en sus cotizaciones después de la pandemia. Esto refleja la volatilidad y las incertidumbres asociadas con la crisis sanitaria y otros factores económicos y políticos.

En resumen, los resultados obtenidos a través de la RD confirman el impacto negativo de la pandemia en los precios de las acciones y los índices bursátiles analizados. Estos hallazgos han servido para subrayar la complejidad y la magnitud de los efectos económicos de la crisis sanitaria en diferentes sectores y mercados.

## CAPÍTULO 6: CONCLUSIONES

Tras el análisis y trabajo realizados, se han obtenido las siguientes conclusiones:

- i) Luego de completar el capítulo de *feature selection*, se ha podido concluir que, a través del análisis de *mutual information* la pandemia sí tuvo un impacto en las variables más importantes relacionadas con la cotización de IAG. Estos resultados destacan la importancia de considerar el contexto y los eventos específicos al analizar y predecir los precios de las acciones en la industria de la aviación.

En este sentido, se observó una clara diferencia en la influencia de las aerolíneas y los índices de mercado durante los dos períodos. Antes de la pandemia, las aerolíneas desempeñaban un papel destacado en la cotización de IAG, mientras que los índices de mercado tenían una influencia menor. Sin embargo, después de la pandemia, el IBEX 35 se posicionó como una variable de mayor relevancia, lo que indica la importancia de considerar la confianza y las expectativas de los inversores en la recuperación económica y el crecimiento del sector de la aviación en España.

- ii) El Capítulo 3 ha proporcionado una perspectiva interesante sobre la variación de los modelos de predicción en función de los datos previos y posteriores a la pandemia. Durante el período anterior a la pandemia, el modelo de regresión basado en el kernel *Exponencial GPR* demostró un desempeño destacado, con un RMSE de 0.11976 y un  $R^2$  de 0.98. Sin embargo, en el período posterior a la pandemia, se observó que el mejor modelo sería aquel basado en el kernel *Squared Exponential GPR*, con un RMSE de 0.13052 y un  $R^2$  de 0.97.

A pesar de esta variación en los modelos seleccionados, se encontró un patrón común en ambos períodos: el modelo GPR destacó como el mejor modelo en ambas configuraciones. Sin embargo, es importante tener en cuenta su comportamiento a nivel experimental.

En general, los GPR muestran un buen desempeño para intervalos cortos, pero, a largo plazo, tienden a ajustarse a la baja si los valores previos siguen una tendencia

alcista. En este sentido, en comparación con las predicciones después de la pandemia, las predicciones realizadas antes mostraron una disminución de dos puntos en relación a los valores reales. En este caso, los precios de cierre eran considerablemente más altos, y el modelo de regresión los ajustó a la baja, explicando así la diferencia de dos puntos mencionada.

- iii) Además, se exploró la aplicación de modelos de aprendizaje automático clasificatorios para predecir las velas de IAG. Los resultados obtenidos confirmaron de nuevo que los modelos varían en función del período considerado. Durante el período previo a la pandemia, se identificó que el mejor modelo fue un *Ensemble*, con una precisión del 77.6% en la clasificación de las velas. Por otro lado, en el período posterior, el modelo que obtuvo mejores resultados fue el *Quadratic Discriminant*, con una precisión del 81.4%. Estos hallazgos resaltan la importancia de considerar el contexto de la pandemia al seleccionar y aplicar modelos de aprendizaje automático tanto en la clasificación de las velas de IAG como en la predicción de su cotización.

Se evaluaron también los modelos de series temporales y se comparó el desempeño de los métodos implementados en Python y MATLAB. Se encontró que el enfoque de 'pdmarmima' en Python brindó predicciones más precisas en comparación con el proceso manual en MATLAB. Específicamente, se determinó que el modelo  $SARIMAX(2, 1, 1)x(2, 1, 0, 5)$  fue el más efectivo en la predicción de la cotización de IAG en el período analizado.

- iv) En última instancia, con los resultados obtenidos a través de la técnica de *regression discontinuity* se confirmó de manera concluyente que la pandemia tuvo un efecto significativo en la cotización de IAG con un coeficiente de - 4.781.

Además, se observó un patrón inverso en el volumen de transacciones de IAG, lo cual resulta notable. A pesar de la disminución significativa en el precio de las acciones, el volumen de transacciones, cuyo coeficiente fue de 111740540.975, experimentó un aumento sustancial. Este hallazgo puede estar relacionado con diversos factores. Una posible explicación es que la volatilidad y la incertidumbre

generadas por la pandemia crearon oportunidades para los inversores y especuladores en el mercado de valores.

Por último, se extendió el análisis a otras variables, constatando que la pandemia tuvo un impacto negativo en los precios de cierre de todas las variables estudiadas. Sin embargo, es importante destacar que los impactos fueron heterogéneos, es decir, cada variable experimentó un efecto distinto en magnitud y dirección. Esto demuestra la importancia de considerar las particularidades de cada variable al evaluar su respuesta a eventos significativos como la pandemia.

En este trabajo también se presentaron ciertas limitaciones que podrían ser abordadas en futuras investigaciones. En el proceso de selección de la muestra inicial, se intentó incorporar datos macroeconómicos a la base de datos seleccionada de *yahoo finance*. Sin embargo, los datos que se pudieron obtener no eran diarios sino mensuales, observando así una reducción significativa en el número de observaciones disponibles. Esta limitación restringió la capacidad de considerar factores macroeconómicos relevantes en el análisis de la cotización de IAG.

Por último, se identifican varias líneas de investigación que podrían enriquecer y expandir el alcance de este trabajo. Una futura línea de investigación podría ser la de algoritmos de *Reinforcement Learning* y el uso de la Ecuación de Bellman como guía para el diseño de estos algoritmos en investigaciones futuras. Estos algoritmos podrían mejorar la toma de decisiones y la optimización de estrategias de inversión al adaptarse a las condiciones cambiantes del mercado y aprender de la retroalimentación recibida.

En el apartado de series temporales se ha empleado el modelo SARIMA para la predicción de los precios de IAG. En este sentido, se propone explorar enfoques más avanzados que permitan una mejor predicción de la volatilidad a corto y medio plazo como los modelos GARCH y ARCH.

El modelo GARCH (*Generalized Autoregressive Conditional Heteroskedasticity*) destaca por su capacidad para poder capturar y predecir la volatilidad en los precios de IAG. Al incorporar la varianza condicional en el análisis, el modelo GARCH ofrece una visión más precisa de los movimientos de precios y permite una gestión más efectiva del riesgo.

Por otro lado, el modelo ARCH (*Autoregressive Conditional Heteroskedasticity*) permite analizar la volatilidad en series temporales y pronosticar su comportamiento futuro. En este trabajo, el modelo ARCH proporcionaría una representación más precisa de la volatilidad real de los mercados. Estos enfoques más avanzados complementarían el análisis realizado con el modelo SARIMA y enriquecerían el estudio de las series temporales de IAG.

Por último, se sugiere realizar estudios econométricos y de tipo causal para examinar las relaciones entre las variables financieras y los eventos económicos globales. Estos enfoques permitirían una comprensión más profunda de los factores que influyen en el desempeño financiero de IAG, y podrían revelar patrones y relaciones causales de mayor relevancia.



## BIBLIOGRAFÍA

[1] Y. Chai, “Analysis of the Impact of COVID-19 on the Tourism Stock Market - Taking Caissa Touristic as an Example”, *Atlantis Press B.V.*, vol. 166, pp. 487-492, 2021.

[2] UNWTO, “International Tourism and Covid-19 | Tourism Dashboard”. [En línea]. Disponible en <https://www.unwto.org/tourism-data/international-tourism-and-covid-19>

[3] L. A. Gil-Alana y C. Poza, “The impact of COVID-19 on the Spanish tourism sector”, *Tourism Economics*, vol. 68, no. 3, pp. 646-653, mayo, 2022. [En línea]. Disponible en <https://journals.sagepub.com/doi/epub/10.1177/1354816620959914>

[4] “El turismo en España y en el mundo, en datos y gráficos”. [En línea]. Disponible en <https://www.epdata.es/datos/turismo-espana-mundo-datos-graficos/272>

[5] M. Skare y D. Riberio Soriano, “Explaining COVID-19 shock wave mechanism in the European service industry using convergence clubs analysis”, *Springer*, no. 16, pp. 283-307, mayo, 2021.

[6] World Health Organization, “Looking back at a year that changed the world: WHO’s response to COVID-19”, 2021.

[7] A.T. Strauss, D. Cartier, B.A. Gunning, B.J. Boyarsky, J. Snyder, D.L. Segev, M. Roush and A.B. Massie, “Impact of the COVID-19 pandemic on commercial airlines in the United States and implications for the kidney transplant community”. *Am. J. Transplant.*, vol. 20, n. 11, pp. 3123-3130, nov., 2020. [En línea]. Disponible en <https://onlinelibrary.wiley.com/doi/epdf/10.1111/ajt.16284>

[8] OACI, “El volumen total de pasajeros en 2020 se redujo un 60 % y el asalto de la COVID-19 a la movilidad internacional no se detiene”, enero, 2021. [En línea]. Disponible en: <https://www.icao.int/Newsroom/NewsDoc2021fix/COM.02.21.SP.pdf>

[9] S. E. Baker, N. Bloom, S. J. Davis, y S. Terry, “COVID-Induced Economic Uncertainty”, *National Bureau of Economic Research*, abril, 2020. [En línea]. Disponible en <https://www.nber.org/papers/w26983>

[10] W. Ding, R. Levine, C. Lin, Chen and W. Xie, “Corporate Immunity to the COVID-19 Pandemic”, *Journal of Financial Economics (JFE)*, pp. 1-11, agosto, 2020.

[11] Y. Zheng, J. Xie, H. Yu and N. Song, “Covid-19's Impact On The Airline Industry In The US Stock Market Based on the Regression Model”, *2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, Taiyuan, China, pp. 268-273, 2021.

- [12] P. Suau-Sanchez, A. Voltres-Dorta y N. Cugueró-Escofet, “An early assessment of the impact of COVID-19 on air transport: Just another crisis or the end of aviation as we know it?”, *Journal of Transport Geography*, vol. 86, pp. 1-8, junio, 2020.
- [13] International Airlines Group, “Acerca de IAG”. [En línea]. Disponible en <https://www.iairgroup.com/es-es/el-grupo/acerca-de-iag>
- [14] International Airlines Group, “Estadísticas de tráfico”. [En línea]. Disponible en <https://www.iairgroup.com/es-es/sala-de-prensa/estadisticas-de-trafico>
- [15] I. Carrillo-Hidalgo, J.I. Pulido-Fernández, J.I. Durán-Román *et al.*, “COVID-19 and tourism sector stock price in Spain: medium-term relationship through dynamic regression models”, *Financ Innov*, vol. 9, no. 8, pp. 1-24, enero, 2023.
- [16] B. Atems y J. Yimga, “Quantifying the impact of the COVID-19 pandemic on US airline stock prices”, *Journal of Air Transport Management*, vol. 97, oct., pp. 1-18, 2021.
- [17] L. Kökény, Z. Kenesei, y G. Neszveda, “Impact of COVID-19 on different business models of European airlines”, *Current Issues in Tourism*, vol. 25, no. 3, pp. 458-474, agosto, 2021.
- [18] L. P. Llorente, “Teoría de la información estadística”, *Estadística Española*, vol. 35, no. 133, pp. 195-268, 1993. [En línea]. Disponible en [https://halweb.uc3m.es/esp/Personal/personas/dpena/publications/castellano/1993EE\\_pardo\\_coment.pdf](https://halweb.uc3m.es/esp/Personal/personas/dpena/publications/castellano/1993EE_pardo_coment.pdf)
- [19] A. Zhu, “Select Features for Machine Learning Model with Mutual Information”, *Towards Data Science*, junio, 2021. [En línea]. Disponible en <https://towardsdatascience.com/select-features-for-machine-learning-model-with-mutual-information-534fe387d5c8>
- [20] J. Brownlee, “Information Gain and Mutual Information for Machine Learning”, *Machine Learning Mastery*, oct., 2019. [En línea]. Disponible en <https://machinelearningmastery.com/information-gain-and-mutual-information/>
- [21] M. Christopher, *Pattern Recognition and Machine Learning*, ed. 1, New York: Springer New York, 2016, pp. 48-58.
- [22] “Comparison of F-test and mutual information”, *scikit-learn*. [En línea]. Disponible en [https://scikit-learn.org/stable/auto\\_examples/feature\\_selection/plot\\_f\\_test\\_vs\\_mi.html#sphx-glr-auto-examples-feature-selection-plot-f-test-vs-mi-py](https://scikit-learn.org/stable/auto_examples/feature_selection/plot_f_test_vs_mi.html#sphx-glr-auto-examples-feature-selection-plot-f-test-vs-mi-py)

[23] H. Chamizo, “Cómo las aerolíneas pueden sufrir más con el descontrol del precio del petróleo”, *La Información*, 2022. [En línea]. Disponible en <https://www.lainformacion.com/mercados-y-bolsas/como-las-aerolineas-pueden-sufrir-mas-con-el-descontrol-del-precio-del-petroleo/2861657/>

[24] BBVA, “Factores que influyen en el precio de las acciones”. [En línea]. Disponible en [Microsoft PowerPoint - Formacion bbva.es factores influyen acciones https://www.bbva.es/estaticos/mult/Ayudas\\_factores\\_acciones.pdf\\_tcm924-528182.pdf#:~:text=econ%C3%B3mica%20del%20mercado%20y%20del%20sector%3A%20las%20acciones,unimpacto%20en%20la%20cotizaci%C3%B3n%20de%20las%20empresas%20afectadas.](https://www.bbva.es/estaticos/mult/Ayudas_factores_acciones.pdf_tcm924-528182.pdf#:~:text=econ%C3%B3mica%20del%20mercado%20y%20del%20sector%3A%20las%20acciones,unimpacto%20en%20la%20cotizaci%C3%B3n%20de%20las%20empresas%20afectadas.)

[25] OECD, “COVID-19 y la industria aérea: impacto y respuestas políticas”, oct., 2020. [En línea]. Disponible en [https://read.oecd-ilibrary.org/view/?ref=1060\\_1060081-7ag4dym0ob&title=COVID-19-y-la-industria-aerea](https://read.oecd-ilibrary.org/view/?ref=1060_1060081-7ag4dym0ob&title=COVID-19-y-la-industria-aerea)

[26] X. Vives, “Competencia en aerolíneas”, *La Vanguardia*, mayo, 2018. [En línea]. Disponible en: <https://www.lavanguardia.com/opinion/20180503/443196633114/competencia-en-aerolineas.html>

[27] R.E. Whaley, “Trading Volatility: At What Cost?”, *The Journal of Portfolio Management*, mayo, 2013.

[28] Efe., “La crisis de Boeing puede tener consecuencias que van desde una caída del PIB hasta a una crisis en el sector auxiliar”, *El Economista*, 2019. [En línea]. Disponible en <https://www.eleconomista.es/empresas-finanzas/noticias/10269152/12/19/Que-esperar-de-la-crisis-de-Boeing-bajada-de-PIB-y-crisis-en-el-sector-auxiliar.html>

[29] IG, “Crypto Trading | What is Cryptocurrency Trading? | IG UK”. [En línea]. Disponible en <https://www.ig.com/es/indices/que-es-el-vix-y-como-operar>

[30] M. L. Escribano, “El precio del queroseno golpea a las aerolíneas antes de la temporada alta”, *La Información*, mayo, 2022. [En línea]. Disponible en: <https://www.lainformacion.com/mercados-y-bolsas/precio-queroseno-golpea-aerolinea-temporada-alta/2866015/>

[31] H. Un Cho, Y. Nam, E. Ji Choi, Y. Jae Choi, H. Kim, S. Bae, J. Woo Moon, “Comparative analysis of the optimized ANN, SVM, and tree ensemble models using Bayesian optimization for predicting GSHP COP”, *Journal of Building Engineering*, vol. 44, dic., 2021. [En línea]. Disponible en <https://www.sciencedirect.com/science/article/abs/pii/S2352710221012699>

- [32] J. Snoek, H. Larochelle y R.P. Adams, “Practical Bayesian optimization of machine learning algorithms”. *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012, NIPS 2012 (Advances in Neural Information Processing Systems; Vol. 4)*, 2012, pp. 2951-2959.
- [33] M.J. Pring, *Technical analysis explained: The successful investor's guide to spotting investment trends and turning points*, vol.4, New York: McGrawHill, 1991.
- [34] J. Wang, “An Intuitive Tutorial to Gaussian Processes Regression”, 2020. [En línea]. Disponible en <https://arxiv.org/pdf/2009.10862.pdf>
- [35] C. Thornton, F. Hutter, H.H. Hoos, K. Leyton-Brown, “Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms”. *Proc. Knowl. Disc. Data Mining*, 2013, 847–855. [En línea]. Disponible en <https://arxiv.org/pdf/1208.3719.pdf>
- [36] I.T. Jolliffe, *Principal Component Analysis*, ed. 2, USA: Springer, 2022.
- [37] S.K. Mathew, and Y. Zhang, “Acoustic-Based Engine Fault Diagnosis Using WPT, PCA and Bayesian Optimization”, *Applied Sciences*, vol. 10, no. 19, 2020. [En línea]. Disponible en <https://www.mdpi.com/2076-3417/10/19/6890>
- [38] E. Lutins, “Ensemble Methods in Machine Learning: What are They and Why Use Them?”, *Medium*, junio, 2018. [En línea]. Disponible en: <https://towardsdatascience.com/ensemble-methods-in-machine-learning-what-are-they-and-why-use-them-68ec3f9fef5f>
- [39] L. Gonzalez, “Matriz de Confusión”, *Aprende IA*, agosto, 2020, [En línea]. Disponible en: <https://aprendeia.com/matriz-de-confusion-machine-learning/>
- [40] “Linear vs. Quadratic Discriminant Analysis – Comparison of Algorithms”, 2021. [En línea]. Disponible en: <https://thatdatatho.com/linear-vs-quadratic-discriminant-analysis/>
- [41] M. D. Cattaneo, N. Idrobo, y R. Titunik, “A Practical Introduction to Regression Discontinuity Designs”, *Cambridge University Press*, nov., 2019.