

Invierte con Inteligencia Artificial:

Metodología para la estimación de precios

En este documento se presenta la metodología utilizada por parte de Eye Performance, AG, ("Eye Performance") o denominado comercialmente "Node.AI" (www.nodetech.ai/), proveedor de la información que Open Bank, S.A. proporciona para su servicio de análisis financiero conforme a la normativa MIFID y Ley de Mercado de Valores. Para obtener precios objetivos, se toma como referencia compañías que componen dos índices bursátiles importantes: el S&P 500 y el STOXX Europe 600. Cada compañía tiene asignados modelos de Machine Learning propios que permiten calcular su precio objetivo para 4 periodos distintos (1, 3, 6 y 12 meses). Estos modelos se entrenan, se ajustan y se validan para lograr la mayor precisión en los cálculos a través de un proceso que consta de 8 pasos, enumerados a continuación y detallados a lo largo del documento:

1. El proceso comienza con la selección del universo de acciones, en este caso, las que componen los índices mencionados anteriormente.
2. En base a estudios financieros, se seleccionan las variables macroeconómicas y de mercado más relevantes que afectan a cada una de ellas y que formarán parte de los modelos.
3. A través de APIs procedentes de proveedores homologados, se extraen los datos de cada una de las variables.
4. Estas variables se analizan con el objetivo de excluir aquellas que sean poco relevantes y que puedan causar ruido en los modelos. Esto se realiza a través del filtrado por correlación y por multicolinealidad, lo que permite constituir los modelos con las variables definitivas. En base al output de cada uno, se crea y se entrena un metamodelo adicional a los anteriores.
5. Se escoge el modelo (o combinación de modelos en base al metamodelo) que obtiene predicciones con mayor precisión a nivel individual (es decir, con menor error en testeo), que podrá variar posteriormente si, al introducir nuevos datos, se concluye que existe otro más preciso.
6. En el siguiente paso, el modelo se valida utilizando diferentes métricas de evaluación, como la precisión, el error cuadrático medio (MSE), el coeficiente de determinación (R^2) y la curva ROC. El objetivo es determinar si el modelo puede extrapolar o generalizar a nuevos datos. Si el rendimiento no es el esperado, es necesario un ajuste de hiperparámetros.
7. Este ajuste implica la exploración de diferentes combinaciones de valores para determinar cuál es la óptima que maximiza el rendimiento del modelo. Esto se hace utilizando la optimización bayesiana, que emplea modelos probabilísticos para determinar la combinación óptima de valores de hiperparámetros.
8. Como último paso de este proceso, se utiliza SHAP, un método que ayuda a aumentar la interpretabilidad ya que permite ver de qué manera afectan las diferentes variables (otras compañías, cambios de moneda, indicadores macroeconómicos, etc) a las estimaciones generadas de cada modelo.

Índice

1. INTELIGENCIA ARTIFICIAL (IA) Y MACHINE LEARNING (ML).....	3
2. DESCRIPCIÓN DEL PROCESO.....	3
2.1. Universo de acciones.....	3
2.2. Construcción de variables.....	3
2.3. Adquisición de variables.....	4
2.4. Reducción de variables.....	4
2.5. Selección del modelo.....	4
2.6. Evaluación del rendimiento del modelo.....	5
2.7. Ajuste de hiperparámetros.....	5
2.8. Interpretabilidad de los modelos.....	6
3. PUESTA EN PRODUCCIÓN.....	6
4. CONCLUSIÓN.....	7
5. CONTROL DE CAMBIOS.....	7
6. Anexos.....	7

1. INTELIGENCIA ARTIFICIAL (IA) Y MACHINE LEARNING (ML)

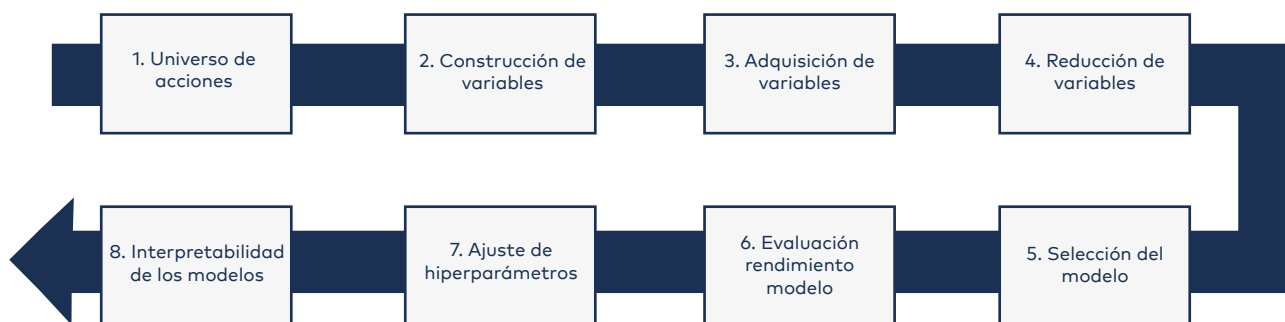
El Machine Learning (ML) es una rama de la Inteligencia Artificial (IA) que desarrolla algoritmos (también llamados modelos) que analizan patrones sobre los datos que reciben como entrada (input), para establecer una relación entre los mismos y, finalmente, la salida deseada (output).

En entornos ligados a los mercados y, más concretamente, en la herramienta "Invierte con Inteligencia Artificial", los inputs son datos o variables de múltiples orígenes: rendimientos del activo financiero y sus tendencias, indicadores técnicos asociados, información sobre las compañías relacionadas con el mercado, datos macroeconómicos y sus evoluciones, etc. Estos datos se utilizan para obtener un output, que en este caso es el dato de rendimiento objetivo para cada acción en distintos horizontes temporales: 1 mes, 3 meses, 6 meses y 12 meses. De esta manera, el ML se usa para aplicar relaciones/patrones entre los datos de entrada y salida, que los humanos no somos capaces de analizar con la misma eficiencia dada la diversidad y cantidad de información analizada.

Una vez definidos los modelos, estos se reentrenan con los nuevos datos recibidos. Es decir, se recalculan teniendo en cuenta la nueva información disponible. De este modo se van ajustando a las nuevas circunstancias de la compañía, del mercado o de la economía en general.

2. DESCRIPCIÓN DEL PROCESO

Tal y como se puede observar en la siguiente imagen, el proceso para obtener las estimaciones consta de 8 pasos.



2.1. Universo de acciones

Antes de comenzar, se debe fijar el universo de acciones sobre las que se van a generar las predicciones. En este caso, este universo está compuesto por las acciones de las compañías que cotizan en el S&P 500 y en el STOXX Europe 600.

Sin embargo, es importante destacar que es necesario una cantidad de datos mínima para la construcción de modelos predictivos robustos y que el universo de acciones está limitado a compañías que lleven cotizando al menos 5 años.

2.2. Construcción de variables

Definido el universo invertible, se construyen las variables, más relevantes. Estas conformarán el input de los modelos de ML para generar la predicción de los precios de cada acción que compone el universo.

Para ello, en una primera fase, se seleccionan manualmente estas variables de acuerdo al conocimiento experto de analistas financieros. Más tarde, se realiza un proceso de generación de nuevas variables (o features) mediante métodos cuantitativos, en lo que se conoce como la fase de feature engineering. En esta, se añaden variables que tienen en cuenta factores como la tendencia, el momentum, la volatilidad de la serie, así como métricas de compañías correlacionadas con la compañía de la que queremos calcular la predicción.

2.3. Adquisición de variables

La adquisición de estas variables se realiza mediante APIs procedentes de proveedores homologados. Los datos son extraídos, transformados y cargados en un sistema de almacenamiento, permitiendo actualizar la información de manera eficiente y efectiva. Además, se establece un sistema de monitoreo para detectar y resolver problemas.

2.4. Reducción de variables

De las más de 100.000 variables que se manejan en el entrenamiento de los modelos (ver ejemplos en el Anexo 1), se realiza un lavado para eliminar aquellas que presentan una elevada correlación o multicolinealidad. También se excluyen aquellas que son poco relevantes y que pueden introducir ruido en los modelos, lo que permite constituir los 5 modelos independientes con las variables definitivas.

A nivel de variables, existen dos tipos:

- Filtrado por correlación, por el cual se eliminan aquellas variables que muestren una alta correlación con otras.
- Filtrado por multicolinealidad, por el cual se eliminan aquellas variables que muestren una correlación alta con otras dos o más. Se entiende que la interpretabilidad que aporta esa variable puede derivarse de las otras dos (o N) variables.

A nivel de modelo, se realiza una eliminación recurrente de variables (recursive feature elimination) que busca encontrar el subconjunto de ellas que genere mayor rendimiento.

Fruto del proceso anterior, se obtiene un modelo que es más sencillo y con mejor rendimiento. El análisis predictivo es individualizado para cada acción. Se utilizan un total de 5 modelos de ML que analizan cada una de ellas de manera independiente. Una vez los 5 modelos han analizado la acción en particular y generado una predicción para el mismo, y en base a su output, se entrena un metamodelo con las predicciones derivadas de estos modelos. Este metamodelo no aprende de los datos históricos como los otros modelos, si no en base a los 5 modelos anteriores. Esto le permite, incluso, aprender de modelos con menor poder predictivo.

La práctica ha demostrado que obtiene mejores resultados. Por ejemplo, si el Modelo 1 suele dar una predicción contraria a los otros modelos, el metamodelo entenderá esta contrariedad y será capaz de utilizarla en su análisis y predicción. Este metamodelo se entrena en una fase posterior al entrenamiento de los 5 modelos anteriores determinando la importancia que designa a cada una de estas 5 partes en su output final.

2.5. Selección del modelo

Finalmente, se escoge el modelo (o combinación de modelos en base al metamodelo) que obtiene predicciones con mayor precisión a nivel individual (es decir, con menor error en testeo). Es importante destacar que la selección del modelo más preciso para cada acción y para cada uno de los horizontes temporales (1, 3, 6 y 12 meses) no es estática.

Al introducir en los modelos nuevos datos históricos y otros datos añadidos, es posible que haya acciones cuyo mejor modelo ya no sea el anteriormente escogido. En ese caso, se cambia el modelo aplicado para dicha acción.

2.6. Evaluación del rendimiento del modelo

Para evaluar el rendimiento del modelo, los datos se dividen en dos tipos: conjuntos de entrenamiento, que se utilizan para ajustar los parámetros del modelo, y conjuntos prueba, que se utilizan para evaluar el rendimiento del modelo. La división se realiza para evitar el sobreajuste que ocurre cuando el modelo se ajusta demasiado a los datos de entrenamiento y no puede generalizar de forma correcta nuevos datos.

El tamaño de los subconjuntos depende del tamaño total del conjunto de datos y de la complejidad del modelo. En general, se utiliza una proporción del 70-80 % de los datos para entrenamiento y del 20-30 % para prueba, asegurando siempre que ambos subconjuntos de datos tienen un volumen suficiente para el correcto entrenamiento de los modelos.

El conjunto de datos de prueba se utiliza para evaluar el rendimiento del modelo después de que se haya entrenado con el conjunto de datos de entrenamiento. Sin embargo, es importante tener en cuenta que el conjunto de datos de prueba no se utiliza durante el entrenamiento, si no que se utiliza para evaluar el rendimiento final del modelo.

La validación se puede realizar utilizando diferentes métricas de evaluación, como la precisión, el error cuadrático medio (MSE), el coeficiente de determinación (R^2) y la curva ROC. La elección de la métrica de evaluación depende del tipo de problema que se esté resolviendo. Por ejemplo, para un problema de regresión como éste, el MSE y el R^2 son las métricas más comunes.

El objetivo de la validación del modelo es determinar si el modelo puede extrapolar o generalizar a nuevos datos. Si el rendimiento del modelo en el conjunto de prueba es bueno, entonces se puede concluir que el modelo es capaz de generalizar bien. Si el rendimiento del modelo es pobre, entonces es necesario ajustar los parámetros del modelo o seleccionar un modelo diferente para resolver el problema.

2.7. Ajuste de hiperparámetros

El ajuste de hiperparámetros es un proceso importante en el aprendizaje automático que implica la selección de los valores óptimos para los hiperparámetros de un modelo.

Los hiperparámetros, diferentes de los parámetros del modelo, se ajustan automáticamente durante el entrenamiento del modelo. Son valores fijos que se establecen antes del entrenamiento del modelo y afectan a su rendimiento. Ejemplos de hiperparámetros comunes podría ser: la tasa de aprendizaje, el número de capas en una red neuronal, el tamaño del lote o el número de árboles en un bosque aleatorio.

Este ajuste implica la exploración de diferentes combinaciones de valores de hiperparámetros para determinar la combinación óptima que maximiza el rendimiento del modelo. Esto se hace utilizando técnicas de búsqueda de hiperparámetros. En este caso, se utiliza la optimización bayesiana, que utiliza modelos probabilísticos para determinar la combinación óptima de valores de hiperparámetros.

2.8. Interpretabilidad de los modelos

La falta de interpretabilidad en los modelos de ML puede llevar a decisiones sesgadas o erróneas y puede ser difícil de justificar. Por ello, existen varios métodos para aumentar la interpretabilidad de los modelos de ML, como el uso de algoritmos de explicación, la visualización de los datos utilizados para entrenar el modelo y la identificación de las características más importantes utilizadas por él.

Sin embargo, es importante tener en cuenta que el aumento de la interpretabilidad a menudo viene con un coste en términos de precisión del modelo. Los modelos de ML más complejos y precisos pueden ser menos interpretables, mientras que los modelos más simples y fáciles de entender pueden ser menos precisos. Por lo tanto, es importante equilibrar la precisión y la interpretabilidad al seleccionar un modelo de ML para un problema en particular.

En este caso, se utilizan los valores de SHAP (Shapley Additive exPlanations) para saber de qué manera afectan las diferentes variables (otras compañías, cambios de moneda, indicadores macroeconómicos, etc) a las predicciones generadas de cada modelo.

La evaluación de los resultados de un modelo de ML consiste en un análisis completo de los resultados del mismo. En este paso, se analizan los resultados de la evaluación del modelo y se determina si el modelo está proporcionando resultados precisos y útiles.

El análisis de los resultados incluye los siguientes pasos:

- **Análisis de métricas:** se analizan las métricas seleccionadas para evaluar el rendimiento del modelo. Si el modelo está proporcionando resultados precisos y útiles, las métricas deberían estar cerca de los valores ideales. Si las métricas no son aceptables, se deben revisar los pasos anteriores y considerar ajustar los parámetros del modelo.
- **Análisis de errores:** se analizan los errores cometidos por el modelo. Se pueden identificar patrones en los errores y utilizar esta información para ajustar el modelo y mejorar su rendimiento.
- **Análisis de importancia de características:** se analiza la importancia de las características utilizadas en el modelo. Esto puede ayudar a identificar las características más importantes para el problema en cuestión y mejorar el rendimiento del modelo.
- **Pruebas adicionales:** si los resultados no son aceptables, se pueden realizar pruebas adicionales para determinar si el modelo es adecuado para el problema en cuestión. Por ejemplo, se pueden realizar pruebas de sensibilidad para evaluar cómo cambia el rendimiento del modelo en diferentes escenarios.
- **Toma de decisiones:** en función de los resultados de la evaluación, se pueden tomar decisiones sobre el uso del modelo. Si el modelo está proporcionando resultados precisos y útiles, se puede utilizar para tomar decisiones informadas. Si el modelo no está funcionando correctamente, se deben realizar ajustes para mejorar su rendimiento o considerar el uso de un modelo diferente.

3. PUESTA EN PRODUCCIÓN

Es el paso final del proceso de creación de un modelo predictivo y solo puede ser acometido cuando la validación de los requisitos técnicos del modelo arroja un resultado positivo.

Este último paso implica, por una parte, la propia instalación o integración del modelo. Se realiza en el entorno en el cual se desarrollan los procesos autónomos que permiten la generación de nuevas predicciones diariamente. Por otro lado, ofrecer a los usuarios estos datos generados diariamente mediante una interfaz de comunicación, en el caso de nuestra herramienta, a través de una API que se ajusta a los límites de la arquitectura REST (Representational State Transfer). Por último, instalar un mecanismo de control de los modelos, realizando pruebas recurrentes y monitorizando cada uno de ellos.

4. CONCLUSIÓN

Como hemos podido leer a lo largo del documento, el proceso de estimación de precios implica la selección, adquisición, y transformación de variables, seguido del entrenamiento y optimización de modelos de Machine Learning. Es un proceso robusto, donde los modelos se validan y mejoran constantemente. Esto permite una mayor precisión y eficacia en la obtención de precios objetivos, tratando de aportar más información a inversores para que puedan adelantarse a los mercados, y a comprar y vender según sus intereses.

5. CONTROL DE CAMBIOS

Este documento se revisa cada vez que aplica un cambio o, en su defecto, con carácter anual.

Versión	Responsable documento	Fecha cambio
1	Departamento de Inversiones de Openbank	Junio 2024
2	Departamento de Inversiones de Openbank	Julio 2024

Versión	Descripción del cambio
1	Se enumeran ejemplos de las variables que conforman los modelos
2	Se actualiza la información legal del documento

6. Anexos

Anexo I: Ejemplos variables

Clasificación de variables:

- Datos de mercado (derivados de precio y comúnmente utilizadas por el análisis técnico del precio de una acción)
- Datos fundamentales
- Datos identificativos
- Datos macroeconómicos (indicadores globales)
- Datos técnicos

Ejemplo de variables

- US Economic surprise 160 days back
(sorpresas económicas en Estados Unidos en los últimos 160 días)
- Shanghai shipping index moving average
(Media móvil del índice de tráfico del puerto de Shanghai)
- Volume moving average % difference T120 T60
(diferencia de medias móviles de volumen negociado en periodos de 120 días frente a 60 días)
- Ch credit impulse % difference 40 days
(diferencia % en 40 días del impulso del crédito en China)
- Global economic surprise % difference 120 days 60 days
(diferencia % entre sorpresas en la economía global en los últimos 120 días y 60 días)
- Global economic surprise moving average
(media móvil de sorpresas en la economía global)
- US used vehicle index % difference 40 days
(diferencias % en los últimos 40 días en en índice de ventas de vehículos usados en USA)
- USD CNY 240 days back
(tipo de cambio dólar yen hace 240 días)
- Bond10yr 2yr spread moving average difference
(media móvil de la diferencia de tipos de interés entre el bono a 10 años y a 2 años)
- Variables en relación al precio de competidores o compañías que afecten directamente al precio de la acción

AVISO LEGAL

Herramienta "Invierte con Inteligencia Artificial"

La herramienta "Invierte con Inteligencia Artificial" proporcionada por Open Bank, S.A. ("Openbank") a través de la información recibida de Eye Performance, AG ("Eye Performance" o también denominado comercialmente "Node.AI"), tiene la consideración de "Informe de Inversión" o "Análisis Financiero" a efectos de la normativa MiFID y la Ley del Mercado de Valores. Openbank (entidad perteneciente a Grupo Santander) es una entidad inscrita en España, autorizada y regulada por la Comisión Nacional del Mercado de Valores y el Banco de España con el código 0073.

Eye Performance una empresa tecnológica, que no se encuentra regulada ni supervisada por ningún organismo supervisor. Asimismo, en la actualidad su capital social no está participado por por ninguna entidad perteneciente a Grupo Santander.

La simple puesta a disposición de la información sobre precios estimados a un cliente o un posible cliente, a través de la herramienta "Invierte con Inteligencia Artificial", se facilita exclusivamente a efectos informativos y no implica la prestación de un servicio de asesoramiento en materia de inversión por lo que en ninguna circunstancia deberá usarse ni considerarse como una oferta de venta ni como una solicitud de una oferta de compra ni como una recomendación.

Los clientes usuarios de la herramienta "Invierte con Inteligencia Artificial" deberán adoptar sus propias decisiones de inversión, siendo conscientes de que los valores o instrumentos mencionados en este documento pueden no ser adecuados para sus objetivos de inversión o situación financiera. Asimismo, deberán tener en cuenta que la evolución pasada de los valores e instrumentos, así como las rentabilidades pasadas a de las inversiones no garantizan la evolución o rentabilidades futuras. Toda la información emitida por Openbank está a disposición de los clientes en la página web, a través del correo electrónico: inversiones@openbank.es

Los datos, estimaciones, previsiones y recomendaciones contenidas en la herramienta "Invierte con Inteligencia Artificial" se calculan, actualizan y publican diariamente, sin previo aviso, en el área privada de Openbank y, con carácter mensual, en la parte pública de la web (<https://www.openbank.es/inversiones/inteligencia-artificial-para-invertir-bolsa>), en ambos casos para cuatro horizontes temporales: uno, tres, seis y doce meses. Asimismo, los términos "Comprar" o "Vender" tienen un significado operativo, con objeto de proporcionar a los clientes el acceso directo al flujo de compra o venta de la acción, en caso de tener la acción en cartera.

Información importante

Openbank no recibe, directa o indirectamente, ningún tipo de compensación por la promoción del precio estimado de cada acción. El pago que realiza Openbank a Eye Performance, por el cálculo de los precios estimados, se realiza con cargo a los recursos propios del Banco, sin repercutir este coste a los clientes. El sistema retributivo de Eye Performance no está vinculado, directa o indirectamente, al sentido del precio estimado mostrado, ni a las contrataciones del servicio de bróker de Openbank.

La información proporcionada en la herramienta "Invierte con Inteligencia Artificial" para cada uno de las acciones, no ha sido comunicada al emisor directa o indirectamente.

Openbank no ha suscrito contratos de liquidez o de creación de mercado con ninguno de los emisores o compañías analizadas en la herramienta "Invierte con Inteligencia Artificial". La información proporcionada en la herramienta "Invierte con Inteligencia Artificial" ha sido obtenida de fuentes consideradas como fiables, si bien Openbank no garantiza la seguridad de las mismas.

Openbank.es 91 177 33 10 fb.com/openbank.es 2 Open Bank, S.A. Domicilio social: Plaza de Santa Bárbara 2, 28004, Madrid – Registro Mercantil de Madrid, folio 202, tomo 5308, hoja M – 87030 – N.I.F. A28021079 @openbank_es Open Bank, S.A. no posee una posición larga o corta neta sobre los emisores de las acciones listadas en la herramienta "Invierte con Inteligencia Artificial" que sobrepase el umbral del 0.5% del capital social total emitido por el emisor, calculado de conformidad con el artículo 3 del Reglamento (UE) nº 236/2012 y con los capítulos III y IV del Reglamento Delegado (EU) nº 918/2012 de la Comisión Europea, de 5 de julio de 2012. Asimismo, ninguna de las entidades emisoras de las acciones listadas en la herramienta "Invierte con Inteligencia Artificial", poseen una participación igual o superior al 5% del capital en Open Bank, S.A. En los últimos 12 meses, Open Bank, S.A., no ha participado como asegurador y/o colocador y/ director/codirector en ofertas públicas de valores o emisiones, ni tampoco ha percibido por tanto remuneraciones basadas en ello, ni tampoco ha recibido compensaciones por la prestación de servicios de inversión o auxiliares de ningún emisor. Grupo Santander puede participar como asegurador y/o colocador y/ director/codirector en ofertas públicas de valores o emisiones, y percibir por tanto remuneraciones basadas en ello, así como compensaciones por la prestación de servicios de inversión o auxiliares de los emisores de las acciones listadas en la herramienta "Invierte con Inteligencia Artificial", todo ello de acuerdo con la [Política de conflictos de interés](#) de Grupo Santander.

La información proporcionada por Openbank en la herramienta "Invierte con Inteligencia Artificial", ha sido elaborada y divulgada de acuerdo con las políticas de gestión de conflictos de interés del Grupo Santander. En relación con la producción y divulgación de análisis de inversión o recomendación de inversión, Grupo Santander tiene [normas internas de conducta](#) que incluyen, entre otros aspectos, procedimientos para evitar conflictos de interés, incluidos procedimientos de murallas chinas y, en caso oportuno, la posibilidad de establecer restricciones específicas en la actividad de análisis. Cualquier sociedad del Grupo Santander, directores, comerciales, traders u otros profesionales pueden proporcionar comentarios de mercado o estrategias de negociación, o análisis de inversión, bien sea de forma oral o escrita, que reflejen opiniones o recomendaciones contrarias o distintas a las expresadas la información emitida y contenida en la herramienta "Invierte con Inteligencia Artificial". Asimismo, cualquiera de las entidades de inversión y negociación de Grupo Santander pueden adoptar decisiones de inversión que no sean coherentes con las recomendaciones expresadas en la herramienta "Invierte con Inteligencia Artificial". Ni Openbank, ni el Grupo Santander se hacen responsables de las pérdidas directas o indirectas que se deban al uso de la información emitida o contenida de la herramienta "Invierte con Inteligencia Artificial". La información emitida, así como el contenido de la herramienta de "Invierte con Inteligencia Artificial" no podrá ser reproducido, distribuido ni publicado por ningún receptor del mismo por ningún motivo.

Rentabilidades pasadas no presuponen en ningún caso rentabilidades futuras. Toda inversión en renta variable está sujeta a riesgos de pérdida del principal invertido. Los factores que pueden afectar al precio y a su cotización en mercado son, entre otros, las expectativas sobre el beneficio futuro de la sociedad y su tasa de crecimiento, eventos societarios (pagos de dividendos y cambios en la política de retribución vía dividendos, ofertas públicas de adquisición o venta, ampliaciones de capital) cambios en la calificación crediticia de la compañía suspensiones de cotización en bolsa o salidas a bolsa de otras compañías la cotización de otras acciones correlacionadas, noticias de tipo económico, político y de toda índole que afectan a los mercados financieros los tipos de interés, la economía en general, la confianza de los inversores, etc Asimismo, la fluctuación de la divisa puede incrementar o disminuir el rendimiento de aquellos valores cotizados en divisas distintas al euro. Para más información sobre las características y los riesgos generales de este tipo de instrumentos financieros, aconsejamos consultar el documento "Folleto MiFID" que le ha proporcionado el Banco y que tiene a su disposición en la página web [Información legal - Información para la contratación y documentación de interés | Información MIFID](#)