Programa ejecutivo de inteligencia artificial y Deep Learning

Resumen ejecutivo

Adrián Grasa

Pedro Gómez

Raúl Estevas

TABLA DE CONTENIDO

PROBLEMA DE NEGOCIO 1

¿Por qué ARP - Trader Cognitivo? 1
Caso de uso y propuesta de valor 1

MODELO DE NEGOCIO 4

Misión, visión y valores 4
CANVAS. Esquema del modelo de negocio 6
DAFO 7

ESTRATEGIA 8

Objetivos estratégicos 8
Mapa estratégico 8
KPI 9

ARP – TRADER COGNITIVO 10

Fuentes y modelo de datos a utilizar 10
Métodos analíticos 11
Simulación inversión basada en el uso de modelo de predicción de valores 17

ARQUITECTURA TÉCNICA 20

Diseño e infraestructura 20
Tecnologías a utilizar 21

PLAN DE NEGOCIO 23

Business Case 23

CONCLUSIONES 27

PROBLEMA DE NEGOCIO

¿POR QUÉ ARP - TRADER COGNITIVO?

El proyecto ARP nace con el ánimo de aprovechar los últimos avances tecnológicos y especialmente aquellos relacionados con la Inteligencia Artificial y de manera específica el Deep Learning, creando un sistema de gestión de carteras basado en estas tecnologías y apoyado en el Big Data.

Por un lado, buscamos obtener rentabilidades superiores a los índices de referencia de manera constante permitiendo que nuestros clientes obtengan los mejores retornos de su inversión de manera sostenida en el tiempo, de acuerdo con su nivel de riesgo.

Por otro, buscamos dar solución a una problemática generalizada en los países más desarrollados y que tiene un profundo impacto en nuestro país y es aquella relacionada con las pensiones. Nuestra solución pretende gestionar, no solo fondos de inversión, si no también planes de pensiones de manera privada que sean complementarios a las pensiones públicas y puedan proporcionar una renta adecuada a los trabajadores en el momento en que estos comiencen su jubilación.

Caso de uso y propuesta de valor

En los siguientes cuadros podemos ver un problema generalizado y es que las rentabilidades medias de los fondos de inversión ofrecidos y de los planes de pensiones individuales ofrecidos por las entidades tradicionales, comparando esto con las rentabilidades del Ibex35 vemos que las entidades tradicionales no son capaces de batir al índice de referencia.

RENTABILIDADES MEDIAS ANUALES PONDERADAS DE LOS FONDOS DE INVERSIÓN (31/08/18)

Tipo de Fondo	Mes	2018	1 Año	3 Años	5 Años	10 Años	15 Años	20 Años	25 Años
MONETARIOS	-0,08	-0,41	-0,53	-0,25	0,04	0,71	1,10	1,45	2,39
RENTA FIJA EURO CORTO PLAZO	-0,25	-0,95	-0,92	-0,21	0,16	0,71	1,15	1,51	2,43
RENTA FIJA EURO LARGO PLAZO	-0,48	-1,21	-1,01	0,40	1,73	2,50	2,19	2,42	3,31
RENTA FIJA MIXTA EURO	-0,92	-1,76	-1,37	0,03	1,59	1,63	2,02	1,80	3,03
RENTA VARIABLE MIXTA EURO	-1,72	-2,01	-1,56	0,54	3,83	2,00	3,07	1,89	3,71
RENTA VARIABLE NACIONAL EURO	-2,41	0,56	1,60	3,67	7,59	2,83	5,71	3,58	6,47
RENTA FIJA INTERNACIONAL	-0,55	-1,34	-1,32	-0,27	2,00	2,39	1,87	2,18	2,95
RENTA FIJA MIXTA INTERNACIONAL	-0,43	-1,20	-0,46	-0,04	1,04	0,81	1,16	1,11	2,34
RENTA VARIABLE MIXTA INTERNACIONAL	-0,34	-0,40	1,47	1,42	2,56	2,42	2,48	1,38	3,14
RENTA VARIABLE EURO RESTO	-3,25	-1,54	1,12	3,30	6,25	4,64	5,28	2,71	5,62
RENTA VBLE. INTERNACIONAL EUROPA	-1,83	-0,92	2,55	2,33	6,10	3,11	4,04	1,46	4,25
RENTA VBLE. INTERNACIONAL EEUU	3,52	9,27	17,26	12,00	12,87	9,05	5,91	4,07	2,55
RENTA VBLE. INTERNACIONAL JAPÓN	-0,26	-1,06	11,09	6,24	9,66	4,69	2,93	0,48	-0,78
RENTA VBLE. INTERNACIONAL EMERGENTES	-3,19	-6,00	-1,62	7,36	4,53	1,78	6,76	5,55	4,19
RENTA VBLE. INTERNACIONAL RESTO	-0,63	1,16	6,13	6,21	8,00	7,29	5,52	2,39	4,06
GLOBALES	-0,44	-0,44	1,91	2,06	3,64	2,37	2,39	1,79	3,26
GARANTIZADOS DE RENDIMIENTO FIJO	-0,96	-1,25	-1,07	-0,14	0,73	1,96	2,00	2,27	
GARANTIZADOS DE RENDIMIENTO VARIABLE	-1,20	-1,73	-1,40	0,12	1,36	1,67	1,93	2,08	
DE GARANTÍA PARCIAL	-0,65	-0,50	-0,10	0,40	3,25				
DE GESTIÓN PASIVA	-1,16	-0,35	0,07	1,22	3,33				
RETORNO ABSOLUTO	-0,41	-1,66	-1,17	-0,03	0,77				
FONDOS DE INVERSIÓN LIBRE (FIL)	-1,09	1,01	3,79	6,06	7,41				
FONDOS DE FIL	-0,40	1,72	1,47	-0,66	1,44				
TOTALES FONDOS	**-0,69**	**-0,73**	**0,43**	**1,10**	**2,16**	**2,06**	**2,28**	**1,82**	**2,86**

Rentabilidades medias anuales ponderadas. NETAS de gastos y comisiones (%)

SISTEMA INDIVIDUAL	26 AÑOS	25 AÑOS	20 AÑOS	15 AÑOS	10 AÑOS	5 AÑOS	3 AÑOS	1 AÑO
Renta Fija Corto Plazo	2,30	1,74	1,64	0,86	0,89	0,22	-0,35	**-1,43**
Renta Fija Largo Plazo	3,70	3,27	1,95	1,92	2,40	1,80	0,26	**-1,72**
Renta Fija Mixta	3,88	3,45	1,91	1,86	1,57	1,68	0,41	**-0,81**
Renta Variable Mixta	5,21	4,68	3,38	3,83	3,41	4,08	2,13	**1,89**
Renta Variable			2,16	4,52	4,71	8,39	5,44	**5,54**
Garantizados				2,81	3,23	4,77	1,70	**-0,36**
TOTAL PLANES	**3,90**	**3,47**	**2,29**	**2,54**	**2,45**	**3,24**	**1,47**	**0,54**

RENTABILIDAD DEL IBEX 35

Año	Rentabilidad*
2007	10,71
2008	-36,50
2009	38,27
2010	-12,93
2011	-7,75
2012	2,78
2013	27,75
2014	8,62
2015	-3,55
2016	2,60
2017	11,25
Rentabilidad 10 años	**10,74**

** Rentabilidad calculada sobre el cierre del año anterior (%)*

Este problema no solo es propio de las entidades nacionales si no que se trata de un fenómeno generalizado, con un porcentaje de fondos que son capaces de superar de manera consistente al índice S&P en sus diferentes categorías, en el mejor de los casos solo un 6’88% de dichos fondos con capaces de hacerlo.

Fund category	Benchmark	Percentage of funds that outperformed their benchmark for three-years	Percentage of funds that outperformed in the subsequent year	Percentage of funds that outperformed for two subsequent years	Percentage of funds that outperformed for three subsequent years
Large-cap	S&P 500	30.83%	33.93%	13.62%	5.17%
Mid-cap	S&P MidCap 400	25.65%	30.39%	10.35%	3.24%
Small-cap	S&P SmallCap 600	30.58%	35.25%	13.30%	4.60%
International	S&P 700	23.89%	36.68%	15.56%	6.88%
Emerging markets	S&P/IFCI Composite	25.24%	38.39%	15.48%	5.22%

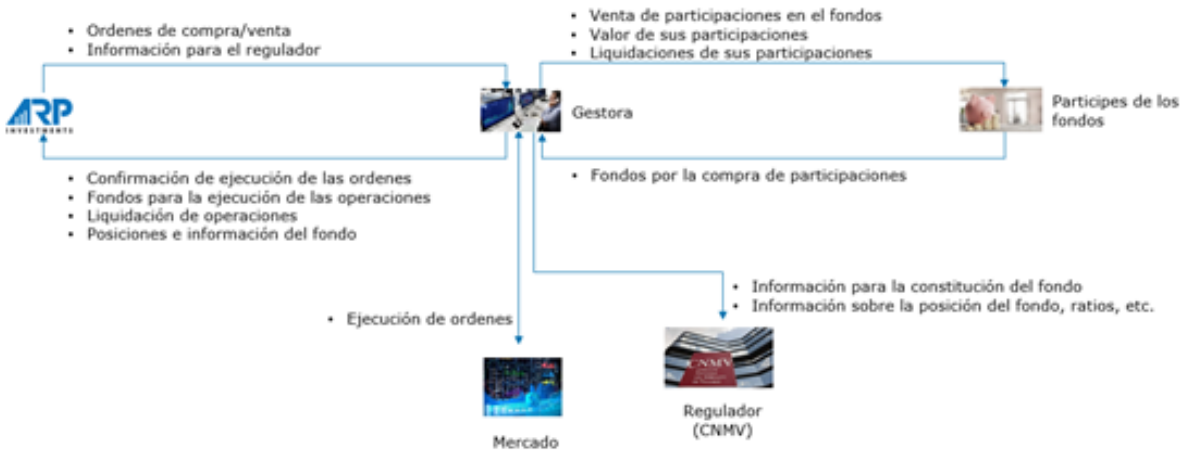
Si a esto le sumamos las comisiones que se les aplican a los clientes, observamos que la rentabilidad de los fondos disminuye aún más, efecto que se agrava por el efecto de la inflación, lo que nos llevaría en ocasiones a rentabilidades negativas.

Nuestra propuesta de valor se basa en ofrecer rentabilidades superiores al retorno del índice SP500, que es el que hemos tomado en nuestro caso como índice de referencia, mediante el uso de la algoritmos basados en Deep learning, que nos permitan prever el comportamiento de los valores que componen este índice de referencia, estando dotado así mismo de flexibilidad para ir ajustando el riesgo de forma sistemática con el paso del tiempo, de tal manera que el modelo tenga capacidad para adaptarse de forma autónoma a la situación del mercado y todo ello sin la intervención del ser humano.

MODELO DE NEGOCIO

Nuestro modelo de negocio se basaría en una solución B2B en la opción partnership. Esto consistiría en la búsqueda de un gestor de fondos "partner" junto con el cual venderíamos nuestro servicio a clientes finales (los beneficios se repartirían entre las partes). En esta opción los clientes del fondo serían los clientes finales con la diferencia que en este caso tanto la imagen de marca y los aspectos regulatorios ya estarían cubiertos por parte del partner. Esta decisión se ha tomado debido a que se ha considerado la opción más simple de implementar en una primera fase y así poder centrarnos en lo que realmente sería nuestro valor añadido: el desarrollo de modelos de trading algorítmico utilizando tecnologías basadas en IA.

La representación conceptual de los diferentes actores implicados en la solución sería la siguiente:



MISIÓN, VISIÓN Y VALORES

Misión

Nuestra misión es la creación de fondos de gestión que superen de manera recurrente a los índices de referencia, todo esto haciendo uso de tecnologías innovadoras basadas en machine learning e inteligencia artificial y utilizando

fuentes de datos financieras tradicionales y algunas fuentes alternativas como son noticias o RSS.

Visión

La visión de nuestra empresa se centra en convertirse en una gestora de referencia a nivel nacional e internacional es los diferentes aspectos relacionados con la inversión en valores de diferentes mercados a partir del uso de las tecnologías descritas.

Valores

- *Confianza*
- *Trato personalizado*
- *Sencillez*
- *Integridad profesional*
- *Excelencia: Know-how financiero y analítico*

CANVAS. Esquema del modelo de negocio

<table>
<tr>
<td rowspan="2">Aliados Clave

- Socio financiero de referencia.

- Compañías comercializadoras

- Compañías obtención de datos financieros.

- Compañías obtención da datos de noticias y RRSS.

- Empresas de servicios cloud para despliegue de la solución.</td>
<td>Actividades Clave

- Gestión de fondos/ planes de pensiones ajustados a los diferentes niveles de riesgo del cliente.

- Análisis de datos financieros y otras fuentes alternativas.

- Creación de modelos basados en IA para la elaboración de inversiones.

- Automatización de procesos</td>
<td rowspan="2">Propuesta de Valor

- Diseño y gestión de operativa de fondos de inversión que mejoran la rentabilidad respecto a los índices de referencia.

- Modelo disruptivo</td>
<td>Relación con el Cliente

- Relación de confianza a largo plazo con un socio de referencia.

- Acuerdos comerciales con los distribuidores de los fondos.</td>
<td rowspan="2">Segmentos de Clientes

- Socio partner creador del fondo de inversión.</td>
</tr>
<tr>
<td>Recursos Clave

- Personal altamente cualificado con conocimientos financieros y en data science/IA

- Fuentes de datos financieros

- Fuentes de datos sobre noticias y RRSS.

- Servicios cloud para el despliegue de la solución</td>
<td>Canales

- Venta: Presencial

- Soporte: Mail, teléfono.</td>
</tr>
<tr>
<td colspan="3">Estructura de Costes

- RRHH (equipo de desarrollo de la solución, equipo de ventas...)

- Costes de infraestructura (oficinas y equipos)

- Costes de hosting de la solución y pago por uso servicios cloud.

- Obtención de datos (financieros y noticias/RRSS).

- Marketing</td>
<td colspan="2">Estructura de Ingresos

- Comisiones por gestión de fondos.

- Comisión de performance.</td>
</tr>
</table>

DAFO

Fortalezas

- Algoritmo capaz de superar al índice de referencia.
- Uso de IA como tecnología innovadora para la generación de las estrategias.
- Aportaciones inicial reducida en comparación a otros fondos.

Debilidades

- Falta de imagen de marca.
- Necesidad de asociación con un partner ya presente en el mercado

Oportunidades

- Capaces de ser independientes del ciclo económico.
- Sin diferenciar por nivel de inversión de los clientes.
- Solución tecnológica innovadora no utilizada en el sector de una forma generalizada

Amenazas

- Competencia en constante crecimiento.
- Sector muy regulado.
- Competencia consolidada.
- Socio partner tome un control elevado sobre el servicio
- Algoritmo no alcance los objetivos definidos.

ESTRATEGIA

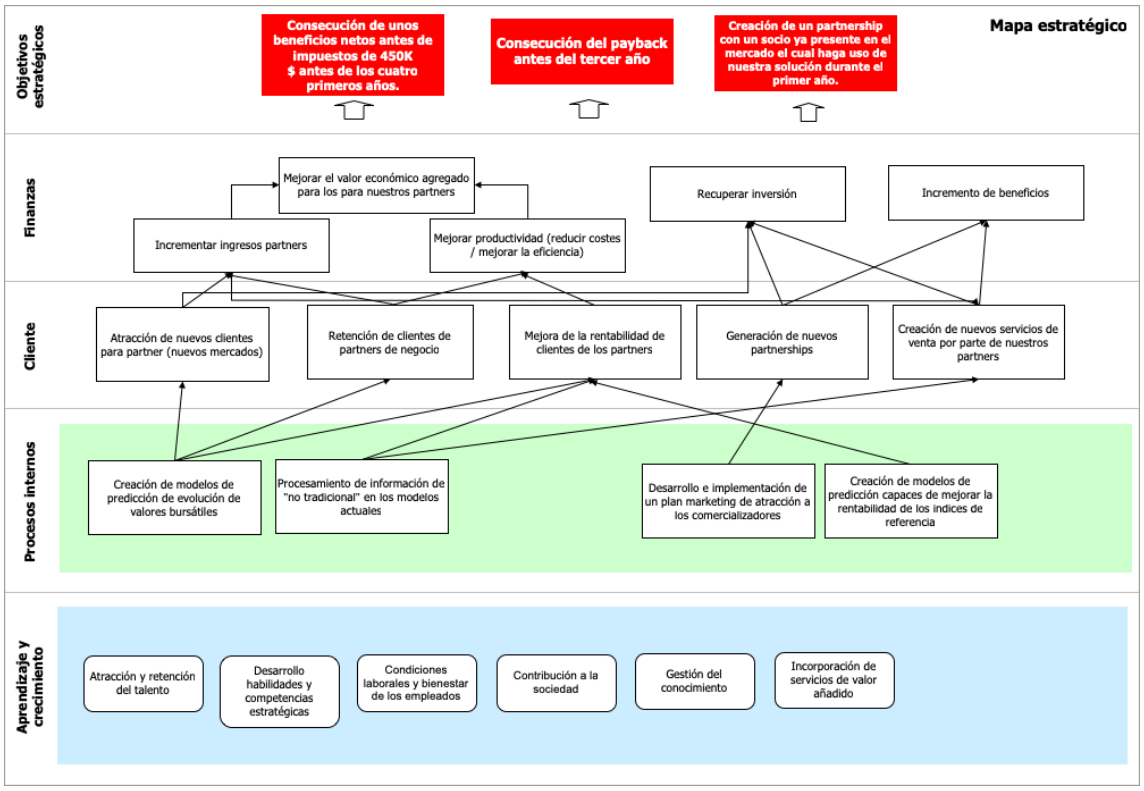
Objetivos estratégicos

Los objetivos estratégicos de nuestra compañía se centrarían en los siguientes puntos:

- *Consecución de unos beneficios antes de impuestos de 450K$ antes de los cuatro primeros años.*
- *Consecución del payback en el segundo año.*
- *Creación de un partnership con un socio ya presente en el mercado el cual haga uso de nuestra solución durante el primer año.*

Para lo cual se ha desarrollado el siguiente mapa estratégico con los diferentes indicadores.

MAPA ESTRATÉGICO



KPI

Financieros

- Rentabilidad obtenida por parte de los partners
- Reducción de costes en la obtención de las recomendaciones de inversión
- Número y volumen de partners se ajusta a lo planificado
- El volumen de cartera de los partners coincide con lo planificado

Clientes

- Incremento de clientes nuevos finales de nuevos partners
- Tiempo en el que los clientes finales mantienen su inversión en los servicios de nuestros socios
- Retorno de la de inversión de nuestro modelo
- El número de acuerdos se ajusta a lo planificado
- Generación de nuevo valor por parte de nuestros clientes

Procesos

- Generación de modelos de predicción que expliquen la evolución de los valores bursátiles
- Inclusión de información no tradicional en los modelos de predicción
- Incorporación de nuevos comercializares
- Generación de modelos de predicción capaces de mejorar la rentabilidad respecto a los índices de referencia.

Intangibles

- Índice de rotación en la empresa
- Conseguir certificaciones técnicas y financieras en las áreas necesarias
- Encuestas de clima laboral
- Repercusión de una parte de los beneficios en la sociedad-
- Planes de formación adaptado a las necesidades de la empresa
- Desarrollo de nuevos servicios

ARP – TRADER COGNITIVO

FUENTES Y MODELO DE DATOS A UTILIZAR

La fuente de datos principal en nuestro proyecto, en base al análisis realizado, va a ser Intrinio ya que nos ofrece una solución completa en cuanto al tipo y volumen de datos ofrecidos.

A continuación, se muestra el listado de información y los datos a utilizar en nuestro modelo.

Nuestro sistema tiene dos partes claramente diferenciadas: Una primera parte de análisis de información no estructurada de noticias y una segunda parte en la que se combinaría el resultado del análisis de información procedente del análisis de noticias (resultado obtenido ya sería información estructurada) y la información financiera "tradicional" sobre cotizaciones (información estructurada).

ANÁLISIS DE NOTICIAS

En esta parte la información de entrada sería la procedente del API de Intrinio sobre noticias financieras en la que a partir del nombre de una cotización y de un rango de fechas de búsqueda, devuelve el listado de noticias sobre dicha cotización en el periodo indicado. Por tanto, de este proveedor obtendríamos la siguiente información de cada noticia:

- *Titulo*
- *Fecha de publicación*
- *URL*

De la información disponible el servicio de análisis de noticias obtenemos la url de la misma, dado que el servicio de NLU acepta el envío de urls para su análisis.

El resultado del análisis de las noticias devuelve numerosa información. Sin embargo, tras la realización de un análisis de esta, la información que se utilizará como entrada del modelo de predicción de valores es la relativa al análisis de sentimiento y emociones.

MODELO DE PREDICCIÓN DE VALORES

En esta parte se describen los datos que formarían parte de la entrada de nuestro modelo, así como el resultado el resultado de este. Como parámetros de entrada tendríamos dos grupos de datos:

- *Información bursátil de una cotización (procedente del API de Intrinio):*
    - *Máximo en un día.*
    - *Mínimo en un día.*
    - *Precios de apertura en un día.*
    - *Precios de cierre en un día.*
    - *Volumen de acciones*
- *Resultado análisis noticias del día:*
    - *Sentimiento: Mediana, máximo y mínimo del día*
    - *Emoción (alegría): Mediana, máximo y mínimo del día*
    - *Emoción (enfado): Mediana, máximo y mínimo del día*
    - *Emoción (disgusto): Mediana, máximo y mínimo del día*
    - *Emoción (tristeza): Mediana, máximo y mínimo del día*
    - *Emoción (temor): Mediana, máximo y mínimo del día*

Dado que en un día concreto pueden existir diversas noticias sobre una cotización, sobre el resultado del análisis de noticias descrito en el apartado anterior se realizaría un preprocesado del resultado para utilizarlo como valores de entrada del modelo de predicción, con el fin de poder obtener más información.

Como comentario general hay que indicar que el listado de datos descritos es un subconjunto de toda la información financiera estructurada disponible. Estos datos han sido utilizados para desarrollar la prueba de concepto descrita a continuación. No obstante, el modelo descrito no presentaría ningún problema, si dicha información estructurada fuera ampliada con nuevos valores estructurados.

SIMULACIÓN DE MÉTODOS ANALÍTICOS

De cara a la elaboración de los métodos analíticos que van a ser capaces de predecir los valores y tendencias de las acciones se han utilizado diferentes tecnologías.

En primer lugar, para tratar de abordar el problema principal "predicción de los valores de una acción en un periodo de tiempo futuro" se trató de dividir el caso en problemas más pequeños. Así pues, identificamos tres partes claramente diferenciadas:

- *Análisis y predicción de valores a partir de información estructurada financiera.*
- *Análisis y procesamiento de información no estructurada procedente de noticias para la mejora de las predicciones sobre los valores.*
- *Análisis y predicción de valores a partir de la combinación de la información estructurada financiera y no estructurada procedente de noticias.*

Análisis y predicción de valores a partir de información financiera

En un primer momento se trató de analizar información estructurada financiera "tradicional" a partir de la cual durante mucho tiempo se ha venido implementado diferentes modelos de predicción de valores.

En nuestro caso, a partir de esta información hemos tratado de crear un modelo de predicción de valores, pero en este caso utilizando el Deep Learning como herramienta.

Para ello, tras un análisis de las diferentes posibilidades que ofrecía este tipo de modelos, nos decantamos por la utilización de redes neuronales recurrentes LSTM (Long Short-Terma Memory).

Las redes neuronales recurrentes LSTM se caracterizan porque sus valores de entrada pueden ser diferentes tipos de variables, pero entre ellas destaca el factor temporal. Es decir, podemos tomar como variables de entrada los valores de una misma variable, tomados en diferentes espacios de tiempos pasados respecto a la predicción que se desea obtener. Unas de estas variables de entrada pude ser incluso la propia variable objetivo en momentos temporales pasados.

Esta característica de este tipo de redes es muy valiosa de cara al problema que queremos abordar ya que las predicciones en bolsa también se ven influenciadas por la información de días anteriores. Son redes neuronales "con memoria".

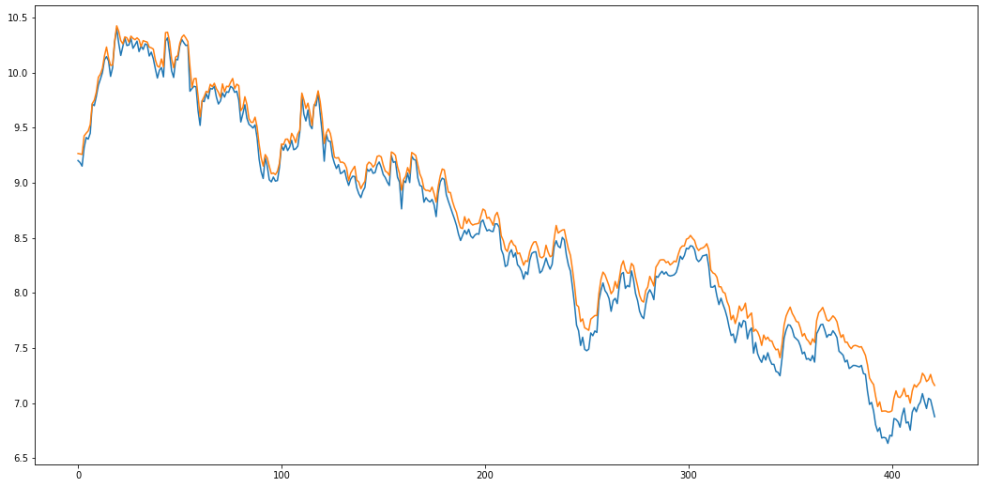
Predicción del valor de cierre de una cotización

Nuestro primer objetivo planteado fue la creación de modelos capaces de predecir el valor de cierre de una cotización para día siguiente.

Para ello creamos una red neuronal con la siguiente estructura.

- *Capa de entrada: 5 neuronas (variables indicadas).*
- *Capa oculta: 50 neuronas LSTM.*
- *Capa de salida: 1 neurona con el resultado.*

A partir de esta estructura y utilizando técnicas de dropout para tratar de evitar el sobre-entrenamiento del modelo, se obtuvieron los siguientes resultados (ejemplo con la cotización de Telefónica):




```
Test R2: 0.989
```

La traducción de las gráficas sería la siguiente:

- *La gráfica azul representa el valor real a lo largo del tiempo (nuestro dataset).*
- *La gráfica naranja representa el resultado de nuestro modelo sobre los valores de entrenamiento utilizados para ajustar el modelo.*

Como se puede ver los valores del R^2 con este modelo presenta cifras entorno al 0,99.

Sin embargo, nuestro objetivo final de cara a la gestión de fondos no es tanto tener con una precisión elevada el valor de la cotización del día siguiente, más

bien sería poder tener predicciones de la tendencia que va a llevar una cotización en un medio plazo (por ejemplo, en la próxima semana). Esta información podría ser usada para la implementación de un sistema de alertas en caso de predecir una variación elevada en el día siguiente.

Predicción de la tendencia del valor de cierre de una cotización en varios días siguientes al día actual

Para la construcción de este modelo aprovechamos la característica de las redes neuronales recurrentes LTSM por la cual son capaces de predecir los valores de una serie temporal no solo para la siguiente unidad de tiempo si no para sucesivos valores futuros.

En este caso se construyó una arquitectura de red neuronal con una capa con 5 neuronas LTSM en la capa oculta para una predicción de siete días en el futuro. Con esto se obtuvieron los siguientes resultados:

Cotización	R^2 día t+1	R^2 día t+2	R^2 día t+3	R^2 día t+4	R^2 día t+5	R^2 día t+6	R^2 día t+7
Apple	0,996	0,991	0,987	0,982	0,977	0,971	0,965
Google	0,994	0,987	0,979	0,972	0,962	0,954	0,944

Como podemos ver la precisión de nuestro modelo va empeorando (como era esperado) conforme nuestras predicciones se van alejando de los valores reales de los cuales disponemos. La precisión en el séptimo día estaba entorno al 94%-96%.

En este punto lo que planteamos era como podíamos mejorar las predicciones de las tendencias de las cotizaciones. Para ello planteamos la inclusión de nuevas variables "no tradicionales" dentro de nuestro modelo como veremos en los siguientes apartados.

Análisis y procesamiento de información no estructurada de noticias

Este apartado se focaliza en el análisis de sentimiento y emociones de noticias financieras sobre las cotizaciones que estamos analizando. Para ello hemos utilizado la herramienta de Watson de IBM, Natural Language Understanding. Por

la cual, a partir de una noticia, el sistema es capaz de devolvernos los resultados sobre los sentimientos y emociones presentes en una noticia y las probabilidades de estas.

Por ello, el modo en el que utilizaríamos el servicio sería, a partir de las noticias del día de una compañía, enviaríamos el texto o la url de la noticia (Intrinio devuelve ambos valores) y obtendríamos los resultados de sentimientos y emociones de esta para combinarlo con el resultado del resto de noticias analizadas.

Para demostrar el uso de este servicio se ha utilizado una noticia de ejemplos:

URL noticia: https://finance.yahoo.com/news/qualcomm-says-apple-7-billion-012941302.html?.tsrc=rss&guccounter=1

Texto:

> (Reuters) - Qualcomm Inc said in a court hearing on Friday that Apple Inc is $7 billion (£5.5 billion) behind in patent royalty payments to the mobile chip firm, which for years supplied parts for the iPhone.
>
> Apple and Qualcomm are involved in a series legal actions. Apple has accused Qualcomm of unfair patent licensing practices. Qualcomm, the world's largest mobile phone chip supplier, has in turn accused Apple of patent infringement.
>
> Qualcomm made its comments about the size of Apple's unpaid royalties in a hearing in one of the cases in federal court in San Diego. Apple has disputed the amount of royalties that it owes to Qualcomm.
>
> Apple has argued that Qualcomm is forcing it to pay for the same patents twice, once when Apple uses Qualcomm's chip in iPhones and then again through patent royalties. Qualcomm has argued that its practices are legal and that Apple is seeking to destroy Qualcomm's business model after agreeing to it for years.

Tras analizar dicha noticia con el servicio NLU de IBM Watson el resultado fue el siguiente:

Análisis	Resultado
Sentimiento	Negativo (-0.29)
Emociones (Alegría)	0.15
Emociones (Enfado)	0.57
Emociones (Disgusto)	0.16
Emociones (Tristeza)	0.20
Emociones (Temor)	0.05

A partir de esta información el resultado podría combinarse con el resultado del resto de noticias de esa compañía. Con estos resultados habríamos transformado la información no estructurada de una noticia, en información estructurada la cual también puede combinarse con la información financiera la cual se había descrito en el apartado anterior.

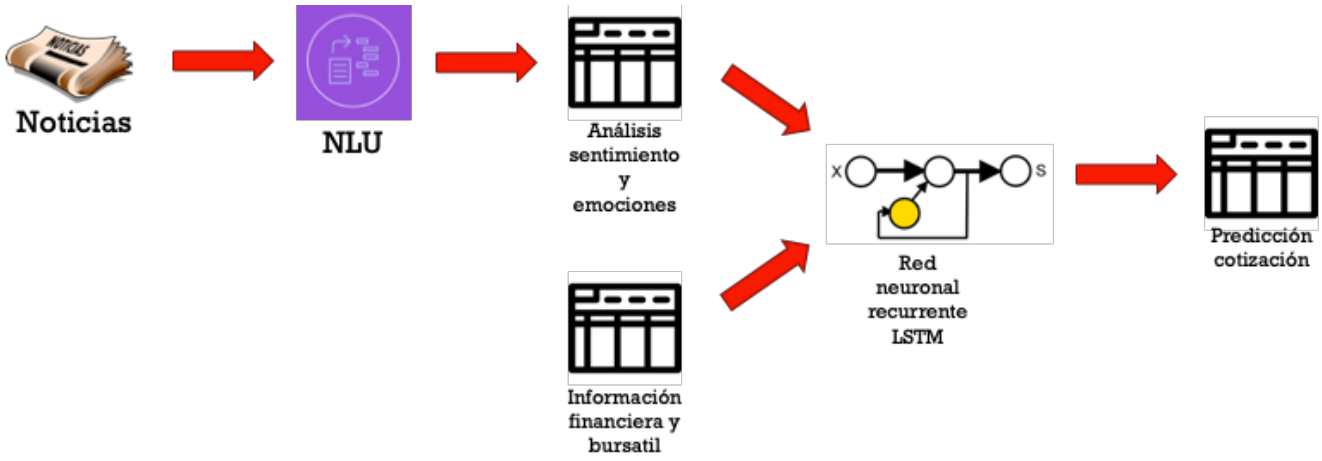
Combinación de información estructurada y no estructurada para la predicción de cotizaciones

En esta parte del análisis se trató de combinar la información financiera que veíamos en el primer apartado con la cual ya teníamos un modelo capaz de predecir tendencias de los valores de cierre de una cotización y tratar de mejorar su precisión con la inclusión de nuevas variables sobre el sentimiento de noticias financieras de un día concreto sobre la cotización que se está analizando.

Para la ejecución de esta combinación se ha partido del mismo modelo presentado en el primer apartado. Es decir, una red neuronal recurrente LSTM (Long Short-Term Memory), para conseguir realizar predicciones de los valores de una cotización durante los días siguientes y así poder identificar tendencias de estas y ejecutar las acciones de compra o venta según corresponda.

A partir de esta información, podríamos obtener las predicciones de los siguientes días al igual que se realizaba en el primer apartado.

De este modo, el flujo de procesamiento de la información dentro de nuestro modelo se puede ver reflejado en alto nivel en el siguiente gráfico:



SIMULACIÓN INVERSIÓN BASADA EN EL USO DE MODELO DE PREDICCIÓN DE VALORES

Una vez elaborado y entrenado el modelo de predicción de valores descrito en los apartados anteriores se trató de aplicar nuestro modelo en el proceso de decisión de inversión dentro de una selección de valores. Este sería el uso principal para el que se utilizaría nuestro modelo.

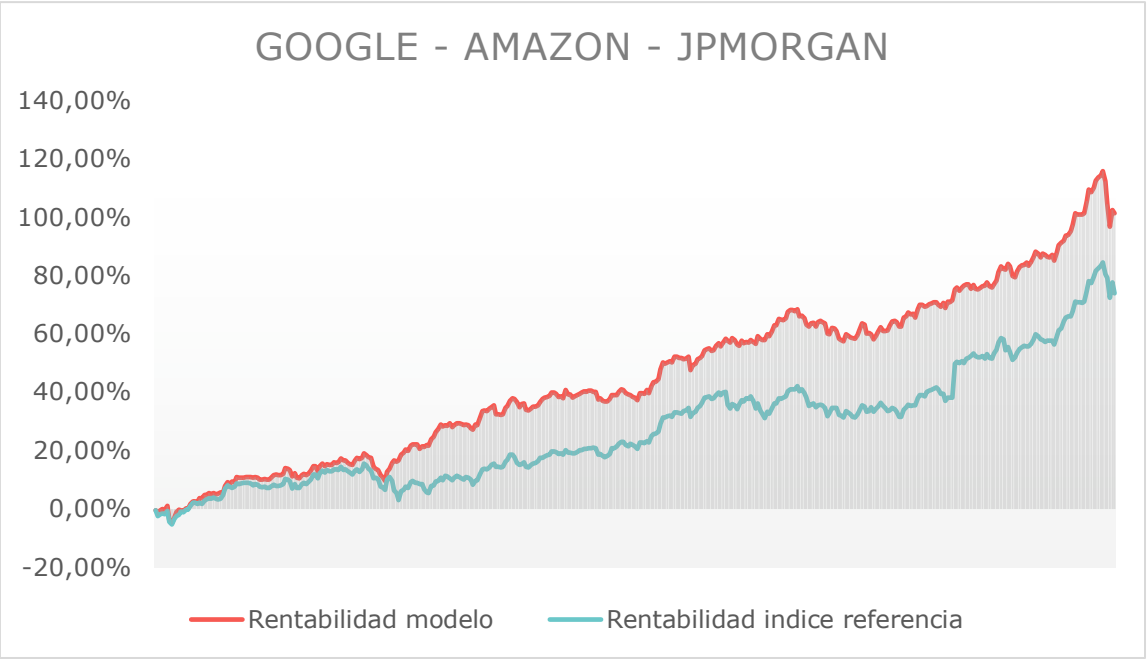
Para tratar de simplificar el problema de cara a la simulación lo que se hizo fue tomar como referencia tres cotizaciones del S&P500 durante un periodo de tiempo pasado (la rentabilidad real combinada de estos tres valores durante este periodo representaría al índice de referencia al cual debemos intentar batir).

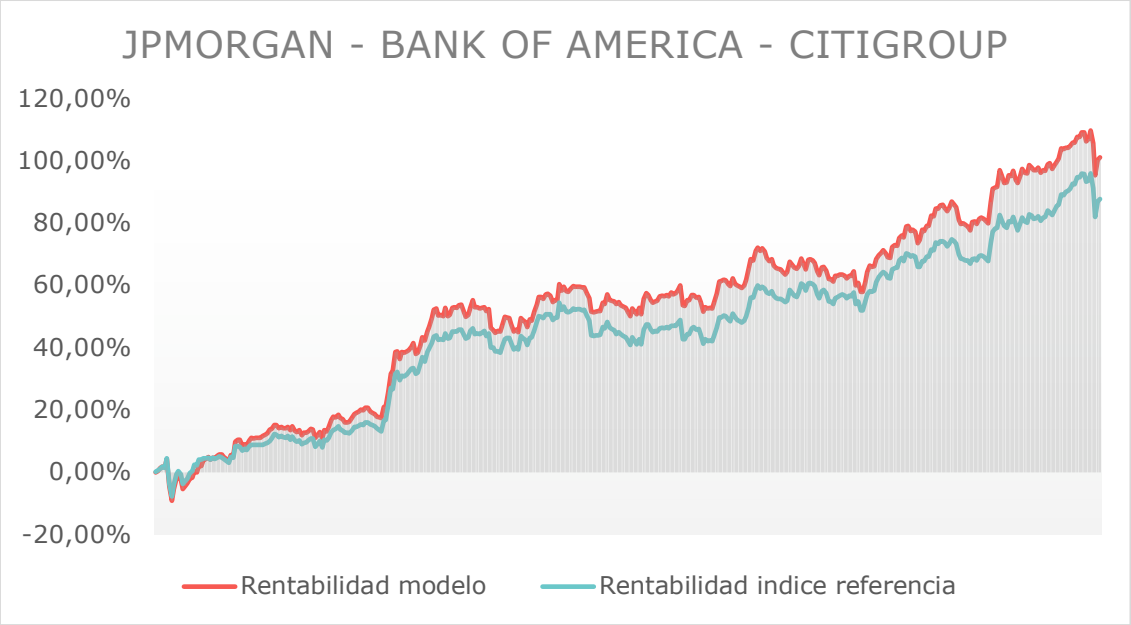
Por otro lado, nuestra política de inversión se fundamentó en la inversión diaria en los dos mejores valores de acuerdo con la predicción del precio de cierre del día siguiente de nuestro modelo. Es cierto que esta política es una simplificación del modelo de inversión que tendría el fondo, tal y como se describió en la política del mismo. Sin embargo, esta simulación nos permite conocer y validar si nuestro modelo funciona correctamente.

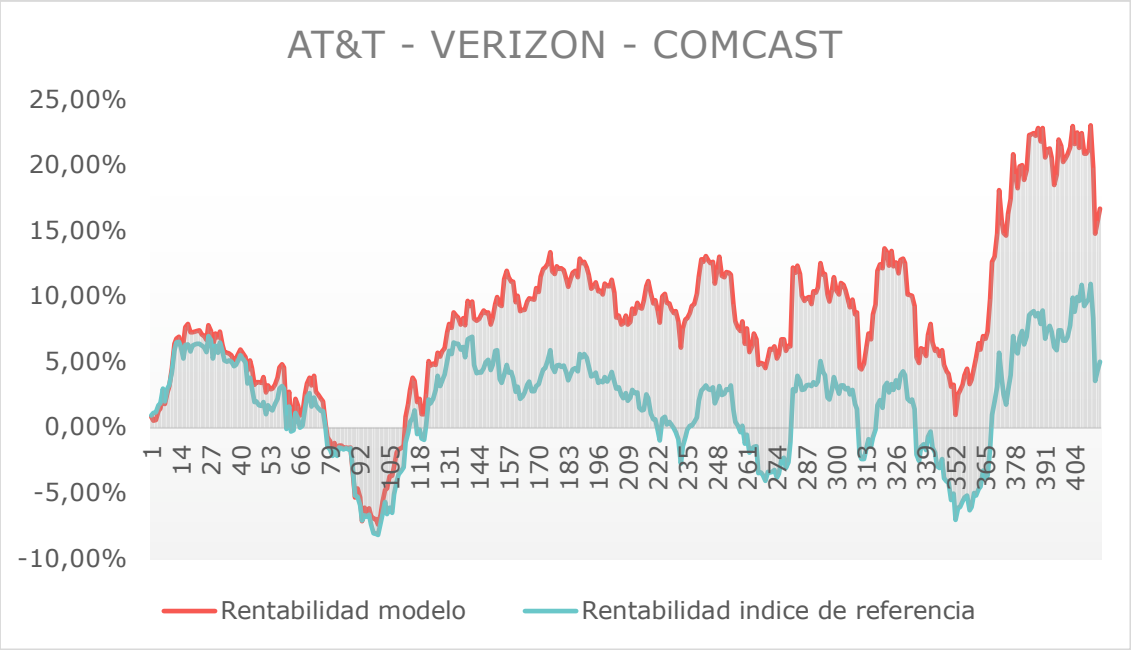
Dicha simulación ha sido repetida utilizando diferentes valores en el mismo periodo de tiempo (desde 16/6/2016 al 07/02/2018) con el fin de confirmar el funcionamiento. A continuación, se resume el resultado de las simulaciones realizadas:

Valores utilizados	Rentabilidad índice de referencia	Rentabilidad simulación	Mejora respecto al índice de referencia
Google – Amazon - JP Morgan	74,04 %	101,42 %	36,98 %
JP Morgan – Bank of America - Citigroup	87,77 %	101,20 %	15,30 %
AT&T - Verizon - Comcast	5,02%	16,69 %	232,62 %

Adicional podemos ver la rentabilidad acumulada diaria de cada una de estas simulaciones frente a la rentabilidad del índice de referencia en cada uno de los casos:







Estas simulaciones se fundamentan en una simplificación de lo que sería la política de inversión con la que se gestionará nuestro fondo de inversión. No obstante, en base a los resultados obtenidos, podríamos afirmar que nuestro modelo planteado funciona y podría ser utilizado tomando un mayor rango de valores a analizar y aplicando las políticas de inversión necesarias, fundamentadas en el resultado de nuestras predicciones. Cabe destacar que estas simulaciones habrían superado nuestro KPI de superar un 10% al índice de referencia.

ARQUITECTURA TÉCNICA

DISEÑO E INFRAESTRUCTURA

Nuestra plataforma se basará en el análisis diario y predicción de la tendencia de todas las acciones del S&P500, estimando las probabilidades de subir o bajar respecto a la posición actual, para realizar dichas estimaciones utilizaremos la información histórica del S&P500 así como todas las noticias financieras que hablen de los valores del índice, esto nos ayudará a predecir posibles comportamientos de los valores basados en noticias tanto positivas como negativas. Nuestro algoritmo tendrá en cuenta como otros tipos de noticias que no hablen directamente de los valores del S&P500 pueden afectar positiva o negativa a dichos valores, como pueden ser noticias políticas, climatológicas, sociales, deportivas, etc.

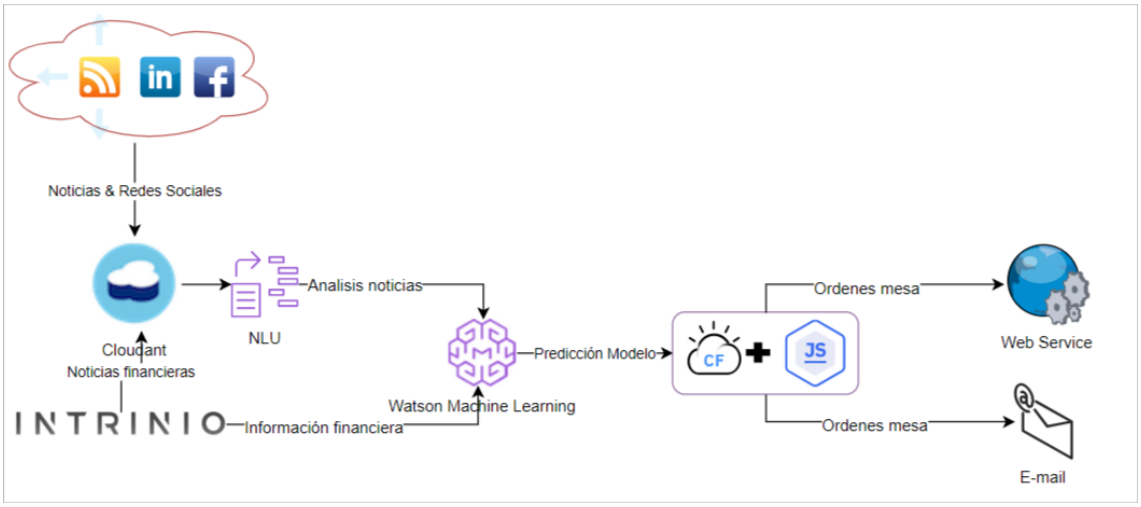
Toda la información y noticias recogidas serán almacenadas en nuestro almacén de datos basado en Cloudant, para posteriormente aplicar nuestro modelo de "Reconocimiento natural del lenguaje" (NLU) que nos devolverá si dichas noticias son relevantes para los valores del S&P500 y en qué medida, ayudando a predecir de forma más exacta el comportamiento de los valores.

Sobre los datos actuales de los valores en bolsa, conjugados con toda la información obtenida de las noticias, hemos desarrollado un modelo de Deep Learning capaz de obtener 2 predicciones compuestas por el porcentaje de variación de los valores para el día siguiente y el porcentaje de variación de los valores para los próximos 7 días.

Esta información será utilizada para definir la cartera tipo recomendada, así como las actuaciones que haya que aplicar sobre las carteras tipo existentes actualmente. Recomendando cambiar los pesos entre los valores de la cartera, así como la venta de algunos de ellos o la compra de otros.

Toda esta información se convertirá en órdenes de compraventa que serán puestas a disposición de la mesa de nuestro partner, que se encargará de ejecutarlas y dejar configurada la cartera tal y como nuestro recomendador indique.

Para hacer llegar toda esta información a la mesa dispondremos de dos modos, uno mediante envío de correo electrónico a la mesa y el otro será un servicio web que se podrá integrar con los servicios de compra venta de la mesa para hacerle llegar la ordenes de manera automática.



TECNOLOGÍAS A UTILIZAR

	Tecnología	Funcionalidad
	Discovery	
	Watson Machine Learning	Plataforma de ejecución del modelo estadístico de predicción
	Natural Language Understanding	Analiza y valora las noticias de la bolsa
	Watson Studio	Plataforma para el entrenamiento del modelo de aprendizaje profundo
	Knowlegde Studio	Plataforma de entrenamiento de los dominios necesarios para NLU
	Db2	Base de datos Relacional
	Cloudant	Base de datos Big Data para el almacenamiento de las noticias y la información recibida de las diferentes bolsas

	Node JS	Integrador de los diferentes servicios que expondrá los resultados del modelo mediante API Rest
	Cloud Foundry Enterprise	Entorno de ejecución de la solución
	Keras	Librería de Deep Learning para el diseño de modelos
	Tensorflow	Librería de Deep Learning para el diseño de modelos

PLAN DE NEGOCIO

BUSINESS CASE

A continuación, se detalla el Business case en el que se aborda la viabilidad del proyecto.

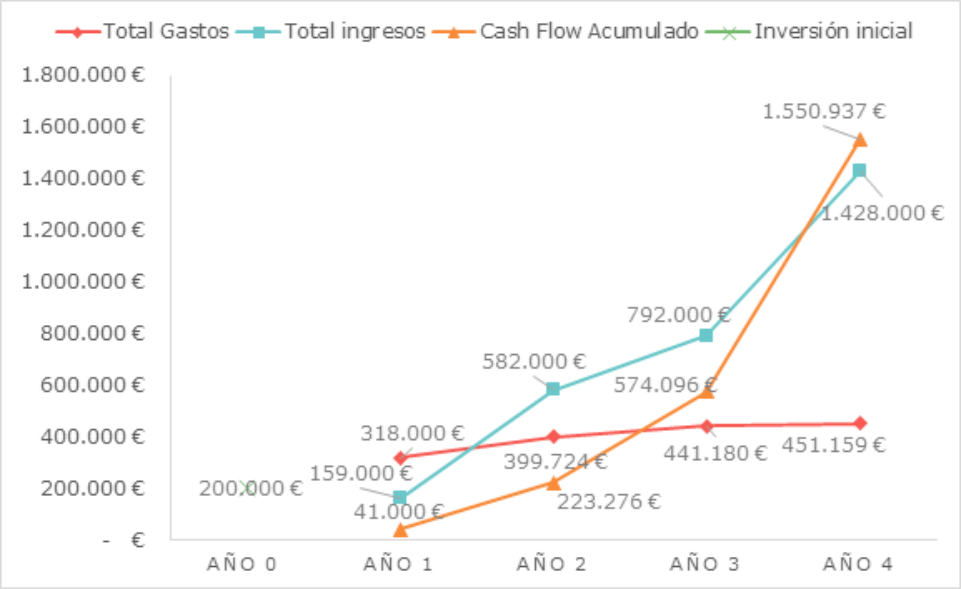
En primer lugar, se presenta una tabla detallada con las principales partidas de gastos e ingresos que se han considerado en el proyecto:

	Año 1	Año 2	Año 3	Año 4	Total
Gastos					
Tecnología	7.000 €	9.500 €	12.000 €	12.000 €	**40.500 €**
Servicios cloud/Infraestuctura	5.000 €	7.500 €	10.000 €	10.000 €	**32.500 €**
Fuentes de datos	2.000 €	2.000 €	2.000 €	2.000 €	**8.000 €**
Gastos estructurales	305.000 €	307.000 €	346.000 €	356.000 €	**1.314.000 €**
RRHH	215.000 €	220.000 €	260.000 €	270.000 €	**965.000 €**
Marketing	50.000 €	45.000 €	40.000 €	40.000 €	**175.000 €**
Oficina	36.000 €	36.000 €	36.000 €	36.000 €	**144.000 €**
Renting equipo informáticos	4.000 €	6.000 €	10.000 €	10.000 €	**30.000 €**
Servicios externos	6.000 €	83.224 €	83.180 €	83.159 €	**255.563 €**
Administración y asesoría legal	6.000 €	6.000 €	6.000 €	6.000 €	**24.000 €**
Devolución crédito	- €	77.224 €	77.180 €	77.159 €	**231.563 €**
Total Gastos	318.000 €	399.724 €	441.180 €	451.159 €	**1.610.063 €**
Inversión inicial	200.000 €	- €	- €	- €	**240.000 €**
Ingresos					
Comisiones de gestión	105.000 €	420.000 €	630.000 €	1.050.000 €	**2.205.000 €**
Comisiones de éxito	54.000 €	162.000 €	162.000 €	378.000 €	**756.000 €**
Total ingresos	159.000 €	582.000 €	792.000 €	1.428.000 €	**2.961.000 €**

Esto da lugar a unos los siguientes beneficios anuales y acumulados:

	Año 1	Año 2	Año 3	Año 4
Beneficios	- 159.000 €	182.276 €	350.820 €	976.841 €
Beneficios acumulados	- 159.000 €	23.276 €	374.096 €	1.350.937 €

Gráficamente la representación de los datos anteriormente expuestos quedaría del siguiente modo:



A continuación se detalla cada una de las partidas antes identificadas:

Inversión inicial

Tras un análisis de la oportunidad en detalle se validó que para la puesta en marcha del proyecto se hacía necesario de una inversión inicial de 200.000 €. Los cuales inicialmente se han planteado obtener a partir de un préstamo bancario con las siguientes condiciones:

Se trataría de un préstamo a 48 meses con unos 12 primeros meses de cadencia y un TAE 6%.

Ingresos

Los ingresos en este proyecto, al igual que en el resto de los gestores de fondos, proceden principalmente de las diferentes comisiones percibidas por la gestión de estos.

En este caso se realizó un análisis de los diferentes tipos de comisiones existentes, así como los límites máximos legales aplicables (se tomó España como país de referencia para este análisis) y se consideró que el reparto de estas se debería realizar al 50 % entre nuestra compañía y nuestro socio junto con la comercializadora.

Así pues, las comisiones que se consideraron fueron (en nuestro caso nos correspondería el 50 %):

- *Comisión de gestión: 1,35%*
- *Comisión de éxito: 9%*

Para el cálculo de la comisión de éxito se consideró que en media nuestros modelos tendrían una ganancia media anual del 8% sobre el capital invertido, (para esto se ha tenido en cuenta la ganancia histórica en bolsa y hemos aplicado nuestro KPI de mejora sobre el índice de referencia).

Adicional para el cálculo de los ingresos se analizó la capacidad de obtención de capital que sería gestionado a través de nuestra compañía y realizamos la siguiente estimación (esta estimación está basada en base al análisis del volumen de fondos gestionados actualmente en España http://www.inverco.es/38/39/101/2018/6):

- *Año 1: 30 M € (Durante 6 meses)*
- *Año 2: 60 M €*
- *Año 3: 90 M €*
- *Año 4: 150 M €*

Con estos volúmenes en el cuarto año representaríamos el 0,056% del mercado en España.

Gastos

El detalle de nuestros gastos sería el siguiente:

Plan de recursos humanos: donde estaríamos contemplando los sueldos de los tres socios de la empresa, un técnico de sistemas y un administrativo y a partir del tercer año realizaríamos la contratación un data scientist.

Plan de marketing: donde incluiríamos campañas para dar a conocer nuestro fondo permitiéndonos la consecución de una alianza con un socio estratégico que diera cobertura a nuestro fondo, acuerdos con distribuidoras de fondos para que lo incluyeran en su catálogo de productos y plataformas como Morningstar que permiten publicitar las rentabilidades, riesgos y comparar nuestros fondos con otros vehículos de inversión.

Plan de operaciones: en este apartado incluiríamos gastos de infraestructura tecnológica y fuentes de datos, alquiler de oficina y equipo informático y servicios externos. Es de recalcar que en este apartado tenemos en cuenta del cierre de

un acuerdo con IBM como socio tecnológico mediante su Global Entrepreneur que presta apoyo tecnológico a startups, para iniciar los primeros pasos en IBM Cloud, lo cual reduciría el coste de la implementación tecnológica.

Riesgos analizados

Para la obtención de los principales indicadores también se analizaron los principales riesgos en nuestro proyecto. Identificándose como principales tres de ellos:

- *Algoritmo se comporta por debajo de las expectativas*
- *No conseguir socio financiero en el primer año*
- *La captación de comercializadoras se encuentra por debajo las expectativas.*

Sobre estos tres riesgos se trató de cuantificarlos con el fin de poder ponderarlos dentro del resultado de nuestro de caso de negocio y se aplicaron las siguientes métricas:

Tipo de riesgo	Probabilidad	Impacto			Coeficiente de corrección		
		Inversión	Δ Ingresos	∇ Costes	Inversión	Δ Ingresos	∇ Costes
Algoritmo se coporta por debajo de las espectativas	40%	Muy Alto	Medio	Muy Bajo	0,10	0,02	0,01
No conseguir socio financiero en el primer año	20%	Muy Alto	Muy Alto	Muy Bajo	0,05	0,02	0,00
Captación de comercializadoras por debajo de nuestras espec	40%	Muy Bajo	Muy Alto	Muy Bajo	0,02	0,04	0,01

Principales indicadores del proyecto

Los principales indicadores del proyecto, teniendo también en cuenta el impacto de los principales riesgos identificados, quedarían del siguiente modo:

ROI: 424 %

Payback: Año 2

VAN: 976.270 €

TIR: 124 %

CONCLUSIONES

ARP se presenta como un proyecto innovador dentro del sector en el que se enmarca. Si bien es cierto que la industria del trading y la inversión es un ámbito en el que la tecnología hace tiempo que emergió con el fin de agilizar y maximizar los beneficios (trading algorítmico, HFT…). La tecnología del Deep Learning aplicada con este fin no ha sido utilizada hasta el momento de manera amplia y generalizada. Tal y como se ha descrito anteriormente en la memoria, tan solo algunas startups americanas han sido identificadas desarrollando esta tecnología en el campo de la inversión. Sin embargo, todavía se trata de un nicho a desarrollar en Europa y en especial, España.

Otro elemento diferenciador es el uso de información no tradicional en la obtención de predicciones de valores. La tecnología utilizada permite incorporar información procedente de noticias financieras, RRSS… e incorporarlas como una variable más dentro de nuestro análisis. Todo esto gracias al análisis de sentimientos y emociones que nos ofrece NLU. Históricamente esta información no era utilizada en los análisis algorítmicos tradicionales y se fundamentaba más en el análisis e intuición del inversor con las limitaciones que esto conlleva (volumen de información, sesgos personales…).

Como posibles líneas futuras de mejora y evolución, adicional a lo descrito en el roadmap, se han planteado diversas alternativas las cuales se describen a continuación:

- **Evolución de los modelos**: Evolución del modelo planteado mediante el uso de nuevas fuentes de información y la utilización de nuevas tecnologías con el fin mejorar la precisión y la varianza del resultado de nuestras predicciones.

- **Expansión geográfica**: Inicialmente el proyecto se ha focalizado en su implantación en España junto con un socio ya presente en el sector. No obstante, un modelo similar se podría implantar en otras regiones de Europa o Estados Unidos.
- **Evolución hacia un modelo B2C**: Como se describió en el posicionamiento de la solución, el modelo inicialmente planteado (B2B) tenía como fin podernos focalizar en la parte en la cual teníamos ya

conocimientos (desarrollo de modelos) y delegar el resto de los factores (regulación, captación de clientes…) en nuestro socio, ya presente en el mercado y con un mayor conocimiento del mismo. Lo que se plantearía en esta opción sería, una vez ya desarrollado, implantado y afianzado este modelo, evolucionar hacia un segundo de venta directa al cliente final, abordando desde nuestro rango de actuación todos los aspectos regulatorios y de captación sin necesidad de un socio. Este planteamiento solo sería viable habiendo demostrado al cliente final el buen funcionamiento de nuestra solución y habiendo creado una imagen de marca alrededor de ella.

- **Ampliación del modelo a derivados**: Una de las salidas naturales sería la aplicación del sistema de recomendación a la operativa con derivados. Aunque hay cambios conceptuales como la depositar garantías para relazar las operaciones de compra venta en el caso de los futuros no tendríamos problema para incluir este tipo de producto al ser los mismo también tendenciales y tener subyacentes basados entre otros en acciones y renta fija. En el caso de opciones si tendríamos que incluir modificaciones de más calado al tener las griegas de las opciones un parte importante en su valoración, no obstante las mismas seguirían teniendo como subyacente a activos que ya estamos considerando como antes hemos dicho. No obstante la inclusión de este tipo de activos nos llevaría a tener que realizar cambios en nuestros perfiles de riesgos de los clientes y nuestra manera de operar.