DIAPOSITIVA 1:

- TÍTULO

DIAPOSITIVA 2:

- Esta presentación está estructurada en una introducción, donde se explica el propósito de este trabajo, un marco teórico, en el que se desarrollan las herramientas necesarias para su implementación, una descripción matemática del problema, un esquema del aprendizaje por refuerzo aplicado a la cartera de valores, unos resultados, unas conclusiones y trabajo futuro.

DIAPOSITIVA 3:

- Una cartera de valores es un conjunto de activos financieros en los que se tiene invertido dinero de manera diversificada. Su valor viene dado por la siguiente ecuación, donde $p_{i,t}$ es el precio del activo i en el período de negociación t, y $A_{i,t}$ representa la cantidad de acciones compradas para el activo i en el periodo t.
- Por lo tanto, el proceso de gestión de carteras es un proceso secuencial y continuo que consiste en la compra-venta de activos para obtener el valor más alto para la cartera.
- Debido al carácter secuencial del mismo, es posible modelarlo como un proceso de Markov, resoluble mediante algoritmos de aprendizaje por refuerzo.

DIAPOSITIVA 4:

- El objetivo es aprender una estrategia para reorganizar de manera óptima los activos de la cartera utilizando aprendizaje por refuerzo profundo:
    - El aprendizaje profundo permite extrae características de los precios de mercado para determinar el potencial de crecimiento de un activo.
    - Mientras que el aprendizaje por refuerzo considera la naturaleza secuencial del proceso y los efectos que tiene la reorganización de los activos en la cartera para cada período de negociación

DIAPOSITIVA 5:

- El aprendizaje por refuerzo se puede resumir a través del siguiente esquema donde:
- El AGENTE interacciona con el AMBIENTE (que en este caso es el mercado), y tomala decisión o ACCIÓN que el considera mejor. Por cada ACCIÓN, el AGENTE recibe un REFUERZO, que será mayor cuanto mejor sea la ACCIÓN tomada.
- Cada ACCIÓN repercute en el AMBIENTE, cambiando su configuración o ESTADO.

DIAPOSITIVA 6:

- El aprendizaje profundo consiste en una subcategoría del aprendizaje automático que intenta modelar abstracciones de alto nivel en datos.
- Utiliza arquitecturas computacionales que admiten transformaciones no lineales, siendo la neurona su unidad básica de cómputo, cuyo esquema de funcionamiento aparece resumido en la siguiente imagen.

DIAPOSITIVA 7:

- Por lo tanto, una red neuronal profunda tiene varias capas de neuronas, lo que permite extraer características más complejas de los datos de entrada.

DIAPOSITIVA 8:

- En este caso se ha utilizado una red convolucional, la cual es capaz de extraer las características de los datos de entrada mediante operaciones de convolución aplicadas sobre un filtro cuyos parámetros se van actualizando a medida que la red se entrena.
- Una vez extraídas las características, la red saca una clasificación para cada variable utilizando una capa SOFTMAX.

DIAPOSITVA 9:

- El aprendizaje por refuerzo profundo combina redes neuronales profundas con algoritmos del aprendizaje por refuerzo, permitiendo resolver problemas de una mayor complejidad con un gran número (posiblemente infinito) de estados, que los algoritmos de aprendizaje por refuerzo tradicionales.
- Esto queda reflejado en la siguiente figura
- Por ejemplo, si se utiliza un algoritmo perteneciente única y exclusivamente al campo de aprendizaje por refuerzo, como puede ser Q-learning, y tenemos un espacio de estados y acciones discreto, podemos tener una representación de la función Q en forma de tabla. Utilizando esta tabla el agente computa una secuencia de acciones que eventualmente generarán la recompensa total máxima, siendo capaz de escoger la acción que obtiene los mejores resultados. Pero esto solo es posible para un número reducido de estados, ya que en caso contrario la tabla de la cual se extraería la recompensa total sería enorme.
- En cambio, si se utilizan redes neuronales profundas, el estado se proporciona como entrada y el Q-value de todas las acciones posibles se genera como salida, pudiendo abordar de esta forma problemas con un número mucho mayor de estados.

DIAPOSITIVA 10:

- Durante el período t, el precio de los activos de la cartera varía, cambiando el valor de esta de P_{t-1} a $P_t{}'$. Por lo tanto, al final de este período, el agente reorganiza la cartera comprando y vendiendo activos para tratar de conseguir el mayor beneficio posible. La compra-venta de activos implica modificar el vector de pesos $w_t{}'$, que pasa a ser w_t para la cartera optimizada. Teniendo en cuenta los costes de transacción, el valor final de la cartera se representa como P_t.

DIAPOSITIVA 11:

- Para obtener el máximo beneficio se definen dos indicadores de rentabilidad. La rentabilidad (compuesta y continua, ya que los intereses se reinvierten continuamente), y el Ratio de Sharpe.
- La rentabilidad se calcula a partir de la variación en el valor de la cartera con respecto al período anterior, que es función del cambio en el precio de los activos durante esta sesión ($\vec{y}_t$) y la composición de la cartera antes de reorganizar los activos ($\vec{w}_{t-1}$)

DIAPOSITIVA 12:

- Además, con el objetivo de ***maximizar el beneficio a largo plazo*** la función de refuerzo escogida es la media de un indicador de rentabilidad a lo largo de los períodos de negociación $\Delta t = t - t_0$

DIAPOSITIVA 13:

- El esquema del aprendizaje por refuerzo profundo aplicado al problema de gestión de carteras está compuesto por:
    - Un AGENTE, que se encarga de reordenar la cartera siguiendo una estrategia.
    - Un AMBIENTE o mercado, con el que el agente interactúa.
    - Una ACCIÓN, que contiene la nueva configuración de pesos decidida por el agente para los activos de la cartera.
    - Y un REFUERZO, o valor que recibe el agente por cada acción $\vec{a}_t$.

DIAPOSITIVA 14:

- El agente se encarga de encontrar la estrategia óptima al ser entrenado con el objetivo de maximizar la función de refuerzo R.
- El entrenamiento se lleva a cabo analizando los precios normalizados de los activos de los últimos $n = 50$ días.
- A este conjunto de precios se le denomina ESTADO, y representa la configuración del ambiente para t.

DIAPOSITIVA 15:

- Por lo tanto, el agente, o red neuronal, recibe un tensor estado, que analiza extrayendo el potencial de crecimiento de los activos.
- Este potencial de crecimiento se utiliza para generar una acción o distribución de pesos en la cartera.
- En concreto, la red empleada es la que aparece en la siguiente imagen, y consta de tres capas convolucionales y una encargada de generar los pesos de los activos, y que representa la ACCIÓN del agente.

DIAPOSITIVA 16:

- El entrenamiento se realiza en lotes de 20 días o períodos de negociación.
- Cada lote de datos empieza en un índice temporal aleatorio del periodo de entrenamiento.
- Los datos del tensor estado para cada lote tienen que estar en orden cronológico.
- Se utiliza el algoritmo ε-$greedy$ para asegurar que el agente explore acciones que a priori no hubiese probado.

DIAPOSITIVA 17:

- Se realizan dos entrenamientos:
- El primero utiliza datos de mercado desde principios de 2012 hasta Octubre de 2016, y para este existen tres períodos de prueba:
    - 2016/10/18-2018/05/24 (Octubre de 2016 a Mayo de 2018)
    - 2018/05/25-2019/12/30 (Mayo de 2018 a Diciembre de 2019)
    - 2016/10/18-2019/12/30 (Octubre de 2016 a Diciembre de 2019)
- El segundo utiliza datos de mercado desde principios de 2008hasta Mayo de 2009, y las pruebas se realizaron en los períodos:
    - 2009/05/27-2009/08/05 (Mayo de 2009 a Agosto de 2009)
    - 2009/08/6-2009/12/29 (Agosto de 2009 a Diciembre de 2009)
    - 2009/05/27-2009/12/29 (Mayo de 2009 a Diciembre de 2009)

- Aunque en la memoria se incluye un estudio detallado para cada período, en este caso se analizará únicamente el período más largo, a no ser que los resultados de los otros períodos aporten información nueva.
- Además, cabe destacar que, debido a que el agente necesita los precios de los 50 días anteriores para empezar a invertir, que el período de prueba empiece por ejemplo el día 18 de Octubre implica que el agente empieza a invertir el día 28 de Diciembre.

DIAPOSITIVA 18:

- Por lo tanto, los objetivos del trabajo son:
    - Estudiar la normalización de los datos introducidos en la red neuronal para su análisis.
    - Determinar que agente obtiene mejores resultados:
        - Agente 1: Maximiza la función objetivo media de las rentabilidades obtenidas para cada lote del proceso de entrenamiento.
        - Agente 2: Maximiza la función objetivo media de los Ratios de Sharpe.
    - Introducir el dinero en efectivo como un activo de la cartera de valores para que el agente aprenda a convertir los activos en dinero si el precio de estos cae.
    - Estudiar la adaptabilidad del agente a otros períodos.

DIAPOSITIVA 19:

- Primero se estudia qué normalización asegura mejores resultados. En concreto se analiza el período que va desde Octubre de 2016 a Mayo de 2018, de tal manera que:
    - La figura (a) muestra los resultados para un agente entrenado con precios normalizados por el precio de apretura de cada día.
    - Mientras que la figura (b) muestra los resultados de un agente entrando con precios normalizados por el precio de cierre más alto.
- De estas gráficas también se observa que el agente 1 es mejor que el agente 2, pero esto se discute en la siguiente sección.

DIAPOSITIVA 20:

- Para explicar las diferencias entre la ejecución del Agente 1 para las dos normalizaciones implementadas, se parte de la siguiente gráfica, que contiene la evolución de una cartera compuesta por un único activo.
- Esta gráfica permite extraer las acciones que el agente debería de tomar para obtener el mayor beneficio posible de tal manera que si, el valor de un activo baja, la proporción que ese activo ocupa en la cartera debería disminuir, aumentando la proporción de otro activo que justo en ese instante aumente de precio o por lo menos no disminuya.

DIAPOSITIVA 21:

- Analizando las acciones tomadas por el agente 1
    - (a) antes de entrenar
    - (b) después de entrenar con datos normalizados por el precio de apertura de cada día
    - (c) después de entrenar con datos normalizados por el mismo precio (precio de cierre más elevado)

- se ve que, la única diferencia con las acciones computadas en (b) y en (a) es que en (b) el agente asigna un peso fijo a cada activo captando únicamente la variación producía en cada sesión.
- En cambio, comparando (c) con (b), se ve que la primera gráfica sí que varía ligeramente los pesos de los activos (las líneas rectas de (b) se convierten en curvas en (c) en función de si el precio del activo ha subido con respecto al de hace n periodos de negociación)

DIAPOSITIVA 22:

- Por lo tanto, como conclusión se extrae que es mejor normalizar la serie temporal de precios de cada activo por un único precio (precio máximo de cierre para los períodos estudiados), ya que así la red consigue captar la variación en los precios del activo a largo plazo, y no únicamente la variación diaria.

DIAPOSITIVA 23:

- Una vez estudiada la normalización de los precios se pasa a determinar que agente ofrece mejores resultados. El estudio se lleva a cabo para el período Octubre de 2016 a Diciembre de 2019.
- De la figura de la izquierda se extrae que el Agente 1 obtiene unos beneficios mayores que los obtenidos por el Agente 2.
- La gráfica de la derecha ilustra el comportamiento de una cartera formada por un solo activo, y permite evaluar las acciones que debería de tomar el agente
- En la siguiente gráfica se comparan las acciones tomadas por cada uno de los agentes.

DIAPOSOTIVA 24:

- Mientras que el Agente 1 invierte más dinero en activos cuyo potencial de crecimiento es mayor, pero que a su vez asumen un mayor riesgo (por ejemplo, el activo SLR.MC), el Agente 2 invierte una cantidad mayor de dinero en activos cuya volatilidad es reducida (por ejemplo GRF.MC), con el fin de evitar pérdidas mayores, pero desaprovechando la posibilidad de obtener beneficios más elevados.

DIAPOSITIVA 25:

- La conclusión es que el Agente 2 reduce ligeramente las pérdidas, pero evita volatilidades que pueden generar rentabilidades positivas elevadas ya que lo que maximiza es el Ratio de Sharpe, que mide la rentabilidad obtenida por unidad de Riesgo.

DIAPOSITIVA 26:

- Con el objetivo de conseguir un agente que obtenga el mayor beneficio posible de las volatilidades que generan rentabilidades positivas y a su vez evitar pérdidas importantes, se crea un segundo modelo, al que se va a denominar Modelo 2, que incluye el dinero efectivo como activo sin riesgo de la cartera.
- En concreto, se pretende que, si todos los activos conllevan a pérdidas, el agente sepa venderlos, quedándose con su valor monetario.
- El estudio se realiza en el período Octubre de 2016 a Diciembre de 2019

DIAPOSITIVA 27:

- Aunque en principio los beneficios son inferiores a los obtenidos mediante el primer modelo, la evolución de las acciones del agente, plasmada en la gráfica inferior, parece mucho más dinámica que la computada para el modelo 1.
- Un ejemplo sencillo se puede ver estudiando la evolución del activo SLR.MC. Mientras que el valor para este activo crece en la gráfica superior, su peso asignado por el agente, el cual se ve en la gráfica inferior, es muy elevado, siendo el mayor de todos. No obstante, para el período comprendido entre Mayo y Septiembre de 2018, el valor del activo cae, al igual que su peso, el cual se ve superado por el peso del activo AIR.MC, ya que en este período dicho activo crece.

DIAPOSITIVA 28:

- Como conclusiones más salientables tenemos que:
- Por un lado, el agente es capaz de captar mejor las subidas y bajadas en los precios de los activos, pero tarda un poco en adaptarse a la volatilidad del mercado, acarreando pérdidas
- Además, debido a que la rentabilidad del dinero efectivo no invertido es muy pequeña, la proporción de efectivo en la cartera nunca es muy elevada.

DIAPOSITIVA 29 Y 31:

- A continuación, se pasa a estudiar la adaptabilidad de los Modelos 1 y 2 a los períodos, que van desde Agosto de 2009 hasta Diciembre de 2009, y desde Mayo de 2009 hasta Diciembre de 2009.
- Comparando el Modelo 1 (que no incluye el efectivo) con el Modelo 2 (que sí que lo incluye), se ve que tanto para esta diapositiva, como para la diapositiva 31, que el Modelo 2 es mucho mejor que el 1.
- La explicación se puede ver bien en las diapositivas 30 y 32

DIAPOSITIVA 30 Y 32:

- Mientras que para el modelo 1 la configuración de los pesos es muy parecida a aquella asignada a los períodos comprendidos entre 2016-2020, el Modelo 2, al ser más dinámico, es capaz de cambiar notablemente esta configuración, evitando el sobre-ajuste que produce el modelo 1.

DIAPOSITIVA 33:

- Por lo tanto, el agente que incluye el efectivo es capaz de adaptarse mejor a los períodos en donde el comportamiento de los activos es muy diferente al visto durante el entrenamiento 2012-2016.
- En cambio, el agente que no incluye el efectivo sobre-ajusta los pesos de los activos al conjunto de entrenamiento.
- Esto no se veía tan claramente en los períodos que van de 2016 a 2020 debido a que los activos tienen un comportamiento similar al visto durante el entrenamiento.

DIAPOSITIVA 34:

- Con la intención de que la red mejore los resultados obtenidos, eliminando posibilidades de sobre-ajuste se re entrena partiendo de los parámetros de la red obtenidos durante el primer entrenamiento.

- El período que va desde principios de 2008 hasta Mayo de 2009 es el que se ha escogido para reentrenar la red ya que el comportamiento de los activos durante este período es muy diferente al visto en el anterior.

DIAPOSITIVA 35:

- El agente del Modelo 1, consigue mejorar sus resultados para los períodos comprendidos entre Agosto de 2009 y Diciembre de 2009, obteniendo resultados ligeramente mejores que los del agente antes de ser entrenado, y más o menos mantiene su ejecución para los períodos comprendidos entre Octubre de 2016 y Diciembre de 2019.
- En cambio, el Agente 2 sí que consigue mejorar notablemente su ejecución en ambos conjuntos. A continuación, se incluyen las mejoras más salientables.

DIAPOSITIVA 36 Y 37:

- El agente del Modelo 2 para el período comprendido entre Agosto de 2009 y Diciembre de 2009 consigue mejorar su ejecución, pero no evitar pérdidas en el valor final. Asimismo, mejora sus resultados para períodos pertenecientes al ciclo 2016-2020. Un ejemplo de esto aparece en la siguiente diapositiva.

DIAPOSITIVA 38:

- Debido a que la mejora se produce en ambos períodos, se puede concluir con el hecho de que el agente no está sobreajustando los parámetros.
- Además, como el Modelo 2 presenta posibilidades de mejora, este se vuelve a reentrenar en 30 lotes más del período que va desde principios de 2008 hasta Mayo de 2009.
- La intención es que consiga captar las variabilidades de este período ya que, se trata de una época económicamente inestable.

DIAPOSITIVA 39:

- Para el período que va de Agosto de 2009 hasta Diciembre de 2009, el agente mejora la ejecución lo suficiente como para evitar pérdidas en el valor final

DIAPOSITIVA 40:

- La misma mejora amplificada se puede ver para el período que comprende desde Octubre de 2016 hasta Diciembre de 2019.
- Sin embargo, en el período que va desde Mayo de 2018 hasta Diciembre de 2019, aunque el agente acaba ganando dinero, acarrea unas pérdidas muy elevadas alrededor de Noviembre de 2018.
- La explicación se puede ver en la siguiente diapositiva.

DIAPOSITIVA 42:

- La principal diferencia que se observa a la hora de reentrar es que el agente aprende a invertir mayores cantidades de dinero en los activos con potenciales de crecimiento muy elevados, siendo capaz de reducir esta cantidad drásticamente cuando el activo sufre una caída importante en su precio (por ejemplo, en el caso de SLR.MC).

- No obstante, el agente no es lo suficientemente dinámico como para evitar pérdidas, ya que no consigue vender las acciones de ese activo con la rapidez suficiente como para no perder dinero.
- Por ejemplo, durante el período que va desde Mayo de 2018 hasta Diciembre de 2019, en las fechas próximas a Noviembre de 2018, el activo SLR.MC tiene picos de subidas y bajadas bastante importantes a los que el agente no es capaz de adaptarse con la rapidez necesaria, aunque sí que sigue un patrón similar en los picos del peso de este activo.

DIAPOSITIVA 43:

- Por lo tanto, este trabajo de fin de máster permite obtener la siguientes conclusiones:
    - Por un lado, el problema puede ser modelizado como un proceso de decisión de Markov, ya que los resultados obtenidos son relativamente buenos dependiendo del período y son justificables.
    - Además, el Agente 1 es mucho mejor que el Agente 2, aunque no consiga gestionar bien el riesgo para casos determinados.
    - El hecho de introducir el dinero en efectivo como activo sin riesgo de la cartera produce mayor dinamismo a la hora de reorganizar la cartera.
    - Debido a que el agente es más dinámico, se adapta mejor a otros períodos de comportamiento muy dispar con el visto por el agente.

DIAPOSITIVA 44:

- Como futuro trabajo se plantea, además del análisis de otras arquitecturas para la red neuronal, el estudio de una política de gestión de riesgos, y la investigación sobre cómo mejorar el *vanishing gradient problem*, que dificulta el proceso de entrenamiento.