



Universidad Politécnica
de Madrid



**Escuela Técnica Superior de
Ingenieros Informáticos**

Grado en matemáticas e informática

Trabajo Fin de Grado

**Aplicación de algoritmos genéticos a la
optimización de carteras**

Autor: Sofía Hernández Montero

Tutor(a): Luis Magdalena Layos

Madrid, junio 2022

Este Trabajo Fin de Grado se ha depositado en la ETSI Informáticos de la Universidad Politécnica de Madrid para su defensa.

Trabajo Fin de Grado

Grado en matemáticas e informática

Título: Aplicación de algoritmos genéticos a la optimización de carteras

Junio 2022

Autor: Sofía Hernández Montero

Tutor:

Luis Magdalena Layos

Matemática aplicada a las tecnologías de la información y comunicaciones

ETSI Informáticos

Universidad Politécnica de Madrid

Resumen

En este trabajo se ha resuelto el problema de optimización de carteras financieras haciendo uso de optimización multiobjetivo y algoritmos genéticos. Para ello, se comienzan explicando algunos conceptos fundamentales, así como las investigaciones que se han hecho hasta el momento relacionadas con el tema. Entre ellas destacan proyectos muy similares a lo que se quiere implementar en este trabajo, pero analizando datos, tipos de carteras o modelos multiobjetivo muy concretos.

Desde este punto de partida, se ha procedido a desarrollar el problema genético y, posteriormente, implementar el código del algoritmo en Python. En el área del problema genético se describen todos los aspectos fundamentales para ser capaz de comprender como funciona el código sin necesidad de leerlo. El fin del algoritmo que se ha creado es encontrar las carteras que consigan maximizar su rentabilidad con el mínimo riesgo posible. Así, la función objetivo a optimizar por el algoritmo va a tener dos componentes que se encuentran en conflicto: maximizar rentabilidad y minimizar riesgo. Es por esto por lo que se habla de problema multiobjetivo, el cual se formula antes de empezar a explicar el algoritmo de forma más concreta.

Como bien es característico de la optimización multiobjetivo, la salida principal del algoritmo es un conjunto de Pareto. En este tipo de funciones no solo hay una solución válida, ya que los valores de las variables que obtengan los mejores resultados para cada uno de los objetivos van a ser diferentes. Es por esto por lo que, en lugar de buscar una única solución que genere el mejor valor para un solo objetivo, se va a tratar de hallar un conjunto de soluciones buenas que tengan en cuenta todos los criterios. La idea de óptimo de Pareto define toda situación en la que no es posible optimizar más una función sin perjudicar a la evaluación de otra. Es un punto donde todos los criterios se encuentran en equilibrio. Dadas dos soluciones a y b, si a no es peor que b y, a la vez, b es no es peor que a, entonces a y b son igual de buenas. Todas las soluciones que no son superadas por otras son llamadas soluciones no dominadas. Una solución domina a otra si es mejor o igual en todos los objetivos y mejor en al menos uno de ellos. Así, se entiende como conjunto de Pareto a todas las soluciones no dominadas.

Además, el algoritmo también genera una serie de gráficos que muestran cómo han evolucionado las soluciones a lo largo de las distintas iteraciones.

Por otro lado, se escogen una serie de datos para utilizar como ejemplo para probar el modelo creado. Todos estos datos son valores históricos de determinados títulos en un periodo de un año. Los datos relativos a los primeros seis meses de dicho año se utilizan para que el algoritmo haga una predicción de cuáles serían las mejores inversiones posibles. Después, los datos correspondientes a los siguientes seis meses se usan para comprobar que hubiera ocurrido si se hubiera seguido alguna de las soluciones encontradas.

Una vez se tiene toda esta información acerca de los resultados que se han conseguido con el algoritmo genético, se hace un análisis para evaluar algunos puntos que se habían definido como objetivo del trabajo. Dichos objetivos son los siguientes:

1. Conseguir clasificar las soluciones obtenidas en función de distintos tipos de inversor.
2. Analizar si el índice de Sharpe, uno de los indicadores más utilizados en el ámbito financiero para establecer como de buena es una cartera en función de su rentabilidad y riesgo, discrimina carteras hacia un tipo de perfil de inversor en concreto.

Abstract

In this paper there is solved the problem of financial portfolio optimisation using multi-objective optimization and genetic algorithms. To do so, I begin by explaining some fundamental concepts, as well as the research that has been done so far on the subject. These include projects very similar to what I want to implement in this work, but analysing very specific data, types of portfolios or multi-objective models.

From this starting point, the genetic problem has been developed and, subsequently, the algorithm code has been created in Python. In the area of the genetic computing, all the fundamental aspects are described to be able to understand how the code works without having to read it. The aim of the algorithm that has been created is to find the portfolios that manage to maximise their profitability with the minimum possible risk. Thus, the objective function to be optimised by the algorithm will have two conflicting components: maximising profitability and minimising risk. Therefore, I speak of a multi-objective problem, which is formulated before starting to explain the algorithm more in depth.

As is characteristic of multi-objective optimisation, the main output of the algorithm is a Pareto set. In this type of functions there is not only one valid solution, since the values of the variables that obtain the best results for each of the objectives will be different. Therefore, instead of looking for a single solution that generates the best value for a single objective, I will try to find a set of good solutions that consider all the criteria. The idea of Pareto optimum defines any situation in which it is not possible to further optimise one function without impairing the evaluation of another. It is a point where all criteria are in equilibrium. Given two solutions a and b, if a is not worse than b and, at the same time, b is not worse than a, then a and b are equally good. All solutions that are not outperformed by others are called non-dominated solutions. A solution dominates another if it is better or equal in all objectives and better in at least one of them. Thus, a Pareto set is understood as all non-dominated solutions.

In addition, the algorithm also generates a series of graphs that show how the solutions have evolved throughout the different iterations.

Furthermore, a set of data has been picked to be used as an example to test the model. All these data are historical values of certain securities over a period of one year. The data for the first six months of that year are used for the algorithm to make a prediction of what would be the best possible investments. The data of the next six months is then used to see what would have happened if any of the solutions found had been followed.

Once the results achieved with the genetic algorithm are available, an analysis has been carried out to evaluate some of the points that had been defined as the objective of the work. These objectives are as follows:

1. To be able to classify the solutions obtained according to different types of inverters.
2. To analyse whether the Sharpe ratio discriminates portfolios towards a particular type of investor profile. This index is one of the most widely used indicators in the financial sphere to establish how good a portfolio is in terms of its profitability and risk.

Tabla de contenidos

| | | |
|----------|--|-----------|
| 1 | Introducción | 1 |
| 2 | Marco teórico | 2 |
| 2.1 | Gestión de carteras | 2 |
| 2.1.1 | Perfil de inversor | 2 |
| 2.1.2 | Concepto de diversificación de una cartera | 3 |
| 2.1.2.1 | Concepto de rentabilidad de una cartera..... | 3 |
| 2.1.2.2 | Concepto de riesgo..... | 3 |
| 2.1.3 | Tipos de cartera de inversión según el perfil de inversor..... | 4 |
| 2.1.4 | Tipos de estrategias según la cartera y el perfil del inversor | 5 |
| 2.1.5 | Evaluación de una cartera | 5 |
| 2.2 | Concepto de optimización..... | 6 |
| 2.2.1 | Optimización multiobjetivo..... | 7 |
| 2.2.1.1 | Optimalidad, conjunto y frente de Pareto | 7 |
| 2.3 | Optimización mediante algoritmos genéticos | 8 |
| 2.3.1 | Conceptos de computación evolutiva..... | 8 |
| 2.3.2 | Algoritmos genéticos | 9 |
| 2.3.2.1 | Representación de un individuo..... | 10 |
| 2.3.2.2 | Métodos de selección | 10 |
| 2.3.2.3 | Proceso de reproducción | 11 |
| 2.3.2.4 | Mecanismos de inserción | 11 |
| 2.3.2.5 | Criterios de parada | 12 |
| 3 | Estado del arte | 13 |
| 4 | Planteamiento y formulación del problema | 15 |
| 4.1 | Diseño del sistema: entradas y salidas | 15 |
| 4.1.1 | Tratamiento y justificación de los datos iniciales..... | 16 |
| 4.1.1.1 | Justificación y obtención de datos elegidos | 16 |
| 4.1.1.2 | Tratamiento de los datos..... | 17 |
| 4.2 | Planteamiento del problema genético..... | 17 |
| 4.2.1 | Representación de los individuos | 17 |
| 4.2.2 | Formulación del problema multiobjetivo | 18 |
| 4.2.3 | Método de selección del algoritmo | 19 |
| 4.2.4 | Proceso de reproducción del algoritmo | 19 |
| 4.2.5 | Mecanismo de inserción del algoritmo y mutaciones | 20 |
| 4.2.6 | Criterios de parada del algoritmo | 20 |
| 5 | Resultados y conclusiones | 21 |
| 5.1 | Ejemplo de salidas del sistema | 21 |
| 5.2 | Clasificación de soluciones para distintos tipos de inversor | 24 |
| 5.3 | Análisis del índice de Sharpe respecto a las soluciones obtenidas | 29 |

| | |
|---|-----------|
| 5.4 Evolución de la rentabilidad de las carteras seleccionadas con el algoritmo genético..... | 30 |
| 5.5 Conclusión general y posibles mejoras | 31 |
| 6 Análisis de Impacto | 33 |
| 7 Bibliografía | 34 |

1 Introducción

Hoy en día, la preocupación de las personas por saber invertir su dinero correctamente cada vez aparece a más temprana edad. La democratización de las finanzas es un proceso que busca descentralizar la industria, de tal forma que la mayor cantidad de personas posible tengan acceso a numerosas opciones para administrar su propio dinero. El periódico El País en 2019 lo define como “el boom de la gestión de carteras” [1]. Así, cada vez son más recurrentes las preguntas acerca de cómo crear y gestionar una cartera financiera.

El año pasado hice prácticas en una gestoría de fondos de inversión, lo que hizo que este tipo de temas captaran mi atención. Además, me planteaba bastante que posible aplicación tendría lo que yo había estudiado en la carrera en esta área. Hasta ese momento habíamos estudiado distintos algoritmos y formas de resolver problemas, pero pocas veces aplicados a la vida real, y particularmente nunca al mundo económico.

De esta forma, en este trabajo se busca aplicar un tipo concreto de algoritmos, los algoritmos genéticos, a los problemas de gestión de carteras. El objetivo es conseguir desarrollar un algoritmo que optimice este proceso, ofreciendo para cada tipo de inversor la mejor cartera posible.

2 Marco teórico

2.1 Gestión de carteras

Se conoce como cartera de inversión a un conjunto de activos financieros que pertenecen a un inversor o sociedad [2]. Dicha cartera puede ser de muchos tipos dependiendo del perfil del inversor. Estos tipos reflejan factores tales como la tolerancia al riesgo, la edad, el horizonte de inversión, la situación personal y profesional, y la estrategia de inversión [3].

2.1.1 Perfil de inversor

El problema de la selección de carteras se basa en gestionar los activos con la mejor estrategia posible para maximizar la rentabilidad con el mínimo riesgo. Para ello no existe un único modelo. El tipo de gestión que cada inversor va a utilizar depende de sus objetivos y una serie de circunstancias personales. A esto se le denomina perfil del inversor, el cual se basa en las siguientes características:

1. La edad. Por lo general una persona más joven siempre puede asumir más riesgos. Se asume que alguien de mayor edad con una familia o hipoteca va a tener más temor de invertir en activos con un mayor riesgo.
2. La situación personal y profesional, así como el horizonte de inversión. Un inversor en una situación profesional estable y sin deudas tendrá mayor disposición a invertir en activos de riesgo. De la misma forma, hay que valorar el instante temporal en el que se encuentra cada persona, es decir, cual es el periodo de recuperación de la inversión (payback) que necesita.
3. En tercer lugar, está la tolerancia al riesgo y la comodidad de inversión. Hay estrategias de inversión que requieren de mucha más dedicación que otras, como es el caso de la gestión activa, de la cual se hablará posteriormente. Hay que tener en cuenta para cada inversor cuánto tiempo está dispuesto a invertir en investigar por sí mismo, o si por el contrario prefiere una estrategia que exija menos dedicación personal [3].
4. Por último está la diversificación, que es la estrategia de invertir capital en diferentes activos para reducir el riesgo de que uno de los activos no funcione tan bien como se estimaba. De esta forma, una cartera diversificada permite elegir más oportunidades de inversión y obtener mejores rendimientos [4].

Una vez bien definido el perfil del inversor, hay que evaluar qué estrategias son las más adecuadas, así como definir cuándo es pertinente hacer un rebalanceo de la cartera, lo cual consiste en reajustar el peso de los distintos activos. Para ello es necesario saber en qué tipo de activos se quiere invertir. Si hay baja aversión al riesgo se puede invertir en activos con rentabilidad esperada elevada, pero con un riesgo alto. Por ejemplo, fondos de inversión cotizados (ETFs, del inglés “exchange-traded funds”) del sector de la biotecnología. Por otro lado, las personas con alta aversión al riesgo deben optar por activos de menor riesgo y de menor rentabilidad esperada, como el oro o

bonos de estado. En todo caso, la estrategia siempre ha de tener en cuenta el conjunto de todos los aspectos mencionados que determinan el perfil de inversor.

2.1.2 Concepto de diversificación de una cartera

La diversificación de carteras es un concepto estratégico de reducción de riesgos. Se basa en distribuir el capital invertido en distintos activos con características distintas, es decir, que están muy poco o nada correlacionados entre sí. De esta forma, el riesgo global de la cartera es inferior a la suma del riesgo de cada uno de los activos de los cuales está compuesta.

Para comparar activos hay que tener en cuenta dos características fundamentales: la rentabilidad y el riesgo.

2.1.2.1 Concepto de rentabilidad de una cartera

La rentabilidad hace referencia a los beneficios que se han obtenido o se pueden obtener de una inversión [5]. En primer lugar, hay que establecer una distinción entre la rentabilidad real y la esperada. Cuando se habla de la rentabilidad de un activo siempre se realiza una suposición. Si estamos en el contexto de que está suposición siempre se cumple, el rendimiento que se espera obtener coincide con el que se obtiene finalmente, es decir, la rentabilidad real coincide con la esperada. Sin embargo, esto no suele ocurrir así, debido a la existencia del riesgo. El inversor necesita poder dar valores al rendimiento y al riesgo de cada activo para poder tomar una decisión sobre sus inversiones [6].

La rentabilidad económica de un fondo sigue la siguiente fórmula:

$$\text{Rentabilidad económica} = \frac{\text{Beneficio antes de interés e impuestos (BAII)}}{\text{Activo total que genera renta al fondo}} * 100$$

Ecuación 1: rentabilidad económica [6]

2.1.2.2 Concepto de riesgo

Se denomina riesgo de un activo a la probabilidad que tiene este de no cumplir con lo esperado. Para evaluar el riesgo de un activo se utilizan medidas estadísticas, más concretamente, la desviación típica, que es la medida de dispersión de un dato respecto de la media. Cuanto más alto sea el valor de la desviación típica, mayor será también el riesgo de este título [7].

La desviación típica depende de la varianza, y formalmente se calcula de la siguiente forma. La varianza es la suma de los cuadrados de las distancias de cada observación respecto de la media, dividido entre el total de observaciones. La desviación típica es igual a la raíz cuadrada de la varianza.

$$\text{Var}(X) = \sigma^2 = \frac{\sum_1^n (x_i - \bar{x})^2}{n} \quad \rightarrow \quad \sigma = \sqrt{\sigma^2}$$

Ecuación 2: riesgo de un activo [7]

siendo x la variable sobre la cual se quiere calcular la varianza, i el número de observación, \bar{x} la media aritmética de la muestra y n el número total de observaciones.

Una vez estudiado el riesgo o volatilidad de cada activo de forma individual, se puede calcular el de la cartera.

El riesgo de una cartera no se calcula como la media ponderada de todos sus títulos, sino que hay otro valor que entra en juego: la correlación entre activos. Sigue la siguiente fórmula matemática:

$$\sigma_{\text{cartera}} = \sqrt{\sum_{i=1}^n x_i^2 \sigma_i^2 + \sum_{i=1}^n \sum_{j=1}^n x_i x_j \sigma_{i,j}} \quad \sigma_{i,j} = \frac{\sum_{k=1}^n (i_k - \bar{i})(j_k - \bar{j})}{n}$$

Ecuación 3: riesgo de una cartera [7]

donde x_i y σ_i son la ponderación y el riesgo individual de cada activo, respectivamente. $\sigma_{i,j}$ es la covarianza (grado de relación) entre los activos i y j . En su respectiva expresión matemática, i y j son los activos, \bar{i} y \bar{j} las medias aritméticas entre todas sus observaciones, y n el número total de observaciones.

2.1.3 Tipos de cartera de inversión según el perfil de inversor

En función de factores observados en el perfil del inversor, tales como la tolerancia al riesgo o el horizonte de inversión, se definen distintos tipos de carteras [2] [8].

1. Conservadora o de ahorro. La composición de la cartera prioriza activos con poco riesgo, invirtiendo la mayor parte del capital en valores de renta fija, donde se generan intereses de forma regular y hay rentabilidad desde el comienzo. El objetivo es maximizar la seguridad del capital.
2. Moderada. Se crea la cartera con una mezcla de activos de renta fija y variable. Así, se busca aumentar el crecimiento de la cartera un poco más frente a la opción anterior, pero existe una gran diversificación de los activos para disminuir el riesgo.
3. De crecimiento o de inversión. La cartera se crea de forma parecida a la anterior, pero priorizando aún más los activos de renta variable. Sin embargo, se sigue manteniendo gran parte del capital en activos de menor riesgo para que su rentabilidad pueda compensar las pérdidas generadas por inversiones más volátiles.
4. De crecimiento agresivo o de renta variable. La mayor parte de activos que componen esta cartera son de renta variable. Es un tipo de cartera con una rentabilidad muy elevada, pero también se asume un gran riesgo.

2.1.4 Tipos de estrategias según la cartera y el perfil del inversor

Cuando un inversor ha escogido la cartera más adecuada para él en función de sus intereses, se pasa a seleccionar una estrategia para gestionarla. Según todos los parámetros y características definidas anteriormente, hay distintas estrategias para gestionar carteras:

1. Gestión activa: esta estrategia tiene como objetivo lograr rentabilidades superiores a los índices bursátiles basándose en una selección estricta y una gestión activa de los activos. A cambio cobra comisiones más elevadas que en el caso de las estrategias pasivas. En este tipo de planteamiento los inversores tratan de predecir los movimientos del mercado, es decir, intentar anticiparse a lo que puede ocurrir.
2. Gestión pasiva: esta estrategia se basa en replicar índices. Requiere menor dedicación y a largo plazo puede ofrecer rentabilidades. No utiliza técnicas de análisis bursátil como en el caso anterior. Un ejemplo del uso de este tipo de gestión de cartera es un fondo indexado.
3. Gestión mixta: combina las 2 anteriores para diversificar entre activos y estrategias de inversión.

Una vez analizada toda esta información, lo último que hay que hacer para tener la cartera terminada es seleccionar los activos y saber cómo y cuándo hacer un rebalanceo, es decir, reajustar la composición de los activos según las necesidades actuales del inversor. Se puede dar la situación de que un activo no está cumpliendo con las rentabilidades objetivo esperadas, o que las necesidades iniciales del inversor han variado. Para ello existen varias formas, por ejemplo, reforzar las posiciones con pérdidas: si se considera que una empresa en la que se está invirtiendo tiene buena perspectiva de futuro, lo lógico es comprar más activos para así poder abaratar el coste medio de adquisición de los títulos de esa empresa. También se pueden reforzar posiciones en auge, o vender posiciones para comprar otras. Por lo general, en el sector se considera que hay que hacer un rebalanceo una o dos veces al año, pero depende siempre del riesgo de cada cartera específica.

2.1.5 Evaluación de una cartera

En el momento en el que se considere pertinente hacer un rebalanceo de la cartera, primero hay que evaluarla para ver en qué punto se encuentra. Para ello los inversores utilizan una serie de índices. Uno de los más comunes es el índice de Sharpe, el cual mide la relación entre la rentabilidad y la volatilidad de la cartera en un determinado periodo de tiempo. Cuando mayor es este índice, mejor es la rentabilidad del título en relación con el riesgo que se ha tomado. Se define mediante la siguiente expresión matemática:

$$\text{Índice de Sharpe} = \frac{\text{Rentabilidad del fondo} - \text{Rentabilidad del activo libre de riesgo}}{\text{Desviación estandar de la rentabilidad del fondo (riesgo)}}$$

Ecuación 4: índice de Sharpe [9]

2.2 Concepto de optimización

Para comprender la realidad de forma más esquemática y simplificada se hace uso de modelos [10]. Dentro de la ciencia hay muchos tipos de modelos: físicos, matemáticos, gráficos, etc. Se utilizan para representar el comportamiento de un determinado sistema, y así poder formalizar experimentos sobre un proceso.

Dependiendo del tipo de modelo hay distintos objetivos. Un modelo de control pretende decidir qué nuevas variables deben amoldar para conseguir un resultado concreto. Por otro lado, un modelo de optimización de un determinado problema se basa en encontrar los valores de los parámetros para los cuales el resultado es el mejor posible. Por ejemplo, en uno de los casos más simples, se puede buscar maximizar o minimizar una función en concreto.

Un problema de optimización se compone de una función objetivo, que permite evaluar la calidad de cualquier posible solución, y un espacio de búsqueda, que es el ámbito en el cual se buscarán las posibles soluciones. Para determinar el ámbito de búsqueda hay que tener en cuenta el grado de restricción del problema, en lo que influyen los factores siguientes. Primero, el número de restricciones. Una restricción es una limitación de una variable. Por ejemplo, que una incógnita tome valores positivos, o que la suma de varias variables no sobrepase un valor en concreto. Además, también influye como de difícil es dar con una solución que satisfaga el problema. Por último, hay que tener en cuenta la dimensión del espacio de búsqueda, así como la dimensión del problema. Esta se define según el número de variables de decisión, el rango de valores que estas pueden tomar y el tamaño del espacio de búsqueda.

Por otro lado, una solución viene definida por el valor de la función objetivo cuando las variables toman unos valores determinados. Esta solución puede ser de varios tipos:

1. Factible: cuando satisface todas las restricciones y, por tanto, las incógnitas del problema están dentro del espacio de búsqueda.
2. Óptima: es una solución factible que genera el mejor resultado posible de la función objetivo.
3. Subóptima: es también una solución factible que hace que la función objetivo tome un buen valor, pero no el mejor.

Haciendo referencia al tipo de cuestión que se está evaluando, los problemas de optimización se clasifican en función del número de soluciones factibles que se pueden encontrar. Así, los problemas pueden ser continuos o combinatorios. Los primeros son aquellos que cuentan con un número infinito de soluciones factibles y suelen trabajar con variables reales. En cambio, en los combinatorios el número de soluciones factibles es finito y se suelen considerar variables de carácter discreto.

Para la resolución de estos problemas se puede hacer uso de técnicas heurísticas y metaheurísticas. Las primeras son procedimientos que utilizan la aproximación intuitiva y el sentido común. Son muy eficientes y funcionan de manera muy rápida, aunque no garantizan encontrar la solución óptima. Sus principales características son la eficiencia, la bondad y la robustez. Se utilizan cuando no existe un método exacto de resolución, o cuando este requiere de

demasiado tiempo para llegar a la solución óptima [11]. Por otra parte, las técnicas metaheurísticas son también una clase de métodos aproximados que no garantizan llegar a la solución óptima. Lo que las diferencia de las técnicas heurísticas es que la búsqueda se orienta en función de la evolución que ha seguido el proceso. Es decir, se intenta huir de óptimos locales. Estas técnicas se utilizan especialmente en procesos de optimización combinatoria, sin embargo, se pueden aplicar a una gran variedad de problemas [12].

2.2.1 Optimización multiobjetivo

En la vida real, los problemas que se van a someter a un proceso de optimización suelen contar con más de un criterio de evaluación. Es decir, se requiere la optimización simultánea de más de un objetivo y, por lo general, todos estos se encuentran en conflicto. Por tanto, los valores de las variables que obtengan los mejores resultados para cada uno de los objetivos van a ser diferentes. Es por esto por lo que en este contexto se necesita redefinir el concepto de solución óptima. En lugar de buscar una única solución que genere el mejor valor para un solo objetivo, se va a tratar de hallar un conjunto de soluciones buenas que tengan en cuenta todos los criterios.

Además, se deben conocer los conceptos de espacio de búsqueda y espacio objetivo. Representan el dominio y el recorrido de las funciones a optimizar, respectivamente. Por otro lado, la dimensión del espacio objetivo coincide con el número de objetivos que se desea optimizar. Un multiobjetivo puede tener múltiples funciones, o una única función vectorial.

Este tipo de optimización se asocia a la teoría de decisión multicriterio (MCDM - multi criteria decision making) [13]. Se basa en la toma de decisiones en problemas de mayor dificultad, que requieren de evaluaciones que están en conflicto. En el desarrollo de esta teoría se destaca la investigación de Benjamín Franklin, Marqués de Condorcet, Georg Cantor y Vilfredo Pareto. Este último tuvo particular importancia, ya que fue quien introdujo el concepto de eficiencia, conocido como el óptimo de Pareto.

2.2.1.1 Optimalidad, conjunto y frente de Pareto

La idea de óptimo de Pareto define toda situación en la que no es posible optimizar más una función sin perjudicar a la evaluación de otra. Es un punto donde todos los criterios se encuentran en equilibrio. Dadas dos soluciones a y b , si a no es peor que b y, a la vez, b no es peor que a , entonces a y b son igual de buenas. Todas las soluciones que no son superadas por otras son llamadas soluciones no dominadas. Una solución domina a otra si es mejor o igual en todos los objetivos y mejor en al menos uno de ellos.

Se entiende como conjunto de Pareto a todo el conjunto de puntos del espacio de búsqueda que sean las mejores soluciones para el problema. Por otro lado, se denomina frente de Pareto al conjunto de soluciones óptimas en el espacio objetivo.

Una solución pertenece al frente de Pareto si no puede encontrarse otra que mejore alguno de los objetivos sin empeorar otro.

Hay diferentes formas para llegar al frente de Pareto. La más utilizada es el método de las ponderaciones. Consiste en otorgar un peso a cada una de las funciones objetivas y construir una nueva función escalar que sea la suma de todas ellas. Para cada ponderación posible se llega a un problema escalar donde hay que minimizar la función resultante.

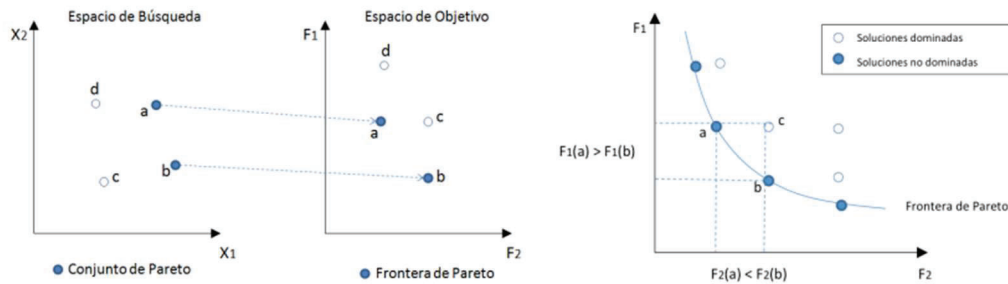


Ilustración 1: representación del espacio de búsqueda y espacio objetivo (a la izquierda). Representación del frente de Pareto y dominancia de soluciones (derecha) [3].

2.3 Optimización mediante algoritmos genéticos

Un algoritmo genético es un algoritmo metaheurístico inspirado en la biología y en la reproducción de poblaciones. Su aplicación se basa en la computación evolutiva [10].

2.3.1 Conceptos de computación evolutiva

La computación evolutiva es, como los algoritmos genéticos, una metaheurística basada en poblaciones e inspirada en la biología. Surge en la década de 1930 [10] y su planteamiento se centra en replicar el proceso de reproducción que se observa en la evolución natural. Consiste en partir de una población de individuos diferentes que son capaces de reproducirse entre ellos. Los descendientes conservan características de los padres y, así, son también distintos entre sí. Además, cada individuo tiene una evaluación en relación con la función objetivo, y es considerado mejor cuanto mejor sea su evaluación.

En cada generación, los mejores individuos tienen una mayor probabilidad de reproducirse. Así, siempre se evoluciona hacia los mejor adaptados, dejando las peores soluciones atrás, imitando el proceso de selección natural enunciado por Chales Darwin en 1805 en su libro “El origen de las especies”. Cabe destacar que cada generación siempre está compuesta por el mismo número de individuos. Cuando se dice que se dejan las peores soluciones atrás, se quiere decir que si un individuo tiene una evaluación muy mala, tendrá una probabilidad baja de reproducirse. Así, es poco frecuente que esos genes se transmitan a la siguiente generación. Finalmente, al crear la nueva generación hay diferentes mecanismos de inserción de los nuevos individuos, los cuales tienden a eliminar a los peores sujetos de la generación anterior.

La reproducción entre individuos de una población tiene su fundamento en las leyes de herencia que estableció Johan Gregor Mendel entre los años 1865 y 1866. La retransmisión de características a los descendientes se produce a través de la información genética de los cromosomas¹. Durante el proceso de reproducción también juegan un papel muy importante las mutaciones. Cuando se transmite un determinado gen, este puede cambiar de forma aleatoria.

Así, en un proceso de computación evolutiva, los elementos de la población representan soluciones a los problemas. Estas soluciones van generando otras nuevas, imitando la reproducción entre especies. Una vez se llega a una condición de parada definida con anterioridad por el usuario, el algoritmo para. Por ejemplo, esto puede ser que ya se haya alcanzado un determinado número de generaciones, o que se haya encontrado una solución por debajo de un umbral de error.

En resumen, un proceso evolutivo se compone de las siguientes fases: la evaluación de los individuos de una generación, la selección de cuales se van a reproducir, la recombinación de los genes en la fase de reproducción (que incluye la mutación aleatoria de algunos genes) y, por último, la inserción de los nuevos individuos creados en la población.

2.3.2 Algoritmos genéticos

Los algoritmos genéticos son un modelo de computación evolutiva que surge en la década de 1970 gracias a John Henry Holland [15]. Se basan en encontrar la solución óptima de un problema utilizando un modelo que imita la selección natural.

Hay distintos tipos de algoritmos genéticos dependiendo de la metodología utilizada. Uno de los más utilizados es el basado en la teoría de evolución de Darwin. Consiste en hacer distintas iteraciones que repliquen el siguiente esquema (al cual se hará referencia posteriormente para explicar las diferentes secciones):

1. Se parte de una población inicial. Son una serie de individuos originados aleatoriamente distintos entre sí.
2. Se evalúa cómo de buenos son dichos individuos, y así se determina su probabilidad de reproducción.
3. En función de las probabilidades definidas en el paso anterior, se selecciona a los elementos de la población que pasarán a la fase de reproducción o cruce.
4. Los individuos seleccionados se reproducen, dando lugar a nuevos individuos hijos. En este momento hay genes que pueden cambiar de forma aleatoria. A esto se le denomina mutación. En el algoritmo se habrá

¹ m. Biol. Filamento condensado de ácido desoxirribonucleico, visible en el núcleo de las células durante la mitosis y cuyo número es constante para las células de cada especie animal o vegetal [14].

definido previamente cual es la probabilidad de que un gen sufra este cambio.

5. Se inserta a los hijos en la población, generando una nueva generación. Esta mantiene el número de individuos que la generación anterior. Por tanto, en el proceso de inserción se descarta a determinados individuos que estaban hasta ese momento.
6. Con una nueva generación ya creada, se comprueban las condiciones de parada del algoritmo y se determina si se continua o no iterando.

Otra metodología muy utilizada en este tipo de algoritmos es la que utiliza las teorías de Jean-Baptiste Lamarck [15]. Este proceso es un parecido al anterior. La única diferencia es que después de la creación de una generación, inicial o intermedia, se realiza una adaptación Lamarckiana. Es un cambio cromosómico que se relaciona con la evolución de los individuos en función de su entorno.

Las principales ventajas de los algoritmos genéticos son su rapidez y sencillez. Como las soluciones se representan de forma vectorial y binaria, se puede trabajar con ellas sin necesitar un trabajo muy complejo. Sin embargo, al igual que se ha mencionado con los algoritmos metaheurísticos, no hay una garantía de encontrar la solución óptima

2.3.2.1 Representación de un individuo

Un individuo representa una solución determinada al problema planteado. Cada uno posee una serie de propiedades (genotipo) y, para representarlos, existen distintas codificaciones.

Un individuo se representa como un vector, donde cada posición es un gen. Por ejemplo, en codificación binaria: el elemento en la posición i del vector vale 1 si el gen número i está activo en el individuo y 0 si no lo está. Es decir, cada posición del vector representa una particularidad que puede estar o no presente en la solución del problema. Un individuo es una solución concreta que tiene las características que están marcadas con un 1.

Ejemplo: Individuo 1 = [0, 1, 0, 1, 1, 0, 0, 0, 0].

Otra forma de representar a un individuo es con una codificación entera, o real, donde cada gen toma valores en un cierto rango y conjunto de números.

Ejemplo: Individuo 2 = [1.5, 2.0, 3.0, 1.2].

2.3.2.2 Métodos de selección

En el paso 3 de esquema que sigue el algoritmo genético (expuesto con anterioridad al principio del punto 2.3.2) se realiza una selección de los individuos que van a reproducirse [10]. Esta elección siempre debe garantizar que los mejores individuos tengan una mayor probabilidad de reproducirse. Sin

embargo, no se puede impedir que ningún individuo se reproduzca, por muy mala que sea su evaluación, ya que se estaría eliminando la parte de aleatoriedad del algoritmo.

Para llevar a cabo el proceso de selección hay distintos mecanismos, dentro de los cuales los más comunes son los siguientes:

1. Selección por torneo: se selecciona a una serie de elementos de forma aleatoria (por lo general 2) y el mejor evaluado se incluye en el grupo de reproductores. Este proceso se repite tantas veces como sea necesario hasta completar el número de reproductores necesarios para originar la siguiente generación.
2. Selección por ruleta: igual que en el anterior método, aleatoriamente se escogen a tantos individuos como sean necesarios para completar el grupo de reproductores. La probabilidad de cada individuo de ser seleccionado va ligada directamente a su evaluación. Es decir, un individuo con una evaluación mejor tendrá más opciones de reproducirse que otro considerado peor.

2.3.2.3 Proceso de reproducción

Una vez se llega al grupo reproductor se pasa a la fase de reproducción [10], la cual se realiza en parejas, que serán los padres. Para determinar los genes del nuevo individuo hijo se utiliza un operador de cruce y, posteriormente el de mutación. El operador de cruce se determina de tal forma que el hijo herede características de cada uno de los padres. Dependiendo de la codificación de los individuos los operadores son diferentes. Por ejemplo, para una codificación binaria se pueden heredar los 1 que tengan los padres en común y seleccionar aleatoriamente el resto de los genes. Una codificación entera o real se puede utilizar diferentes cruces que combinen los números. Por ejemplo, un cruce aritmético donde los genes del hijo sean la media aritmética de los de los padres. De esta misma forma, existen numerosos operadores de cruce que se pueden usar para diferentes códigos. En cada problema se debe seleccionar el que se considere más adecuado para la solución que se desea alcanzar.

2.3.2.4 Mecanismos de inserción

Después de realizar la reproducción se generan nuevos individuos. Estos juegan un papel crucial en formar la siguiente generación, para lo que existen dos modelos [10]:

1. Modelo generacional: en cada iteración del algoritmo se crea una población totalmente nueva. Los individuos hijos que han nacido de las diferentes reproducciones sustituyen totalmente la antigua generación. De esta forma, de las reproducciones salen tantos nuevos individuos como elementos haya en cada generación.
2. Modelo estacionario: se caracteriza por solo reemplazar algunos individuos. Hay muchas formas de seleccionar que individuos van a pasar de generación y cuales se van a eliminar. Por ejemplo, si se quiere reemplazar a n individuos, una forma sería eliminar los n peores. Otra opción puede ser eliminar aleatoriamente entre aquellos individuos que se hayan reproducido, o más sencillamente escoger de forma totalmente aleatoria. De todas formas, para que el algoritmo evolucione lo antes

posible hacia la solución óptima es recomendable intentar priorizar que se mantengan los mejores individuos.

2.3.2.5 Criterios de parada

Como bien se ha mencionado en la introducción de los algoritmos genéticos, no hay garantía de alcanzar la solución óptima. Es por esto por lo que hay que establecer una serie de condiciones que paren las iteraciones del algoritmo. Las más comunes son [10]:

1. Criterios en función del consumo de recursos: haber llegado a un determinado número de generaciones, individuos o tiempo.
2. Llegar a una solución factible que sea suficientemente buena. Por ejemplo, si el objetivo del problema es minimizar el error de la solución, sería que el resultado obtenido se encuentre por debajo de un determinado error.
3. Estancamiento: cuando después de un número concreto de iteraciones la mejor solución encontrada no mejore.

3 Estado del arte

Hasta este punto se ha hablado de 3 conceptos fundamentales que se van a utilizar en esta investigación: gestión de carteras, optimización multiobjetivo y algoritmos genéticos. El desarrollo de la ciencia ha permitido que muchas áreas se estudien, en ocasiones, de forma conjunta. Hasta la actualidad, se han realizado numerosas investigaciones interdisciplinarias que ligan estos conceptos de distintas maneras. A continuación, se va a proceder a explicar brevemente los trabajos más relevantes encontrados sobre estos temas que se han utilizado como punto de partida para este proyecto. Se ha intentado que dichos trabajos sean lo más recientes posibles. No tendría mucho sentido utilizar como base un documento de, por ejemplo, hace 20 años, ya que probablemente esté un poco obsoleto y haya muchos avances desde entonces.

El primer ejemplo usado es un trabajo de diciembre 2015, “Algoritmo evolutivo para la optimización de portafolios de inversión” [2]. Está escrito por Alan Gustavo Plata, Héctor Rodríguez y Jorge Emyr Vázquez en el Instituto Politécnico Nacional de México. Trata de implementar un sistema en el que se empleen técnicas de optimización evolutiva multiobjetivo y programación en paralelo para optimizar carteras siguiendo el modelo de Harry Markowitz (1952). Para ello se han utilizado datos de la bolsa mexicana de valores. Llega a la conclusión de que la situación económica del país dificulta conseguir rentabilidades muy elevadas y analiza cual es mejor algoritmo específico para tratar ese problema.

La diferencia de este trabajo con el mío es que en este proyecto no se van a analizar unos datos en específico, si no que en función de los datos que se tengan, se pretende llegar a un algoritmo que indique cual es mejor conjunto de soluciones dependiendo de factores que decida el usuario, como es la rentabilidad.

En segundo lugar, se ha tenido en cuenta el trabajo realizado en 2016 por Santiago Arévalo González en “Modelo de gestión de carteras de Markowitz usando algoritmos genéticos” [16]. Al igual que en el ejemplo anterior, aunque en mi trabajo no se vaya a hacer uso de este tipo de carteras en concreto, se ha podido ver una aplicación del tipo de algoritmo seleccionado en el área financiera, lo que ha ayudado a plantear el problema. Se llega a 2 conclusiones: primero, que los algoritmos genéticos se adaptan muy bien a este tipo de problemas. Una vez estudiado a fondo el caso se puede modificar el algoritmo con suficiente precisión para que converja en una solución lo suficientemente cercana a la óptima, si no es la solución óptima. Segundo, el número de variables que influyen en la rentabilidad de una cartera es muy abundante. Muchas de ellas cambian con facilidad en el tiempo, por lo que, siempre que proceda, es conveniente realizar estudios donde los parámetros se estudien a lo largo de un periodo.

Tercero está el trabajo de 2007 de César Peñuela y Mauricio Granada, 2 alumnos de máster de ingeniería eléctrica de la universidad tecnológica de

Pereira: “Optimización multiobjetivo usando un algoritmo genético y un operador estilista basado en un ordenamiento no dominado” [17]. Trata del estudio de la herramienta NSGA-II en este tipo de problemas. Se concluye que NSGA-II es un algoritmo que mantiene las soluciones dominantes dentro de todas las generaciones por las que pasan las iteraciones del algoritmo. Así, podría resultar útil cuando no se quiere almacenar las soluciones de forma constante en otro archivo.

Por último, se ha analizado el trabajo “algoritmos evolutivos multiobjetivo para selección de carteras de inversión con restricciones de cardinalidad”, escrito por Feijoo Colomine Duran (Universidad Experimental de Táchira), Carlos Cotta y Antonio J. Fernández (Universidad de Málaga) [18]. Se estudia el mismo problema de gestión de carteras, utilizando como marco experimental los valores reales de la Bolsa de Caracas. Además, se usan y comparan 3 conceptos diferentes de los algoritmos genéticos: NSGA II, SPEA2 e IBEA. La conclusión principal que se ha extraído de este estudio es que dependiendo del método y la cardinalidad utilizada para plantear el problema, el índice de Sharpe ofrece resultados distintos. Es decir, dicho índice califica un algoritmo, IBEA particularmente, como mejor que los otros. Esto ha servido para plantear la siguiente pregunta: dependiendo de las características de un determinado perfil de inversor, será mejor una solución u otra. Pero al considerar todas esas soluciones de forma conjunta nos encontramos con un frente de Pareto donde todas las opciones son igual de buenas. Sin embargo, como ya se ha explicado con anterioridad, en el ámbito financiero se utilizan herramientas como el índice Sharpe para calificar las soluciones encontradas. Entonces, ¿está este índice discriminando las soluciones hacia un perfil de inversor en concreto?

4 Planteamiento y formulación del problema

El objetivo de este trabajo es conseguir optimizar el problema de gestión de carteras centrándose en maximizar la rentabilidad y minimizar el riesgo de los activos. Para cada inversor, en base a su perfil y preferencias, se pretende encontrar la mejor cartera posible.

Para ello, se va a desarrollar un algoritmo genético que busque la solución al problema de optimización multiobjetivo. El algoritmo parte de un archivo con los datos del problema y pretende encontrar un frente de Pareto compuesto por las mejores soluciones posibles. La herramienta que se va a utilizar en la implementación del algoritmo es Python, donde se hará uso de las siguientes librerías: pandas, numpy, random, math y sys.

De forma más concreta, los objetivos específicos a desarrollar son los siguientes:

1. Plantear y formular el problema genético.
2. Implementar la función del problema de optimización multiobjetivo que se a usar para evaluar las soluciones, así como sus restricciones (si las hay).
3. Desarrollar en Python el algoritmo genético que resuelva el problema.
4. Hallar el frente de Pareto con todas las soluciones no dominadas.
5. Analizar que solución es mejor para cada inversor.
6. Evaluar si el índice de Sharpe discrimina las soluciones hacia un perfil de inversor determinado.

Este problema, en cada momento concreto, se puede resolver con los datos que más convenga. De la misma forma, según interese, se pueden añadir más funciones en la optimización multiobjetivo aparte de la rentabilidad y el riesgo. En este trabajo no importan demasiado los datos escogidos, ya que la meta es ver como se desarrollaría y analizaría un problema de este estilo. La elección de los datos se ha basado en escoger títulos con rentabilidades y riesgos muy diferentes entre sí. Así, se pueden encontrar soluciones muy diferentes y evaluar mejor cómo funciona el algoritmo.

Los títulos seleccionados son:

- Aena S.M.E., S.A.
- Atresmedia Corporación de Medios de Comunicación, S.A.
- ADL Bionatur Solutions, S.A.
- Naturhouse Health, S.A.
- Clínica Baviera, S.A.
- 1 Kessler Global, FI.
- Blackrock Global Funds - Global Allocation Fund E2.
- Kutxabank Rf Carteras Fi.
- Bono del Tesoro de España a 5 años.

4.1 Diseño del sistema: entradas y salidas

Los datos de entrada del sistema, necesarios para comenzar a resolver el problema son los relativos a los activos que formarán, o no, la cartera. Estos se encuentran en un fichero xls, donde se indica el nombre del título, así como su

rentabilidad y el riesgo. Los valores utilizados para calcular estos datos se describirán con detalle más adelante en la sección 4.1.1.

Por otro lado, la salida del sistema es un conjunto de soluciones no dominadas, dentro de las cuales el inversionista podrá elegir la que más le convenga. Además, según se ejecuta el algoritmo, se va guardando una serie de información y gráficas de los individuos no dominados. Esto se hace para poder apreciar la evolución que ha tenido el algoritmo hasta llegar a la solución final.

4.1.1 Tratamiento y justificación de los datos iniciales

4.1.1.1 Justificación y obtención de datos elegidos

Como bien se ha mencionado anteriormente, se han elegido una serie de datos para poder evaluar y probar el algoritmo. Esta elección se basa en intentar escoger títulos muy diferentes entre sí en cuanto a la rentabilidad y el riesgo, ya que son las características que se van a evaluar. Así, se ha decidido seleccionar 5 activos de empresas específicas, 3 fondos de inversión, 1 bono soberano, ya que con títulos con riesgos muy diferentes entre sí.

Dentro de los activos, también se ha priorizado la diversidad. Para ello, se han seleccionado de la siguiente manera.

1. Se ha consultado la base de datos Damodaran [19]. Ahí se han observado las betas por industria en Europa. El coeficiente beta (β) mide el grado de variabilidad de la rentabilidad de un título (en este caso una industria) respecto a la rentabilidad del mercado en el que cotiza. Cuando la beta es igual a 1, la industria se mueve en la misma proporción que el índice que se toma como referencia para evaluar al mercado. Cuando la beta es mayor que 1 la industria tiene mayor variabilidad que el índice del mercado. Por ejemplo, si el mercado ha subido un 8% en el último año, la industria ha subido más. De forma contraria, si la beta es menor que uno, la industria tiene menos variabilidad que el índice. Por ejemplo, si el mercado ha subido un 8% en el último año, la industria ha subido menos [20]. Así, se han seleccionado 5 industrias con betas distintas: farmacéutica ($\beta = 1.09$), transporte aéreo ($\beta = 1.00$), meteorológica ($\beta = 0.68$), comercio minorista ($\beta = 1.24$) y salud ($\beta = 0.37$).
2. Se ha visto que empresas se han utilizado para calcular las betas de cada industria y se han seleccionado 5 compañías españolas. Las empresas elegidas han sido las siguientes:
 - a. Aena S.M.E., S.A
 - b. Atresmedia Corporación de Medios de Comunicación, S.A.
 - c. ADL Bionatur Solutions, S.A.
 - d. Naturhouse Health, S.A.
 - e. Clínica Baviera, S.A.
3. Se han consultado los datos históricos de las acciones de cada empresa en el último año (01/04/2021 – 01/04/2022) [21]. La primera mitad de dichos datos, es decir, los que van del 01/04/2021 al 01/10/2021 se utilizarán como datos de entrada al algoritmo. Es decir, el algoritmo genético implementado basará su solución únicamente en la información obtenida de las empresas en este periodo de tiempo. Por otro lado, la segunda mitad de los datos, es decir, los que van del 01/10/2021 al 01/04/2022, se usarán más adelante para comprobar que hubiera

ocurrido si se hubiera hecho caso a la solución encontrada por el algoritmo genético.

Sobre los 3 fondos de inversión que se han seleccionado, se han escogido 3 fondos españoles, uno de renta fija, uno de renta mixta y otro de renta variable. Este tipo de rentas se diferencian en la rentabilidad que se consigue y el riesgo que se asume en la inversión. La renta fija ofrece una rentabilidad limitada con un riesgo muy bajo, mientras que la renta variable ofrece mayores beneficios, pero asumiendo un riesgo más elevado. La renta mixta es un punto intermedio entre ambas [22]. Así, los fondos por los que se ha optado son los siguientes:

1. 1 Kessler Global, FI, cuya gestora es GVC Gaesco Gestión, SGIIC.
2. Blackrock Global Funds - Global Allocation Fund E2, cuya gestora es BlackRock.
3. Kutxabank Rf Carteras Fi, cuya gestora es Kutxabank Gestión, SGIIC.

Por último, como bono soberano también se ha elegido uno español, concretamente el Bono del Tesoro de España a 5 años.

Al igual que se hizo con las diferentes acciones, también se han consultado los datos históricos de los 3 fondos y el bono sobreaño en las mismas fechas.

4.1.1.2 Tratamiento de los datos

Los datos consultados contienen la información diaria histórica de las fechas definidas. A partir de ellos se ha creado un archivo xls que sirve como datos de entrada del problema. Para ello también se ha utilizado Python, concretamente las librerías pandas, numpy y xlswriter.

Para la rentabilidad, se ha visto la variación porcentual diaria. Haciendo la media se ha obtenido la rentabilidad media diaria (dada en porcentaje). Por otro lado, el riesgo de cada activo, como bien se ha explicado en el marco teórico, se corresponde con la desviación típica de dichas variaciones.

4.2 Planteamiento del problema genético

Para comenzar a desarrollar un algoritmo genético, es necesario primero definir una serie de conceptos: la representación de los individuos, la función de evaluación, el método de selección de individuos, el proceso de reproducción y el mecanismo de inserción que se va a seguir, y los criterios de parada del algoritmo.

4.2.1 Representación de los individuos

Un individuo del algoritmo genético simboliza a una posible cartera. Se representa con un vector de 3 dimensiones, donde cada componente del vector es una parte de información de la cartera. Dichos datos son los siguiente:

La primera componente del vector del individuo es un cromosoma, es decir, es la sección que contiene la información genética. Esta se corresponde con la composición de la cartera. Como bien se ha mencionado en la formulación del problema multiobjetivo, cada cartera es un vector, donde sus componentes representan el porcentaje económico que se invierte en ese activo sobre el capital total. Así, el cromosoma tendrá 9 genes, donde cada uno de ellos simboliza a un activo diferente. En el algoritmo se muestra como un vector con de longitud 9, donde cada componente deberá ser un valor numérico en el rango de 0 a 100. Además, la suma de todos los componentes deberá ser igual a 100.

Un ejemplo esta representación es el siguiente:

| | | | | | | | | |
|------|-------|------|------|-------|------|------|------|------|
| 0.26 | 19.73 | 0.00 | 0.00 | 77.66 | 0.01 | 0.16 | 2.18 | 0.00 |
|------|-------|------|------|-------|------|------|------|------|

La segunda parte del vector del individuo corresponde a su evaluación. Esto queda representado por otro vector de longitud 2, donde la primera componente es la evaluación de la rentabilidad y la segunda la del riesgo. Así, ambos valores son siempre numéricos. Por ejemplo, la evaluación de la cartera anterior es la siguiente:

| | |
|------|------|
| 0.31 | 0.09 |
|------|------|

Por último, la tercera componente el vector del individuo contiene la información de su nivel de dominancia. Es un valor entero que indica cuantos individuos de la población lo dominan. Así, cuanto más bajo sea este nivel, mejor es el individuo. Todos los individuos con nivel 0 forman en cada momento el frente de Pareto buscado, ya que representan a los no dominados.

La representación del individuo completo que se ha puesto como ejemplo es la siguiente:

| | | | | | | | | | | | |
|------|-------|------|------|-------|------|------|------|------|------|------|---|
| 0.26 | 19.73 | 0.00 | 0.00 | 77.66 | 0.01 | 0.16 | 2.18 | 0.00 | 0.31 | 0.09 | 0 |
|------|-------|------|------|-------|------|------|------|------|------|------|---|

4.2.2 Formulación del problema multiobjetivo

Cada individuo está formado por la representación de una cartera y su respectiva evaluación y nivel de dominancia. Para evaluar una cartera es necesario definir la función de evaluación que se va a utilizar. Como se van a medir 2 características, la rentabilidad y el riesgo, se requieren 2 funciones. Así, es preciso formalizar el problema de optimización multiobjetivo que se va a seguir.

Se desea maximizar la rentabilidad y minimizar el riesgo de la cartera.

Sea $f(x)$ el riesgo de una cartera y $g(x)$ su rentabilidad. Sea $x = (x_1, x_2, x_3, \dots, x_9)$ la representación de una cartera, donde cada x_k simboliza el porcentaje del capital que se ha invertido en el activo k .

Así, el problema de optimización escrito matemáticamente queda de la siguiente manera:

$$\begin{aligned} & \min f(\bar{x}), \max g(\bar{x}) \\ \text{s.a.} \quad & \sum_{i=0}^9 x_i = 100 \\ & \forall x_k, x_k \in [0,100] \end{aligned}$$

Ecuación 5: expresión matemática del problema de optimización formulado.

4.2.3 Método de selección del algoritmo

En cada generación del algoritmo, se comienza creando una población intermedia. Está formada por una serie de individuos seleccionados de la generación anterior. Estos son los que posteriormente se reproducen, dando lugar a nuevos individuos.

El método de selección utilizado es el de la ruleta, es decir, se hace una selección aleatoria ponderada. Las ponderaciones las delimitan las probabilidades de cruce de los individuos de la población, las cuales son más altas cuanto mejor sea considerado ese individuo. Para esto se prioriza el nivel de dominancia. Cuanto más bajo, mayor probabilidad de cruce. Ante el mismo nivel de dominancia, se favorece al individuo que sea más diferente a aquellos que ya hay seleccionados para la población intermedia. La diferencia entre dos individuos se define como la distancia que hay entre ellos. Así, la diferencia entre un individuo arbitrario 'x' y un conjunto de individuos, es la suma de todas las distancias de 'x' al resto de individuos.

4.2.4 Proceso de reproducción del algoritmo

Una vez obtenida la población intermedia, se seleccionan aleatoriamente parejas de individuos para que se reproduzcan. Dichas reproducciones solo se realizan con la información genética, es decir, la parte de la cartera de cada individuo. La evaluación y el nivel del hijo no se transmiten de forma directa, ya que se calculan a partir del nuevo individuo. Es por esto que, aunque esta información no se herede de la misma forma que los genes, sigue teniendo dependencia de los padres.

La formación de la cartera del individuo hijo sigue el siguiente esquema:

1. Se hereda aleatoriamente el contenido de una posición de cada uno de los padres.
2. El contenido del resto de posiciones es la media aritmética del de los padres.
3. Se analiza cual es la suma de todas las componentes obtenidas y se realiza una proporción con cada componente para que dicha suma sea 100. De esta forma, todas las carteras cumplen la restricción establecida.

A continuación se muestra un ejemplo de dos carteras padres y un posible hijo:

| | | | | | | | | | |
|----------|-------|------|------|------|------|-------|------|-------|-------|
| Padre 1: | 69.14 | 0.00 | 3.99 | 0.25 | 0.00 | 25.62 | 0.63 | 0.01 | 0.35 |
| Padre 2: | 1.76 | 0.00 | 0.00 | 0.00 | 0.00 | 4.31 | 0.01 | 35.94 | 57.97 |

Hijo sin realizar la proporción:

| | | | | | | | | | |
|--|-------|------|------|------|------|-------|------|-------|-------|
| | 35.45 | 0.00 | 2.00 | 0.12 | 0.00 | 14.96 | 0.32 | 17.98 | 29.16 |
|--|-------|------|------|------|------|-------|------|-------|-------|

| | | | | | | | | | |
|-------|-------|------|------|------|------|-------|------|-------|-------|
| Hijo: | 39.00 | 0.00 | 1.00 | 0.00 | 0.00 | 14.00 | 0.00 | 17.00 | 29.00 |
|-------|-------|------|------|------|------|-------|------|-------|-------|

4.2.5 Mecanismo de inserción del algoritmo y mutaciones

Para formar cada una de las nuevas generaciones se sigue el modelo estacionario, es decir, se mantiene la generación anterior y solo se reemplazan algunos individuos. Particularmente, se reemplaza al 50%. El otro 50% que no se sustituye es la élite de la antigua generación. Los individuos que forman dicha élite dependen de cada generación. Por ejemplo, si la cantidad de individuos no dominados de una generación es muy grande, no tienen por qué ser todos seleccionados para formar la élite. Por el contrario, si el número de los individuos no dominados es muy bajo, habrá individuos en la élite que no formen parte del conjunto de Pareto.

De los nuevos individuos, el 80% proviene de reproducciones y el 20% restante de mutaciones. Primero, se selecciona al 80% de individuos de la población intermedia de dos en dos. Cada pareja se reproduce generando dos nuevos individuos, los cuales se introducen en la población. Aquellos individuos que no se han reproducido pasan por un proceso de mutación, donde pueden mutar o mantenerse como están. Concretamente, tienen una probabilidad del 30% de mutar. Después de este proceso, se vuelven a evaluar las carteras para actualizar el individuo y se introducen en la nueva generación.

La mutación de un individuo consiste en intercambiar de forma aleatoria el contenido de dos posiciones del vector que representa la cartera. A continuación se muestra una cartera sin mutar y mutada.

| | | | | | | | | |
|------|------|------|-------|-------|------|------|------|------|
| 0.00 | 0.00 | 3.00 | 15.00 | 62.00 | 4.06 | 6.00 | 7.79 | 2.15 |
|------|------|------|-------|-------|------|------|------|------|

| | | | | | | | | |
|------|------|-------|-------|------|------|------|------|------|
| 0.00 | 0.00 | 62.00 | 15.00 | 3.00 | 4.06 | 6.00 | 7.79 | 2.15 |
|------|------|-------|-------|------|------|------|------|------|

4.2.6 Criterios de parada del algoritmo

Por último, los criterios de parada del algoritmo son los siguientes:

- Se hayan alcanzado diez generaciones sin cambio en el conjunto de soluciones.
- Se llegue a un número de generaciones máximas, concretamente 150.

5 Resultados y conclusiones

5.1 Ejemplo de salidas del sistema

Como ya se mencionó brevemente en el capítulo anterior, los datos de salida del algoritmo son: la solución encontrada y un conjunto de ficheros y gráficas que muestran la evolución de dicha solución. Más específicamente, son los siguientes:

1. El conjunto de Pareto encontrado, es decir, la solución al problema de optimización.
2. Un fichero de texto con la información de la evolución del conjunto de Pareto por generación, es decir, un archivo donde se guarda la solución encontrada en cada iteración del algoritmo. Se ha utilizado cuando se necesitaba analizar que ocurría de generación en generación de forma específica. Fundamentalmente, este análisis se ha usado para comprobar que no hubiera errores en el código del algoritmo que produjeran estancamientos de las soluciones.
3. Un fichero de texto igual que el anterior, pero que guarda la información cada 10 generaciones. En caso de que el número de generaciones no sea múltiplo de 10, también añade la solución final. Pese a que los datos de este documento ya estén en el anterior, resulta muy útil cuando se quiere ver la evolución de la solución del algoritmo de forma más general.
4. Un conjunto de gráficas que muestran visualmente la información del fichero de texto explicado en el punto anterior. La gráfica originada en una generación específica representa todas las soluciones encontradas desde el inicio del algoritmo hasta ese momento. Los individuos no dominados de todas las generaciones anteriores a la actual se muestran con un punto de menor tamaño y diferente color y forma que el conjunto de Pareto de esa generación. A continuación se expone un ejemplo:

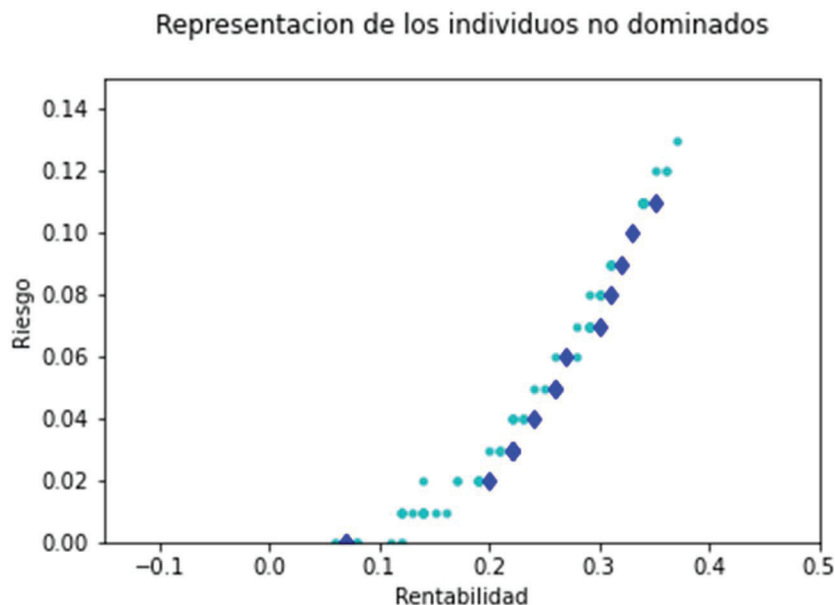


Ilustración 2: ejemplo de gráfica generada por el algoritmo que representa la evolución del conjunto de Pareto hasta una determinada generación.

Como se puede observar en la gráfica, la menor rentabilidad encontrada está entre el 0.00% y el 0.10%, mientras que la mayor no alcanza el 0.40%. Estos valores, al igual que en el fichero de datos de entrada del algoritmo, son las variaciones diarias porcentuales del precio del título. Por otro lado, el riesgo mínimo encontrado es 0.00%, mientras que el máximo se sitúa entre 0.12% y 0.14%.

Además, hay que destacar que ambas variables se redondean a 2 decimales, es decir, en todas las gráficas los datos siempre se van a mover en intervalos de 0.01 respecto a ambos ejes. Esto es más obvio en eje y que en el x, ya que el rango de valores que toman los puntos dentro del eje y es mucho menor que en el x.

Particularmente, el conjunto de Pareto representado en la gráfica anterior es el siguiente (información extraída del fichero que contiene los datos de las soluciones cada 10 generaciones):

Carteras seleccionadas generación 90

| | | |
|--|----------------------|-----------|
| <i>[array ([0.00, 0.00, 0.00, 9.00, 89.00, 0.00, 1.00, 0.00, 1.00]),</i> | <i>[0.35, 0.11],</i> | <i>0]</i> |
| <i>[array ([5.00, 4.00, 1.00, 22.00, 64.00, 100., 3.00, 0.00, 0.00]),</i> | <i>[0.27, 0.06],</i> | <i>0]</i> |
| <i>[array ([0.00, 1.00, 0.00, 9.00, 85.00, 0.00, 1.00, 4.00, 0.00]),</i> | <i>[0.33, 0.10],</i> | <i>0]</i> |
| <i>[array ([3.00, 0.00, 3.00, 33.00, 48.00, 0.00, 3.00, 7.00, 3.00]),</i> | <i>[0.22, 0.03],</i> | <i>0]</i> |
| <i>[array ([3.00, 3.00, 3.00, 33.00, 48.00, 2.00, 3.00, 1.00, 4.00]),</i> | <i>[0.22, 0.03],</i> | <i>0]</i> |
| <i>[array ([0.00, 1.00, 0.00, 42.00, 47.00, 0.00, 0.00, 10.00, 0.00]),</i> | <i>[0.22, 0.03],</i> | <i>0]</i> |
| <i>[array ([79., 0.00, 2.00, 18.00, 0.00, 0.00, 0.00, 1.00, 0.00]),</i> | <i>[0.07, 0.00],</i> | <i>0]</i> |
| <i>[array ([20.00, 2.00, 0.00, 35.00, 39.00, 0.00, 4.00, 0.00, 0.00]),</i> | <i>[0.20, 0.02],</i> | <i>0]</i> |
| <i>[array ([0.00, 1.00, 1.00, 27.00, 71.00, 0.00, 0.00, 0.00, 0.00]),</i> | <i>[0.30, 0.07],</i> | <i>0]</i> |
| <i>[array ([0.00, 0.00, 0.00, 25.00, 75.00, 0.00, 0.00, 0.00, 0.00]),</i> | <i>[0.31, 0.08],</i> | <i>0]</i> |
| <i>[array ([4.00, 0.00, 0.00, 28.00, 55.00, 0.00, 5.00, 6.00, 2.00]),</i> | <i>[0.24, 0.04],</i> | <i>0]</i> |
| <i>[array ([4.00, 0.00, 0.00, 31.00, 60.00, 0.00, 5.00, 0.00, 0.00]),</i> | <i>[0.26, 0.05],</i> | <i>0]</i> |
| <i>[array ([0.00, 2.00, 0.00, 41.00, 57.00, 0.00, 0.00, 0.00, 0.00]),</i> | <i>[0.26, 0.05],</i> | <i>0]</i> |
| <i>[array ([17.00, 1.00, 0.00, 20.00, 48.00, 0.00, 5.00, 8.00, 1.00]),</i> | <i>[0.22, 0.03],</i> | <i>0]</i> |
| <i>[array ([79.00, 0.00, 2.00, 18.00, 0.00, 0.00, 0.00, 1.00, 0.00]),</i> | <i>[0.32, 0.09],</i> | <i>0]</i> |
| <i>[array ([68.00, 1.00, 5.00, 22.00, 0.00, 3.00, 0.00, 0.00, 1.00]),</i> | <i>[0.07, 0.00],</i> | <i>0]</i> |
| <i>[array ([71.00, 1.00, 1.00, 23.00, 0.00, 0.00, 0.00, 0.00, 4.00]),</i> | <i>[0.07, 0.00],</i> | <i>0]</i> |

Teniendo la información completa sobre todos los individuos, la mayor rentabilidad alcanzada es 0.35%, con un riesgo de 0.11%. De forma opuesta, la menor rentabilidad se sitúa en un valor de 0.07%, con un riesgo nulo. El mayor valor que alcanza la rentabilidad coincide en el mismo individuo que también tiene el riesgo más elevado, así como la menor rentabilidad coincide, de la misma forma, con el menor riesgo. Esta característica es necesaria que se

cumpla siempre en el frente de Pareto. Así, podrían aparecer otras soluciones, pero estarían dominadas.

A continuación se va a mostrar el conjunto de gráficas de la ejecución del algoritmo de la cual se ha extraído el ejemplo anterior, para poder tener una visión más general del desarrollo del frente de Pareto. Concretamente, la ilustración 2 se corresponde con la generación 90.

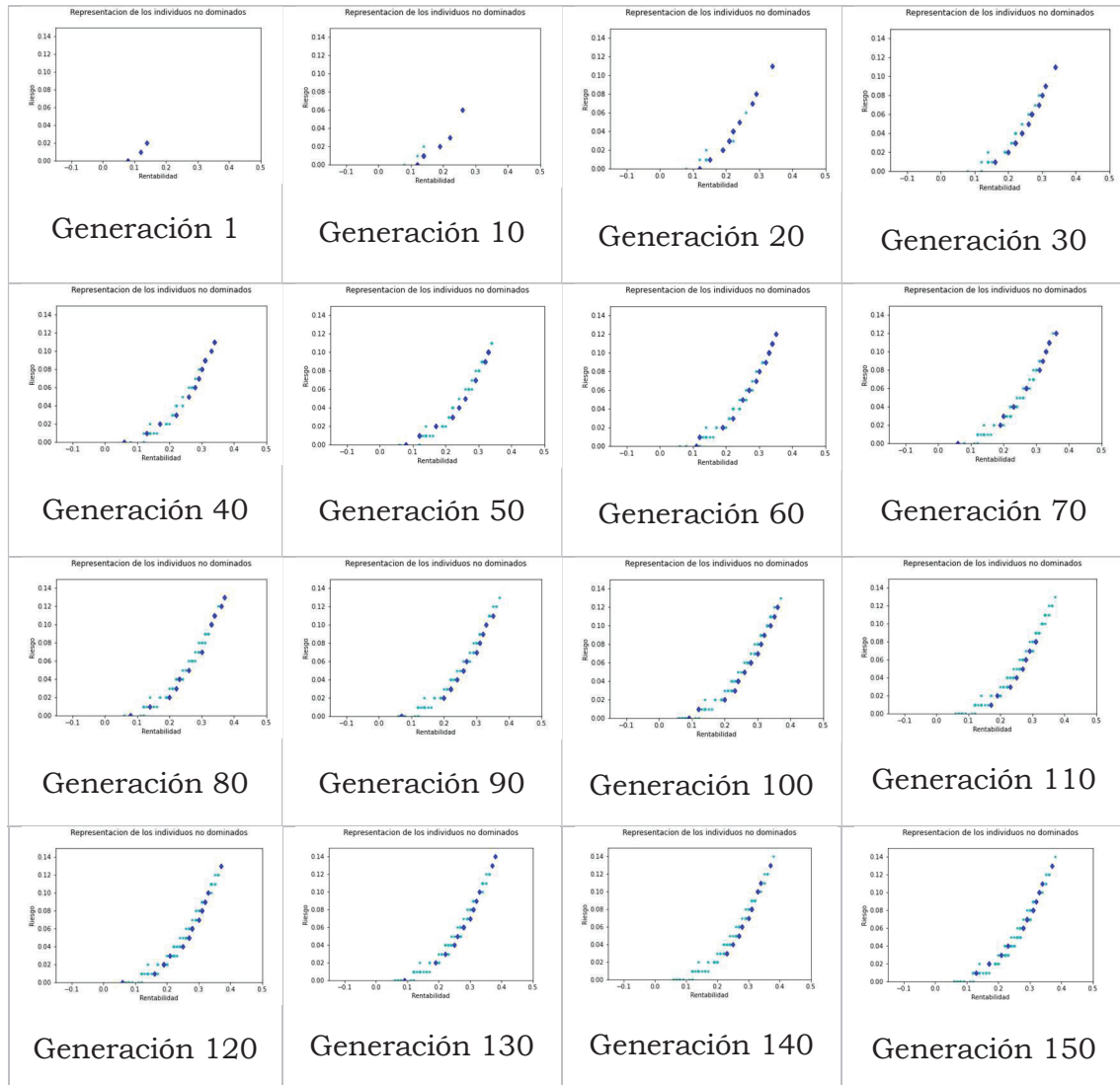


Ilustración 3: ejemplo de evolución del conjunto de Pareto en el algoritmo genético.

Observando detenidamente las gráficas se destaca:

1. Las soluciones se van desplazando lentamente hacia la parte derecha inferior de la gráfica. Es precisamente en esta área donde se encontrarán los mejores resultados, ya que se está maximizando la rentabilidad (eje x) y minimizando el riesgo (eje y).
2. Acerca de la evolución del algoritmo, a simple vista puede parecer que el conjunto de Pareto apenas cambia entre generaciones, es decir, que la solución evoluciona poco. En la generación 90, por ejemplo, la cartera obtenida con rentabilidad más alta tenía una evaluación de 0.35% de

rentabilidad y 0.11% de riesgo. En la última generación se llega a un 0.37% y 0.13%, respectivamente. Aumentar un 0.02% la rentabilidad puede parecer poco significativo, sin embargo, se trata de cifras diarias. Si se extrapola a datos anuales, se obtendría un 7.3% de rentabilidad. Poniendo un ejemplo, si la inversión inicial realizada por el inversor es de 500000€, y el negocio va bien, se correspondería con tener una ganancia de 36500€ al año.

3. En la composición de una nueva generación siempre está presente el 50% de los mejores individuos de la generación anterior, los cuales reciben el nombre de élite. Sin embargo, hay ocasiones donde, de una gráfica a la siguiente, se pierden individuos no dominados. Esto se explica porque el número de individuos que forman parte de cada generación es uno de los parámetros que se escoge al principio del código del algoritmo. Este se puede variar según quiera la persona que controla el código. Todas las pruebas se han hecho con esta variable fijada a 30 individuos. Si la longitud del conjunto de Pareto de una determinada generación es mayor que 15, entonces, por mucho que todos los individuos sean igual de buenos y, por tanto, todos puedan estar en la élite de dicha generación, no pasarán todos a formar parte de la siguiente. Es decir, se perderá alguna solución.

5.2 Clasificación de soluciones para distintos tipos de inversor

Uno de los objetivos planteados inicialmente era ser capaz de clasificar las soluciones obtenidas por distintos tipos de perfiles de inversor.

Haciendo un breve repaso de los principales conceptos explicados sobre esa área en el marco teórico, se puede clasificar a los inversores utilizando factores como la edad, situación personal y laboral, tolerancia al riesgo, horizonte de inversión, etc. En función de estas características, se definen distintos tipos de carteras: conservadora, moderada, de crecimiento y de crecimiento agresivo. En cada una, se prioriza un tipo de activo, según diversos factores, donde destacan el riesgo y la rentabilidad. Hay inversores que están dispuestos a asumir más riesgos para conseguir mayores beneficios. Por el contrario, otros inversores prefieren no potenciar tanto el crecimiento financiero, asumiendo un riesgo menor.

La solución obtenida en el ejemplo dado es el siguiente conjunto de Pareto (se numeran los individuos y, por consiguiente, las carteras para que en el resto de la sección sea más fácil referirse a ellas):

| | | |
|--|------------------|----|
| 1. $[array ([0.00, 23.00, 14.00, 23.00, 37.00, 3.00, 0.00, 0.00, 0.00])$, | $[0.17, 0.02]$, | 0] |
| 2. $[array ([0.00, 32.00, 0.00, 0.00, 68.00, 0.00, 0.00, 0.00, 0.00])$, | $[0.29, 0.07]$, | 0] |
| 3. $[array ([0.00, 3.00, 0.00, 22.00, 74.00, 0.00, 1.00, 0.00, 0.00])$, | $[0.31, 0.08]$, | 0] |
| 4. $[array ([0.00, 0.00, 3.00, 79.00, 14.00, 0.00, 0.00, 4.00, 0.00])$, | $[0.13, 0.01]$, | 0] |
| 5. $[array ([0.00, 1.00, 3.00, 23.00, 71.00, 0.00, 2.00, 0.00, 0.00])$, | $[0.29, 0.07]$, | 0] |
| 6. $[array ([1.00, 0.00, 0.00, 29.00, 68.00, 0.00, 0.00, 2.00, 0.00])$, | $[0.29, 0.07]$, | 0] |

| | | |
|--|---------------|----|
| 7. [array ([0.00, 0.00, 2.00, 23.00, 75.00, 0.00, 0.00, 0.00, 0.00]), | [0.31, 0.08], | 0] |
| 8. [array ([0.00, 0.00, 1.00, 33.00, 65.00, 0.00, 0.00, 0.00, 1.00]), | [0.28, 0.06], | 0] |
| 9. [array ([0.00, 0.00, 28.00, 0.00, 70.00, 0.00, 2.00, 0.00, 0.00]), | [0.29, 0.07], | 0] |
| 10. [array ([3.00, 2.00, 0.00, 14.00, 80.00, 0.00, 1.00, 0.00, 0.00]), | [0.32, 0.09], | 0] |
| 11. [array ([32.00, 0.00, 0.00, 0.00, 66.00, 0.00, 0.00, 2.00, 0.00]), | [0.28, 0.06], | 0] |
| 12. [array ([0.00, 0.00, 0.00, 23.00, 75.00, 0.00, 1.00, 1.00, 0.00]), | [0.31, 0.08], | 0] |
| 13. [array ([0.00, 0.00, 0.00, 30.00, 67.00, 0.00, 3.00, 0.00, 0.00]), | [0.29, 0.07], | 0] |
| 14. [array ([0.00, 1.00, 4.00, 33.00, 52.00, 0.00, 0.00, 2.00, 8.00]), | [0.23, 0.04], | 0] |
| 15. [array ([1.00, 0.00, 0.00, 13.00, 86.00, 0.00, 0.00, 0.00, 0.00]), | [0.34, 0.11], | 0] |
| 16. [array ([0.00, 2.00, 1.00, 27.00, 69.00, 0.00, 0.00, 0.00, 1.00]), | [0.29, 0.07], | 0] |
| 17. [array ([0.00, 2.00, 2.00, 26.00, 70.00, 0.00, 0.00, 0.00, 0.00]), | [0.29, 0.07], | 0] |
| 18. [array ([1.00, 0.00, 0.00, 0.00, 96.00, 0.00, 0.00, 3.00, 0.00]), | [0.37, 0.13], | 0] |
| 19. [array ([0.00, 2.00, 0.00, 14.00, 83.00, 0.00, 1.00, 0.00, 0.00]), | [0.33, 0.10], | 0] |
| 20. [array ([0.00, 22.00, 13.00, 28.00, 35.00, 2.00, 0.00, 0.00, 0.00]), | [0.17, 0.02], | 0] |
| 21. [array ([1.00, 0.00, 17.00, 36.00, 45.00, 1.00, 0.00, 0.00, 0.00]), | [0.21, 0.03], | 0] |
| 22. [array ([0.00, 0.00, 0.00, 84.00, 15.00, 0.00, 0.00, 0.00, 1.00]), | [0.33, 0.10], | 0] |

Y su representación gráfica es la siguiente:

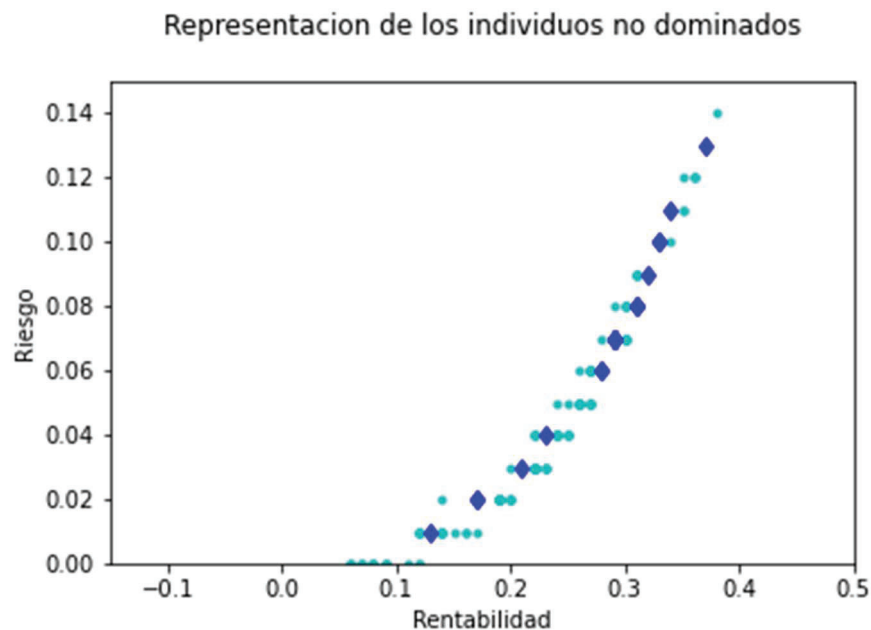


Ilustración 4: gráfica del frente de Pareto hallado como solución en el ejemplo dado.

A continuación, se divide la gráfica en distintas secciones para analizar qué tipos de carteras se tienen en cada uno de los casos.

Representación de los individuos no dominados

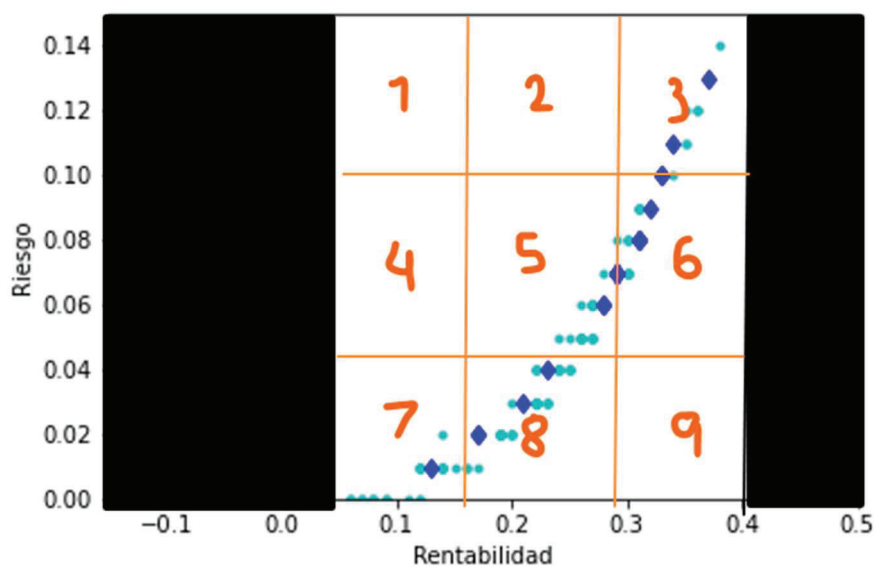


Ilustración 5: clasificación inicial de carteras por secciones.

Respecto a la rentabilidad, la primera columna (casillas 1, 4 y 7) se corresponde con las carteras cuyo crecimiento es menor, nunca llegando a ser nulo. Según una cartera se desplaza una columna a la derecha, dicho crecimiento va aumentando, correspondiéndose las casillas 2, 5 y 8 con una rentabilidad moderada, y las 3, 6 y 9 con una rentabilidad alta.

Por otro lado, en relación con el riesgo, la fila superior (casillas 1, 2 y 3) es donde se encuentran las carteras de riesgo alto, la fila de en medio (casillas 4, 5 y 6) se corresponde con un riesgo moderado y, por último, las casillas restantes, situadas en la fila inferior, representan un riesgo bajo.

Entonces, sabiendo los distintos tipos de carteras según el perfil de inversor ¿En qué casilla se ubica cada uno de estos tipos?

Primero, hay 3 casillas que se descartan de cualquier elección: 1, 2 y 4, ya que nadie desea invertir con una rentabilidad muy baja y un riesgo muy alto. De hecho, no hay ningún individuo cuya cartera se encuentre en esta área. De la misma forma, parecería lógico eliminar también las casillas 5 y 7. En el primer caso, si un inversor quiere crear una cartera con riesgo moderado, se intentaría maximizar la rentabilidad fijado el tipo de riesgo, lo cual le llevaría a carteras en la casilla 6, no 5. Lo mismo ocurre con las carteras en la casilla 7: si un inversor desea invertir asumiendo un riesgo bajo, se intentarían aumentar los beneficios todo lo posible dentro de este riesgo, lo que le conduciría a carteras de la casilla 8. Sin embargo, la curva que forma el frente de Pareto muestra que, aunque sea por poco, las carteras en las casillas 5 y 7 toman riesgos inferiores que las que están en las casillas 6 y 8, respectivamente. Podría haber inversores para los que un 0.01% de diferencia entre algún valor pudiera significar una diferencia. Por tanto, estas casillas se siguen teniendo en cuenta.

Segundo, la casilla ideal en la que todo inversor querría estar es la 9, donde se maximizan los beneficios con un riesgo mínimo. Sin embargo, no se pueden encontrar carteras con estas características.

Así, sobre la gráfica, la clasificación de carteras por casillas queda de la siguiente manera:

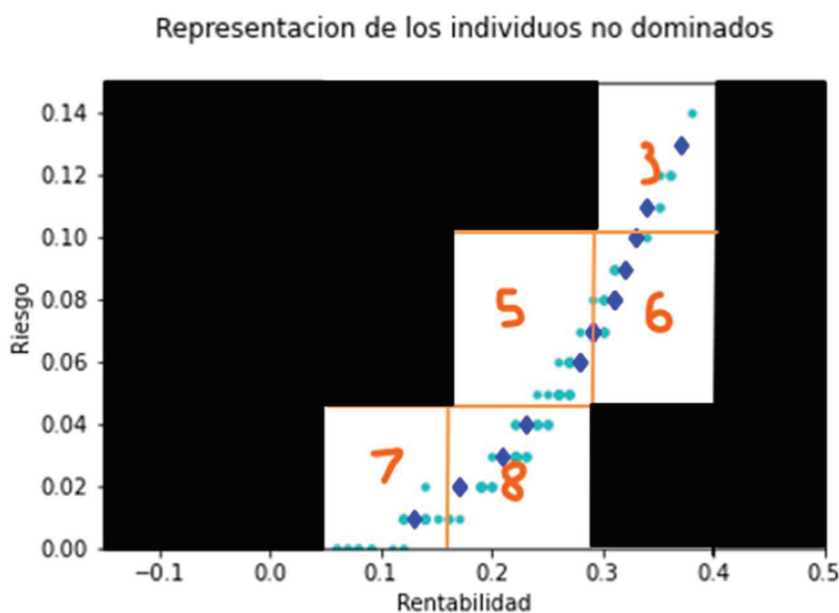


Ilustración 6: clasificación final de carteras por secciones.

La casilla número 7 se corresponde con un tipo de cartera conservadora, donde la tolerancia al riesgo es muy baja. La única cartera obtenida con estas características es la siguiente:

4. [0.00, 0.00, 3.00, 79.00, 14.00, 0.00, 0.00, 4.00, 0.00] → Rentabilidad = 0,13%, riesgo = 0,01%

Si se expresa el vector que simboliza la cartera en activos, la inversión queda distribuida de la siguiente forma:

- La mayor parte del capital (79%) iría a Naturhouse Health, S.A.
- El 14% del capital total se invertiría en las acciones de la Clínica Baviera, S.A.
- Porcentajes más bajos del capital monetario, 4% y 3%, se invertirían en Kutxabank Rf Carteras Fi y ADL Bionatur Solutions, S.A., respectivamente.

Resulta sorprendente que la mayoría de la inversión se realice sobre activos de empresas específicas, ya que, teóricamente, son aquellos cuyo riesgo es más elevado. Esto puede ocurrir por el periodo de tiempo del que se han extraído los datos. Pese a que el riesgo de una determinada empresa pueda ser muy alto, durante ese periodo puede haberse mantenido muy estable y, por tanto, en el primer análisis de datos, donde se calculaban la rentabilidad y riesgo para cada activo, se haya medido como de riesgo bajo. Aun así, es posible que para este tipo de cartera haya otra combinación de activos que logre un mejor resultado, pero que el algoritmo no la haya encontrado. Sin embargo, también hay que tener en cuenta que la solución ofrecida está basada en un modelo genético. Como bien se explicó en el marco teórico, este tipo de implementaciones se caracterizan por no encontrar siempre la mejor solución de forma exacta. Se pretende llegar a un resultado aproximado que, pese a no ser el mejor, sea suficientemente bueno como para poder utilizarlo.

Las casillas número 5 y 8 se relacionan con carteras moderadas. Dependiendo si la tolerancia al riesgo es más alta o baja se cambiaría de una casilla a otra, siendo la 5 aquella con un riesgo más elevado.

De forma análoga a la cartera de la casilla 7, las obtenidas en esta parte de la gráfica son las siguientes:

1. [0.00, 23.00, 14.00, 23.00, 37.00, 3.00, 0.00, 0.00, 0.00] → Rentabilidad = 0,17%, riesgo = 0,02
2. [0.00, 32.00, 0.00, 0.00, 68.00, 0.00, 0.00, 0.00, 0.00] → Rentabilidad = 0,29%, riesgo = 0,07%
5. [0.00, 1.00, 3.00, 23.00, 71.00, 0.00, 2.00, 0.00, 0.00] → Rentabilidad = 0,29%, riesgo = 0,07%
6. [1.00, 0.00, 0.00, 29.00, 68.00, 0.00, 0.00, 2.00, 0.00] → Rentabilidad = 0,29%, riesgo = 0,07%
8. [0.00, 0.00, 1.00, 33.00, 65.00, 0.00, 0.00, 0.00, 1.00] → Rentabilidad = 0,28%, riesgo = 0,06%
9. [0.00, 0.00, 28.00, 0.00, 70.00, 0.00, 2.00, 0.00, 0.00] → Rentabilidad = 0,29%, riesgo = 0,07%
11. [32.00, 0.00, 0.00, 0.00, 66.00, 0.00, 0.00, 2.00, 0.00] → Rentabilidad = 0,28%, riesgo = 0,06%
13. [0.00, 0.00, 0.00, 30.00, 67.00, 0.00, 3.00, 0.00, 0.00] → Rentabilidad = 0,29%, riesgo = 0,07%
14. [0.00, 1.00, 4.00, 33.00, 52.00, 0.00, 0.00, 2.00, 8.00] → Rentabilidad = 0,23%, riesgo = 0,04%
16. [0.00, 2.00, 1.00, 27.00, 69.00, 0.00, 0.00, 0.00, 1.00] → Rentabilidad = 0,29%, riesgo = 0,07%
17. [0.00, 2.00, 2.00, 26.00, 70.00, 0.00, 0.00, 0.00, 0.00] → Rentabilidad = 0,29%, riesgo = 0,07%
20. [0.00, 22.00, 13.00, 28.00, 35.00, 2.00, 0.00, 0.00, 0.00] → Rentabilidad = 0,17%, riesgo = 0,02%
21. [1.00, 0.00, 17.00, 36.00, 45.00, 1.00, 0.00, 0.00, 0.00] → Rentabilidad = 0,21%, riesgo = 0,03%

Se aprecia que, pese a que hay muchas carteras con la misma evaluación, sus composiciones son diferentes. Es decir, hay muchas combinaciones de activos y fondos que llevan al algoritmo al mismo resultado.

Analizando de forma general la composición de estas carteras, destaca que todas coinciden en el activo en el que más se invierte: Clínica Baviera, S.A.

Por otro lado, la casilla número 6 contiene aquellas carteras “de crecimiento”. De forma teórica, con carteras similares a las anteriores en las que se busca priorizar activos de renta variable para conseguir una rentabilidad más elevada. Las carteras obtenidas en esta área en este ejemplo en concreto son las siguientes:

3. [array ([0.00, 3.00, 0.00, 22.00, 74.00, 0.00, 1.00, 0.00, 0.00]), [0.31, 0.08], 0]
7. [array ([0.00, 0.00, 2.00, 23.00, 75.00, 0.00, 0.00, 0.00, 0.00]), [0.31, 0.08], 0]
10. [array ([3.00, 2.00, 0.00, 14.00, 80.00, 0.00, 1.00, 0.00, 0.00]), [0.32, 0.09], 0]
12. [array ([0.00, 0.00, 0.00, 23.00, 75.00, 0.00, 1.00, 1.00, 0.00]), [0.31, 0.08], 0]
19. [array ([0.00, 2.00, 0.00, 14.00, 83.00, 0.00, 1.00, 0.00, 0.00]), [0.33, 0.10], 0]
22. [array ([0.00, 0.00, 0.00, 84.00, 15.00, 0.00, 0.00, 0.00, 1.00]), [0.33, 0.10], 0]

Haciendo un análisis análogo al realizado en el caso anterior, estas carteras se parecen en que todas invierten prácticamente la totalidad de su capital en Clínica Baviera, S.A. y Naturhouse Health, S.A. Se diferencian de las anteriores en que en estas prácticamente no hay inversión en ninguno de los fondos, ni en bonos del tesoro de España. Mientras que en el caso de carteras

moderadas se podía invertir hasta un 8% en alguno de estos títulos, ahora solo se alcanza un 1% en algunas carteras. Esto coincide con la teoría, ya que la mayoría de los fondos son de renta fija, mientras que los activos de renta variable ofrecen más riesgo y rentabilidad.

Por último, la casilla número 3 se corresponde con carteras de crecimiento agresivo, donde siempre se prioriza la rentabilidad. En esta zona del gráfico se ubican 2 carteras:

15. $[array ([1.00, 0.00, 0.00, 13.00, 86.00, 0.00, 0.00, 0.00, 0.00]), [0.34, 0.11], 0]$
 18. $[array ([1.00, 0.00, 0.00, 0.00, 96.00, 0.00, 0.00, 3.00, 0.00]), [0.37, 0.13], 0]$

Las evaluaciones obtenidas confirman que son aquellas carteras cuyo riesgo es más elevado, pero también su rentabilidad. Así, con este ejemplo, se puede concluir que, por lo general, el riesgo y la rentabilidad son variables con un alto grado de correlación. Cuando una incrementa su valor, la otra también lo hace.

En estos casos se invierte casi la totalidad del capital en las acciones de la Clínica Baviera. La composición de cartera cuya rentabilidad es la máxima en este ejemplo muestra cómo se invierte un 96% del capital en este activo.

5.3 Análisis del índice de Sharpe respecto a las soluciones obtenidas

Esta sección calcula el índice de Sharpe de todas las carteras obtenidas como soluciones y analiza si dicho índice prioriza algún tipo de cartera.

Para calcular el índice de Sharpe de un título se necesita medir su rentabilidad respecto de la de un título de referencia sin riesgo. Por sencillez, se ha supuesto que la rentabilidad de referencia de un título sin riesgo es nula.

Así, los índices de Sharpe de todas las soluciones obtenidas son los expuestos a continuación. A partir de ahora se identificará a cada cartera por su número dado en la sección 5.2, ya que una vez clasificadas no es relevante su composición.

| Número de identificación de la cartera | Tipo de cartera según la evaluación | Índice de Sharpe |
|--|-------------------------------------|------------------|
| 4 | Conservadora | 13.00 |
| 1 | Moderada | 8.50 |
| 2 | Moderada | 4.14 |
| 5 | Moderada | 4.14 |
| 6 | Moderada | 4.14 |
| 8 | Moderada | 4.66 |
| 9 | Moderada | 4.14 |

| | | |
|----|-------------------------|------|
| 11 | Moderada | 4.66 |
| 13 | Moderada | 4.14 |
| 14 | Moderada | 5.75 |
| 16 | Moderada | 4.14 |
| 17 | Moderada | 4.14 |
| 20 | Moderada | 8.50 |
| 21 | Moderada | 7.00 |
| 3 | De crecimiento | 3.88 |
| 7 | De crecimiento | 3.88 |
| 10 | De crecimiento | 3.56 |
| 12 | De crecimiento | 3.88 |
| 19 | De crecimiento | 3.30 |
| 22 | De crecimiento | 3.30 |
| 15 | De crecimiento agresivo | 3.09 |
| 18 | De crecimiento agresivo | 2.85 |

Tabla 2: relación de cada cartera con su índice de Sharpe

El índice de Sharpe establece una relación entre la rentabilidad y el riesgo de las carteras para analizar el rendimiento de las inversiones. Afirma que, cuanto mayor es el índice, mejor es dicha inversión en cuanto a rendimiento.

Los valores calculados y mostrados en la tabla indican que la mejor cartera es la conservadora, con una clara diferencia. De esta forma, se concluye que este índice favorece claramente carteras con un riesgo muy bajo. Sin embargo, como ya se ha mencionado en otras ocasiones, es posible que este resultado sea muy sensible a los datos utilizados.

5.4 Evolución de la rentabilidad de las carteras seleccionadas con el algoritmo genético

Esta sección mide cómo habría evolucionado la rentabilidad de las carteras seleccionadas con el algoritmo genético en un periodo posterior al utilizado en la simulación.

La solución descrita se ha obtenido con datos históricos de los títulos en un periodo de 6 meses, concretamente desde principios de abril hasta principios de octubre de 2021.

Para medir la evolución posterior de la rentabilidad de las carteras consideradas mejores se han utilizado los datos de octubre de 2021 a abril de 2022. Como ya se explicó en el apartado de tratamiento de datos, a partir de esta información se ha calculado la rentabilidad y el riesgo de cada título.

A continuación, se muestran los resultados obtenidos para una cartera de cada tipo: conservadora, moderada, de crecimiento y de crecimiento agresivo. Se ha intentado escoger, en cada caso, aquella cartera más representativa de cada uno de los tipos, en otras palabras, la más cercana al promedio de las carteras del mismo tipo.

A continuación me referiré a la rentabilidad esperada como la resultante del periodo de 6 meses inicial del análisis; y, por otro lado, me denomino rentabilidad obtenida, la medida en los 6 meses posteriores. Para ilustrar los resultados, he supuesto una inversión inicial de 10.000€ en cada tipo de cartera.

| Tipo de cartera | Rentabilidad esperada | Rentabilidad obtenida |
|-------------------------|------------------------------|------------------------------|
| Conservadora | 23.73% → +2373 € | -3.65% → -365 € |
| Moderada | 52.93% → +5293 € | 10.5% → +1050 € |
| De crecimiento | 56.58% → +5657.5 € | 12.78% → +1277.5 € |
| De crecimiento agresivo | 67.53% → +6752.5 € | 18.25% → +1825 € |

Tabla 2: rentabilidad esperada y obtenida para cada tipo de cartera.

Como se puede observar, las rentabilidades obtenidas son muy inferiores a las esperadas, llegando incluso, en uno de los casos, a suponer pérdidas. El resultado diferente en distintos periodos de tiempo se debe a la evolución de los distintos mercados financieros que ocurre en estos periodos. Esto es un acontecimiento inevitable y siempre va a ocurrir. Por eso, se concluye que no hay que evaluar las variables en términos de valor absoluto, ya que este depende de los precios del mercado. Consecuentemente, la utilidad del algoritmo genético creado no se puede medir de esta forma.

Para evaluar y controlar la solución por los efectos derivados de la evolución del mercado, se debería comparar la rentabilidad observada de cada tipo de cartera con la rentabilidad de un índice de mercado representativo de este tipo de cartera. De renta variable, por ejemplo, se puede coger el IBEX 35². Así, lo que se evaluaría es el diferencial respecto de esa cartera tipo. Esto indicaría que tipo de cartera ha evolucionado mejor, teniendo en cuenta el desarrollo del mercado. Dicho ejercicio no se ha intentado hacer en este trabajo, porque los datos escogidos para el análisis han sido a modo de ejemplo.

5.5 Conclusión general y posibles mejoras

Con la realización de este trabajo se ha comprobado que se puede diseñar un algoritmo genético con los criterios buscados que resulte útil en el área de carteras financieras. Sin embargo, los resultados expuestos en toda la sección 5 muestran que las predicciones son muy sensibles a los datos que se toman como referencia. Probablemente haber utilizado un ciclo económico de 6 meses

² El IBEX 35 (acrónimo de Iberia Index) es el principal índice bursátil de referencia de la bolsa española elaborado por Bolsas y Mercados Españoles (BME). [23].

sea muy poco para poder apreciar la evolución de cada título dentro de su respectivo mercado. Después de analizar los resultados obtenidos, se concluye en que hubiera sido más conveniente utilizar datos en un periodo igual o superior a 1 año.

Por otro lado, otra posible mejora sería haber utilizado un método diferente para determinar la dominancia de los individuos. En el desarrollo que se ha hecho del algoritmo, la dominancia de cada individuo se define mediante el nivel de dominancia. Dicho parámetro, respecto a un individuo 'x', indica cuantos de los individuos que hay actualmente en la población dominan a 'x'. Esto nos lleva a un conjunto de soluciones como la de la ilustración 4. Sin embargo, en esta gráfica se puede ver como la representación de las mejores carteras sigue una tendencia sobre el plano. Se podría hallar la curva que representa esta tendencia y utilizarla como frente de Pareto. Esto aumentaría el número de soluciones y, por tanto, opciones de inversión, que se están ofreciendo.

Para ello basta con crear el cierre convexo de los puntos representados en la gráfica. Un conjunto S es convexo cuando, para todo par de puntos pertenecientes al conjunto, se cumple que el segmento que une esos 2 puntos está contenido en dicho conjunto S. Así, habría que utilizar la parte del cierre convexo que une a los puntos.

Hacer un cierre convexo sobre la ilustración 4 quedaría de la siguiente manera:

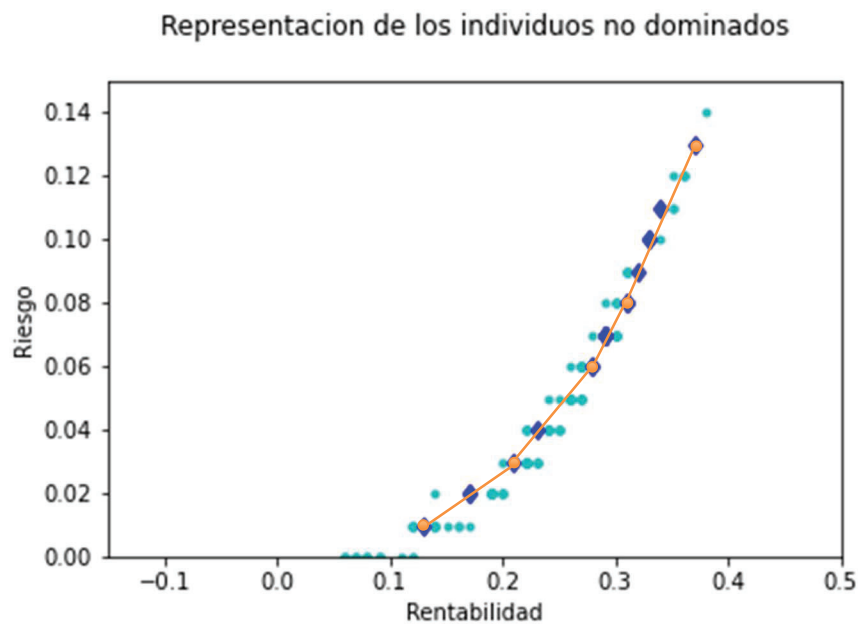


Ilustración 7: cierre convexo.

Así, las soluciones obtenidas para el problema de optimización planteado serían todas aquellas carteras situadas sobre la curva de color naranja. Hay elementos que actualmente se encuentran en el frente de Pareto que desaparecerían, todos aquellos rombos azules no marcados en naranja. Al ampliar el número de carteras que se que se están ofreciendo, aparecen nuevas soluciones que dominan a algunas de las actuales.

6 Análisis de Impacto

Una vez terminado el trabajo, se confirma haber cumplido todos los objetivos planteados inicialmente en la sección 4. Estos iban desde ser capaz de formular un problema genético basado en la optimización de carteras, a desarrollar un algoritmo que resuelva el problema y analizar los respectivos resultados. Así, a nivel personal, la realización de este trabajo ha tenido un gran impacto en mi aprendizaje sobre el tema. A lo largo de la carrera nunca había tenido la oportunidad de aplicar los conocimientos técnicos que iba adquiriendo de matemáticas y programación al área financiera. Además, he podido trabajar con bases de datos y herramientas que no había usado hasta el momento.

Por otro lado, considero que esta investigación también puede resultar una gran aportación a nivel empresarial y económico. El proceso de optimización reduce la pérdida de tiempo y recursos, mejorando así la eficiencia del negocio. Cambiar la gestión de procesos de una empresa puede suponer un cambio total en las ganancias obtenidas. El objetivo 8 de los Objetivos de Desarrollo Sostenible de la Agenda 2030 (ODS) se relaciona con promover el trabajo decente y crecimiento económico, a lo cual contribuye el hacer los procesos empresariales más eficientes y rentables.

Desde otro punto de vista, este proyecto también puede beneficiar a una población a nivel social y cultural. Hay muchas personas que probablemente no sepan como invertir su dinero por sí solos, por lo que un sistema digital que lo haga por ellos podría ayudar a dar el paso de empezar a invertir. Se tiene la confianza en la fiabilidad de un sistema automático para saber que la inversión escogida es la mejor, basada en todos los datos históricos económicos de las empresas.

Por último, también se podría aplicar este tipo de método a cualquier otro problema que implique buscar un equilibrio entre un beneficio y su respectivo coste. Por ejemplo, a la selección de políticas públicas relacionadas con reducir las emisiones de dióxido de carbono. Se pueden hacer políticas que disminuyan las emisiones netas de carbono más rápidamente, pero con el riesgo de reducir el crecimiento económico mundial. Así, se puede estudiar esta opción y compararla con otras políticas que ofrezcan combinaciones alternativas de crecimiento de corto plazo y objetivos medioambientales a medio-largo plazo. Con esto se estarían cumpliendo todos los ODS relacionados con la sostenibilidad, acción por el clima y medio ambiente.

7 Bibliografía

[1] Aparicio, L. (2019). El 'boom' de la gestión de carteras. elpais.com. Consultado el 7 de abril de 2022, de https://elpais.com/economia/2019/11/07/actualidad/1573118820_740551.html.

[2] Plata Godínez, A., Rodríguez González, H. and Vázquez Juárez, J. (2015). Algoritmo evolutivo para la optimización de portafolios de inversión [PDF]. Ciudad de México: Instituto Politécnico Nacional ESCOM. Consultado el 6 de marzo de 2022, de <https://tesis.ipn.mx/bitstream/handle/123456789/21135/Algoritmo%20evolutivo%20para%20la%20optimizaci%C3%B3n%20de%20portafolios%20de%20inversi%C3%B3n.pdf?sequence=5&isAllowed=y>.

[3] Redondo, A. (2020). GESTIÓN DE CARTERA: Qué Es y Cómo Crearla Correctamente [Video]. YouTube. Consultado el 13 de febrero de 2022, de https://www.youtube.com/watch?v=WLHp-K9kwrY&ab_channel=ElClubdeInversi%C3%B3n.

[4] Diversificación de riesgos: qué es - Diccionario de Economía. Eleconomista.es. (2022). Consultado el 14 de febrero de 2022, de <https://www.eleconomista.es/diccionario-de-economia/diversificacion-de-riesgos>.

[5] Sevilla Arias, A. (2015). Rentabilidad. Economipedia.com. Consultado el 31 de marzo de 2022, de <https://economipedia.com/definiciones/rentabilidad.html>.

[6] BBVA. Aula Banca Privada. La importancia de la diversificación. [PDF] Banco BBVA. Aula Banca Privada [PDF]. Consultado el 31 de marzo de 2022, de <https://www.bbva.es/estaticos/mult/importancia-de-la-diversificacion.pdf>.

[7] Riesgo de una cartera. Guiasjuridicas.wolterskluwer.es. Consultado el 8 de abril de 2022, de https://guiasjuridicas.wolterskluwer.es/Content/Documento.aspx?params=H4sIAAAIAAAEAMtMSbF1jTAAASmZQwNLtbLUouLM_DxbIwMDS0NDIwOQQGZapUt-ckhlQaptWmJOcSoAqoAccjUAAAA=WKE.

[8] Cuatro tipos de carteras para distribuir el patrimonio | BBVA. BBVA NOTICIAS. (2016). Consultado el 31 de marzo de 2022, de <https://www.bbva.com/es/cuatro-tipos-carteras-distribuir-patrimonio/>.

[9] ¿Qué es el ratio de Sharpe? - Self Bank. Selfbank.es. Consultado el 8 de abril de 2022, de <https://www.selfbank.es/centro-de-ayuda/fondos-de-inversion/que-es-el-ratio-de-sharpe#:~:text=Se%20calcula%20dividiendo%20la%20rentabilidad,rentabilidad%20en%20el%20mismo%20periodo.&text=Cuanto%20mayor%20es%20el%20Sharpe,ha%20tomado%20en%20la%20inversi%C3%B3n>.

[10] Magdalena, L. (2021). Modelización mediante técnicas evolutivas. Madrid: ETSIInf - UPM. Consultado el 12 de abril de 2022.

[11] Díaz Fernández, A. (1996). Optimización heurística y redes neuronales. Ediciones Paraninfo S.A. Consultado el 12 de abril de 2022.

[12] Martí, R., (2003), pp. 355–356. Multi-start methods. In Handbook of Metaheuristics, F. Glover and G. Kochenberger, Eds., vol. 57 of International Series in Operations Research & Management Science. Springer US. Consultado el 14 de abril de 2022.

[13] López, J., (2013). Optimización Multi-objetivo: Aplicaciones a problemas del mundo real [PDF]. Argentina: Facultad de informática, Universidad Nacional de la Plata, pp.47-58. Consultado el 20 de abril de 2022, de http://sedici.unlp.edu.ar/bitstream/handle/10915/27920/Documento_completo_.pdf?sequence=1&isAllowed=y.

[14] Real Academia Española: Diccionario de la lengua española, 23.^a ed., [versión 23.5 en línea]. Consultado el 8 de abril de 2022, de <https://dle.rae.es>.

[15] Entendiendo los algoritmos genéticos. Un caso de uso en el entorno organizacional. Planetachatbot.com. (2017). Consultado el 10 de abril de 2022, de <https://planetachatbot.com/entendiendo-los-algoritmos-geneticos-caso-de-uso-entorno-organizacional/#:~:text=Los%20algoritmos%20gen%C3%A9ticos%20nacieron%20en,en%20simular%20la%20selecci%C3%B3n%20natural>

[16] Arévalo González, S. (2016). Modelo de gestión de carteras de Markowitz usando algoritmos genéticos. [PDF] Argentina: Universidad Nacional de Entre Ríos. Consultado el 1 de marzo de 2022, de <https://repository.usta.edu.co/bitstream/handle/11634/11507/Ar%C3%A9valosantiago2016.pdf?sequence=1&isAllowed=y>.


[17] Peñuela Meneses, C. and Granada Echeverri, M. (2007). Optimización multiobjetivo usando un algoritmo genético y un operador estilista basado en un ordenamiento no-dominado (NSGA-II) [PDF]. Pereira: Universidad tecnológica de Pereira. Consultado el 11 de abril de 2022.

[18] Colomine Durán, F., Cotta, C. and Fernández, A. (2022). Algoritmos evolutivos multiobjetivo para selección de carteras de inversión con restricciones de cardinalidad [PDF]. Universidad Nacional Experimental de Táchira y Universidad de Málaga. Consultado el 11 de abril de 2022, de https://www.researchgate.net/profile/Antonio-Fernandez-Leiva/publication/236658740_Algoritmos_Evolutivos_multiobjetivo_para_seleccion_de_carteras_de_inversion_con_restricciones_de_cardinalidad/links/54201eec0cf241a65a1b05e4/Algoritmos-Evolutivos-multiobjetivo-para-seleccion-de-carteras-de-inversion-con-restricciones-de-cardinalidad.pdf.

[19] Damodaran, A. (2022). Damodaran Online: Home Page for Aswath Damodaran. Pages.stern.nyu.edu. Consultado el 18 de abril de 2022, de <https://pages.stern.nyu.edu/~adamodar/>.

[20] Puig, X. Beta de una acción: qué es - Diccionario de Economía. Eleconomista.es. Consultado el 18 de abril de 2022, de <https://www.eleconomista.es/diccionario-de->

Este documento esta firmado por



| | |
|-------------------------------|---|
| Firmante | CN=tfgm.fi.upm.es, OU=CCFI, O=ETS Ingenieros Informaticos - UPM, C=ES |
| Fecha/Hora | Thu Jun 02 00:56:38 CEST 2022 |
| Emisor del Certificado | EMAILADDRESS=camanager@etsiinf.upm.es, CN=CA ETS Ingenieros Informaticos, O=ETS Ingenieros Informaticos - UPM, C=ES |
| Numero de Serie | 561 |
| Metodo | urn:adobe.com:Adobe.PPKLite:adbe.pkcs7.sha1 (Adobe Signature) |