MEGI

Mestrado em Estatística e Gestão de Informação

Master Program in Statistics and Information Management

A Fresh Green Index in the World

Building and optimizing a Vegan and Sustainable Index Fund using a Genetic Algorithm and a Heuristic Local Search

Francisca Inês Augusto Lousão

Dissertation presented as partial requirement for obtaining the master's degree in Statistics and Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

A FRESH GREEN INDEX IN THE WORLD: BUILDING AND OPTIMIZING A VEGAN AND SUSTAINABLE INDEX FUND USING A GENETIC ALGORITHM AND A HEURISTIC LOCAL SEARCH

by

Francisca Inês Augusto Lousão

Dissertation presented as partial requirement for obtaining the master's degree in Statistics and Information Management, with a specialization in Risk Analysis and Management

Advisor: Doctor Mauro Castelli, Associate Professor, NOVA IMS

September 2022

“One of the first conditions of happiness is that the link between man and nature shall not be broken.”

Leo Tolstoy

ACKNOWLEDGMENTS

Sendo este o término de uma das etapas mais desafiantes da minha vida, até à data, resta-me agradecer a todos aqueles que, de certa forma, me incentivaram e apoiaram a concluí-la com tanto gosto.

Sem especial ordem, agradeço primeiramente ao Miguel. Obrigada por prescindires tantas vezes do teu (e do nosso) tempo para me estimulares a ser mais e melhor. Por todo o apoio incessante desde o primeiro dia, dos projetos mais pessoais aos académicos. Sem ti, tinha sido demasiado difícil.

Aos meus pais. Obrigada por me proporcionarem, desde sempre, a melhor educação possível, que me tornou na pessoa que sou hoje. Pela motivação, que me incentivou, desde cedo, a querer chegar mais longe. Pela mão que me deram e que nunca me deixou ficar mal. Esta tese é vossa e graças a vocês.

Aos meus avós, que me apoiam incessantemente. Por terem sido sempre uma presença tão boa e tão constante na minha vida.

À Sara, a minha irmã. Obrigada por, sem saberes, me fazeres ver que tenho em ti a maior e mais verdadeira amizade. Obrigada pelo apoio constante e por todo o *tough love*. Agradeço-te imensamente pela companhia e pelo encorajamento constantes.

À minha restante família e aos meus amigos, dos mais antigos aos mais recentes. Não conseguindo mencionar-vos individualmente, agradeço-vos por, de perto ou de longe, me motivarem com palavras carinhosas que permanecerão comigo para sempre.

Agradeço, por último, mas não menos importante, ao meu orientador, Mauro Castelli, pelo ânimo com que sempre me ofereceu *feedback*. Agradeço-lhe Professor, por acreditar neste projeto tanto quanto eu.

Dedico as minhas últimas palavras a mim mesma. Por nunca ter desistido, mesmo quando esse parecia ser o caminho mais fácil. Por me conhecer desde sempre como lutadora e devota àquilo em que acredito. Tenho em mim que quando se faz algo de que se gosta, tudo se torna mais simples.

On to better things, it was about time!

ABSTRACT

The curiosity of investors regarding Environmental, Social and Governance (ESG) factors has seen a growth in the last few years (Alcoforado, 2016), as the world faces some of its biggest problems to date, such as Climate Change and Ecological Collapse. As these issues are not to be taken lightly, individuals have started to act, in the hopes of creating a 'greener' world. As individuals hope to align with principles such as Sustainability and Veganism, the proposed project hopes to build a Vegan and Sustainable Index Fund, as "An investment is not an investment if it is destroying our planet." (Shiva, 2017).

The aim of the proposed work is, consequently, to build and optimize an Industry and Geographical diversified Index Fund, using a Genetic Algorithm (GA), demonstrating this through the incorporation of Vegan and Sustainable companies, in addition to the global top-50 ESG ranked firms. Index Funds, which are mutual or Exchange-Traded Funds (ETF), are known to be passively managed portfolios, which have been broadly used in hedge trading (Orito, Inoguchi, & Yamamoto, 2008).

This study uses historical data from Vegan, Sustainable and ESG-ranked companies as sample data, replacing traditional optimization methods using a Genetic Algorithm.

The GA method was applied to a sample of 61 assets, regarding vegan and sustainable companies, further obtaining a well-diversified and non-centred asset allocation. The obtained results confirm the possible efficiency of genetic algorithms, given their high-speed convergence towards a better solution. A few functions were presented in the algorithm, for example the penalty function method, to perform portfolio optimization which expects to maximize profits and minimize risks. Some flaws have been identified in regard to the method applied.

KEYWORDS

Index fund; Vegan; ESG; Genetic Algorithm; Heuristic Local Search

INDEX

LIST OF EQUATIONS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

ANN	Artificial Neural Networks
ARMA	Autoregressive Moving Average
ARIMA	Autoregressive Integrated Moving Average
BP	Basis-point
CAGR	Compound Annual Growth Rate
DCF	Discounted cash flow
DL	Deep Learning
EA	Evolutionary Algorithms
EM	Emerging Markets
ESG	Environmental, Social and Governance
ETF	Exchange-Traded Fund
FTSE	Financial Times Stock Exchange
GA	Genetic Algorithm
GP	Genetic Programming
HLS	Heuristic Local Search
LR	Linear Regression
ML	Machine Learning
MOEA/D	Multi-Objective Evolutionary Algorithm based on Decomposition
PETA	People for the Ethical Treatment of Animals
QIGA	Quantum-inspired genetic algorithm
SASB	Sustainability Accounting Standards Board
SVM	Support Vector Machines
TCFD	Task Force on Climate-related Financial Disclosures
TE	Tracking Error
VEGAN	US Vegan Climate Index
WCED	World Commission for Environment and Development

1. INTRODUCTION

1.1. Background and Problem Identification

In recent years, several publications have grabbed the attention of the public, concerning Climate Change and Animal Cruelty. It is well established that in case our generation does not change its erroneous actions, the earth will not be fit for future generations to thrive. In 2019, Amina Mohammed, United Nations Deputy Secretary General, penned a post named "Making decisions to save our planet now, for future generations" which underlines the urgency to act. Even though the above-mentioned topics can be delicate, it is essential to introduce them to the overall humanity and specific industries which have been neglecting the health and livelihood of the globe's inhabitants (McMichael, Campbell-Lendrum, Corvalán, Ebi, Githeko, Scheraga & Woodward, 2003).

To this day, various worrying statements have been released. One of these, by the well-known Kip Andersen, mentions the following: "Cows and other farmed animals produce a substantial amount of methane from their digestive processes. Methane gas from livestock is 86 times more destructive than CO_2 from vehicles.". Over the years, the beforementioned has been confirmed through different organizations, namely the Food and Agriculture Organization of the United Nations which, in 2006, claimed "Animal agriculture is responsible for 18 percent of greenhouse gas emissions, more than the combined exhaust from all transportation." (Andersen & Kuhn, 2014).

Furthermore, vegan, vegetarian, and plant-based diets are associated with great reductions in greenhouse-gas emissions. A shift to a plant-based life is believed to reduce mortality, as well as greenhouse gases by 10% and 70%, respectively, by 2050 (Physicians Committee for Responsible Medicine, n.d.). Additionally, the global plant-based food market is projected to reach a worth of around 74.2 billion dollars by 2027 (Meticulous Research, 2020). The success of such sector stands as the reason investors are particularly interested, as they witness the growth of companies within the segment.

The above stated can be substantiated by several studies. One of these studies is the one developed by Research and Markets which, in March 2021, released a report concerning the Vegan Food Delivery market, underlining its size, share, the impact Covid-19 represented and, lastly, a forecast regarding 2027. This forecast can be shown below (Figure 1), representing the result of a comprehensive research, using in-depth data and contemporary analysis. This was accomplished at a global, regional, and key country level, further split by different sub-segments of the named industry. As seen below, the industry is expected to grow around 1.99 times, considering 7 years, at a Compound Annual Growth Rate (CAGR) of 10.3%.

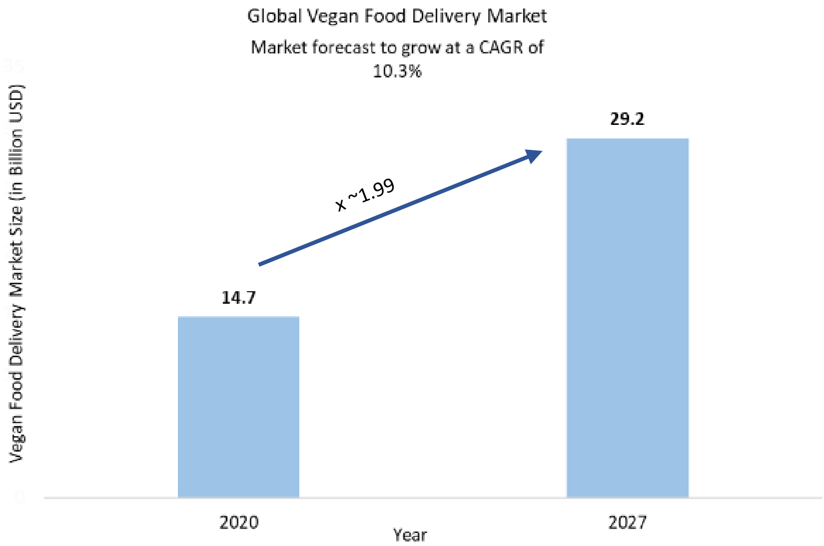


Figure 1. Global Vegan Food Delivery Market. Adapted from a published report entitled '2021 Vegan Food Market'.

As financial markets grow bigger, more investors consistently attempt to build systematic approaches intended to predict prices and movements (Vieira, 2020). The development of this project will support individuals who wish for an outsider to assume responsibility for screening companies and selecting the best investments. Consequently, an Index Fund will be created, considering vegan and ESG factors, the latest aligning positive returns with long-term impacts, not only on society but also on the environment and the performance of a business.

It is worth mentioning that, although one Vegan Climate Index has been developed, the United States Vegan Climate Index (VEGAN), it is believed to be crucial for investors to buy from geographically diversified stocks, as it lowers the overall level of investment risk (Barclays, n.d.). Figure 2 showcases the possibility of portfolio diversification when considering different geographies. It should be once again noted the percentages shown are purely illustrative, as these should not be considered as guidance to be followed.

This should not be overlooked as each geographic area can act autonomously, considering the impact of external factors. A gap will be filled in the research field by creating additional knowledge through the intended universal Vegan and Sustainable Index.

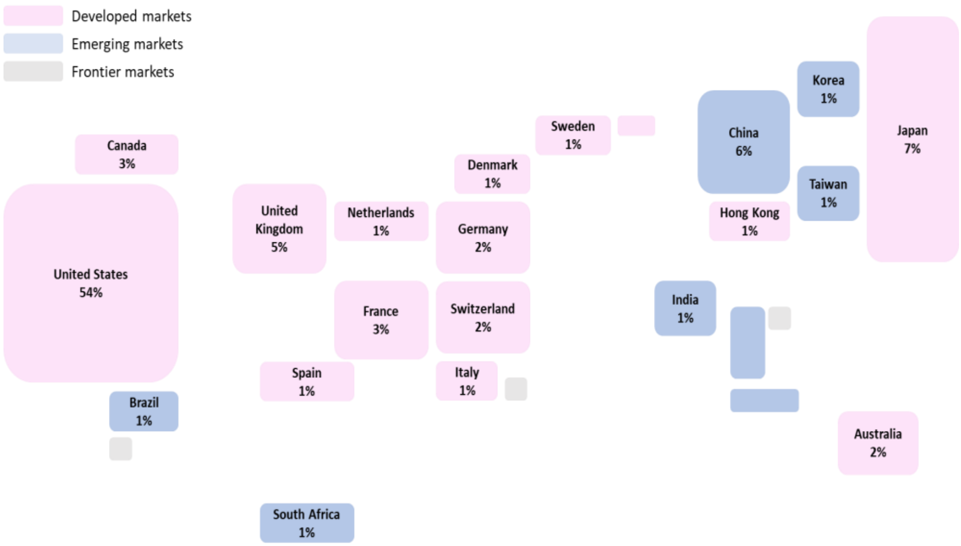


Figure 2. Adapted from image retrieved from Parsec Financial. Data provided by Bloomberg. For illustration purposes only.

In addition to the above stated, there is another index which, according to Stoxx, "(…) offers a representation of the leading global companies in terms of environmental, social and governance criteria, based on ESG indicators provided by Sustainalytics. (…)". This index, entitled 'STOXX Global ESG Leaders', is made of the following three ESG sub-indices: (i) the STOXX Global ESG Environmental Leaders; (ii) the STOXX Global ESG Social Leaders; and (iii) the STOXX Global ESG Governance Leaders indices.

1.2. Research Objectives

The thesis, titled 'A Fresh Green Index in the World: Building and optimizing a Vegan and Sustainable Index Fund using a Genetic Algorithm and a Heuristic Local Search', studies the Vegan Stock Market, emphasizing the theme of optimizing a global and sector-diversified index, built around Vegan and Sustainable Publicly Traded Stocks, in addition to the Top-50 ESG ranked Publicly Traded Stocks.

To achieve the beforementioned fund, stocks were chosen thoroughly. In the data collection phase, there was a retrieval of datasets, with subsequent cleaning and treatment processes. Obtaining the required data with success lead to the implementation and optimization of the algorithm, therefore achieving the proposed index. Prior to this, the *N* issues included in the index will be selected by an HLS. After obtaining the required index, an analysis of results was developed, in addition to a conclusion.

It should be addressed the factor which differentiates the index to be presented is the vast link between its vegan nature, as well as its ESG background (Figure 3).

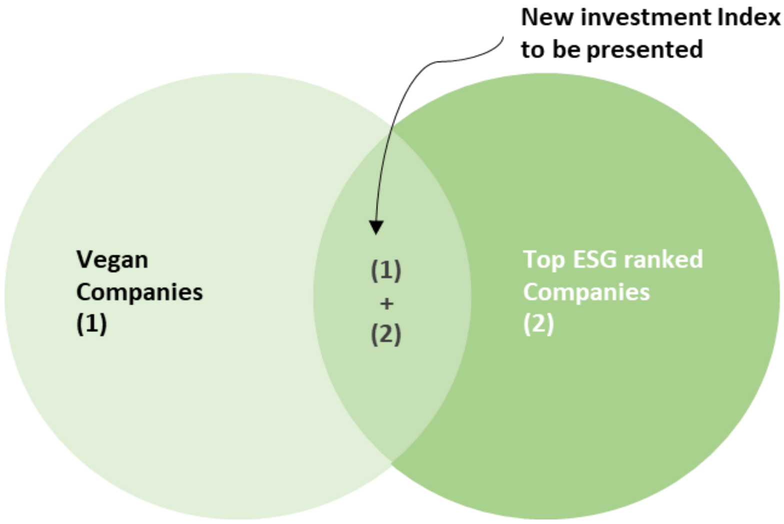


Figure 3. Venn diagram resembling the placement of the index to be developed. Source: own research.

Nowadays, companies around the world have been employing Advanced Analytics to sustain their decision-making process. On the other hand, conventionally, businesses have used Statistics and Business Intelligence for this purpose. However, as technology is advancing, more complex models are gaining popularity among its users. The main reason behind the increasing interest in Machine Learning (ML) and Deep Learning (DL) models is the fact these achieve superior prediction accuracy (Napiórkowska, 2019). For this reason, the proposed portfolio was achieved through the implementation of a GA portfolio scheme.

Readers should, therefore, expect to be presented with a GA portfolio scheme to optimize the established index fund, as Genetic Programming (GP) "(…) has been considered one of the most successful existing computational intelligence methods and capable to obtain competitive results on a very large set of real-life application against other methods." (Oliveira, 2016:29). The mentioned scheme was expanded and, consequently, delivered an optimal selection of stocks using key variables. It is also an objective of this thesis to present information in tables, as well as graphs.

1.3. Study Relevance and Importance

Climate Change and Animal Cruelty have been enormous problems for several years. According to the European Environment Agency, "Climate change is one of the biggest challenges of our times.".

The option of associating the investment sector with the Vegan industry is due to a substantial personal desire, along with a worldwide need for transformation. Moreover, it is worth stating the immense importance this practical study brings, not only to sustainability-driven investors, but also to the financial sector, given financial agents are able to profit from ML based systems to plan and monitor financial investments more accurately and, therefore, achieve higher returns. As mentioned, this thesis adds new meaning to existing information, taking into consideration its output will generate unseen conclusions in the beforementioned sectors.

1.4. Dissertation Design

Taking into consideration the end goal this research conveys, the methodology applied, which is data-driven, was divided into 5 stages, being those the following: (i) Identifying the problem; (ii) Developing a solution to the identified problem; (iii) Composing a research design; (iv) Collecting, treating, and analyzing retrieved data; and (iv) Assembling and improving the report and, subsequently, its final presentation. The first stage, (i), identifies the question and highlights its reasonings. The ultimate identified research question is “What will be the optimized Index Fund, only considering the incorporation of Vegan and Sustainable companies, in addition to the global top-50 ESG ranked firms?”. The second phase, (ii), saw the development of a ML Genetic Algorithm, used to optimize the desired Index Fund. Throughout the development of stage 3, the exploratory study commenced, considering the identification of key variables, along with the sample size. This phase accounted for the definition of the structure of the equation to be introduced in the model, as well as the different variables to be collected from the sample. Phase 4 is the central phase. The data, retrieved from Yahoo Finance was pre-processed to be inserted into the algorithms. In the final phase, the evaluation of the obtained results of the questions defined was established. The last phase consists of a summary of information in a written report, to be presented to the public.

1.5. Expected Contributions

As the outcome of this research is believed to be successful, there are three main contributions. First, the work is expected to contribute to the advancement of human knowledge through the addition of a new Green Index. Second, future research should be pinpointed, persuading individuals into developing a leading topic. Finally, the biggest contribution would be to globally impact the actions of future investors, attaining a healthier planet.

The author would like to address that considering Index Funds, mutual or Exchange-Traded Funds (ETF), are known to be passively managed portfolios, the concepts will be intertwined solely for the purpose of this study.

2. LITERATURE REVIEW

There is a vast amount of literature regarding the forecasting of stock markets including distinct approaches ranging from advanced Machine Learning algorithms to traditional statistical models. During the writing of this dissertation different studies were conducted and deeply analysed to identify the best approach to be followed, as distinct markets and resources require specific models.

To form a portfolio consisting of the highest expected return for a specific level of risk tolerance, modern portfolio theory delivers a sophisticated paradigm. Harry Markowitz, the inventor of the beforementioned theory, initially formulated the fundamental theorem of Mean–Variance portfolio framework. Such framework is known to explain the trade-off between mean and variance, each being representative of the expected returns and risk of a specific portfolio, respectively. Although Markowitz's theory only uses the mean and variance to describe the characteristics of return, it became one of the foundations of modern portfolio theory (Fama, 1970; Hakansson 1970, 1974; Merton, 1990; Mossin, 1969). Following this initial step, great progress in portfolio theory and practice was made.

The valuable literature on optimization algorithms (utilized to achieve a better optimal solution) uses, for the most part, the optimization methods of Markowitz, such as the mean-variance model, as well as the factor and neural network models, among others.

It was in 1984 that (Perold, 1984) suggested using the mean-variance model for index tracking. (Haugen & Baker, 1990) believed beta, volatility and determination coefficients could be used to calculate the tracking capability of said model. (Gaivoronski, Krylov, & van der Wijst, 2005) examined the difference between tracking a portfolio and a benchmark index with several methods. For their part, Yu et al. (2006) improved the Markowitz model for index tracking.

Secondly, the factor model can also be used to track indices. In 1989, Rudd proposed a portfolio selection method, then creating a portfolio with beta, relevant to an index. Corielli and Marcellino (2006) presumed stock prices as being driven by the factor model. Canakgoz and Beasley (2009) used a mixed integer programming model.

Some scholars, as indicated in the following paragraphs, applied cointegration and other optimization methods to track indices. Cointegration was used to optimize the weight of the tracked portfolio. The tracking error (TE) measurement was used to assess the difference between the return of the tracked portfolio and the return of the target index, further evaluating the performance of index tracking. Meade and Salkin (1990) used the quadratic programming to minimize this TE.

Jansen and van Dijk (2002) studied the tracking error minimization by limiting the number of constituent stocks in the tracking portfolio. Gaivoronski et al. (2005), for their part, discussed numerous measurement methods for tracking this error.

Jeurissen & van den Berg (2005) produced an index tracking approach using a hybrid genetic algorithm, defining TE as a measure used to evaluate fitness (Mezali, 2013). The weights associated with a particular stock in the tracking portfolio were, therefore, decided by using a genetic algorithm. The mentioned authors presented an index tracking model, one which was based on the reduction of the variance of the difference between the return of the tracking portfolio and index return. This model, a quadratic program, did not consider transaction costs, to be mentioned further along in this thesis.

To sum up, most academics have made use of traditional optimization algorithms, including the mean-variance and factor models, previously mentioned, in addition to the cointegration optimization method. Only a small number of academics researched other methods, such as the neural network algorithm.

A drawback was that, as most specialists built the mentioned tracking models based on stock prices or returns, they did not, however, i. Compare the tracking model based on stock prices with that based on stock returns; and ii. Consider the sample stock selections of the tracked portfolio. Instead, researchers leaned on the rank of market value or even random sampling to select sample stocks. The results of these models could, therefore, be affected as some valuable information could be overlooked.

Regardless of the algorithm in use, it is known that ML can be "(…) a respectable instrument that can allow the better understanding of all relationships between underlying variables (…)." (Vieira, 2020:7).

As for Genetic Programming, which belongs to the ML domain, it is not a new topic considering its origins go back to, at least, the 1950s. Ever since then, there has been a great number of research papers on the topic. One of the first descriptions of this optimization algorithm was made by Koza (1992) stating that "Genetic Programming is a problem-solving approach in which computer programs evolve producing new individuals, based on Darwinian principle, changing the syntax of the parents without taking into account the semantics of the individuals, in order to solve a problem." (Koza, 1992).

In Genetic Programming, one of the most important insights was the Genetic Semantic (Koza, 1992). This assessment provided new insights concerning the relation between program syntax and semantics, search operators and fitness landscape. Nevertheless, this methodology held a strong limitation because it produced offspring larger than their parents, hence creating very large-sized programs. To overcome this limitation, Vanneschi, Castelli, and Silva (2013) implemented a new approach, making it feasible to use such operators in a useful and efficient way. This new approach was further explored, seeing its applications presented in real life problems related to the electricity and pharmacokinetics fields (Vanneschi, Castelli & Silva, 2014). The most common application of GP is the forecast of stock prices, as the excellent results show how much of a powerful instrument for forecasting prices it can truly be.

3. SUSTAINABILITY, THE FINANCIAL WORLD AND MACHINE LEARNING

3.1. Sustainability in the Corporate World

Organizations are progressively going further beyond volunteering and philanthropy to implant sustainability actions into their core business strategies. Prominent transnational corporations around the world perceive sustainability as being a business priority, considering the current worldwide environmental and social challenges (Mercer, 2014).

3.2. Evolution of Sustainable Development

It is of extreme importance to mention the broad, and dialectical concept of Sustainable Development which offsets the necessity for economic growth, with social equity and environmental protection. The term, which was popularized in the book entitled 'Our Common Future', in 1987, by the World Commission for Environment and Development (WCED), described the concept as a development that understood the needs of present-day generations, without compromising the capability of future generations to fulfill their needs. In the beforementioned book, the WCED recognized the success of such concept could not be left to policymakers and government regulators. The industry had, as it does to this day, a considerable role to play. The opinions defended by the authors were that, even though firms had been the engines for the economic development, they also needed to become more initiative-taking in regard to balancing the drive between social equity and environmental protection. This was due to the fact organizations had been the source of certain unsustainable conditions. In addition, firms had access to the required resources to address the problems. The reaction from the industry came in stages as everyone debated the idea of what sustainable development in action should consist of. In 2021, Forbes published an article titled 'Why Corporate Strategies Should Be Focused on Sustainability', urging for an action to be undertaken as "Although 90% of executives think sustainability is important, only 60% of companies have a sustainability strategy." (Forbes, 2021). The article defended the need for growth in corporate sustainability as this can "(…) add brand value, meet consumer demands, increase efficiency, attract valuable talent and create new opportunities." (Rafi, 2021).

3.3. Introduction to ESG factors

Progressively, ESG (an abbreviation for Environmental, Social, and Governance) considerations, understood to be non-financial, such as animal harm and exploitation, as well as fossil fuel, environmental damage, or human rights, are being applied as part of an analysis process which hopes to detect material risks and growth opportunities in global and sector-diversified companies.

Although ESG metrics are not usually mandatory in financial reporting, companies increasingly make disclosures in annual or sustainability reports (CFA Institute, n.d.). A substantial amount of institutions, such as the Sustainability Accounting Standards Board (SASB) and the Task Force on Climate-related Financial Disclosures (TCFD), among others, are currently working to define materiality and develop standards towards simplifying the future incorporation of the abovementioned factors into an investment process.

3.4. Preliminaries of a Vegan Company

According to The Vegan Society, veganism is considered "(…) a philosophy and way of living which seeks to exclude—as far as is possible and practicable—all forms of exploitation of, and cruelty to, animals for food, clothing, or any other purpose; and by extension, promotes the development and use of animal-free alternatives for the benefit of animals, humans and the environment." (Society, n.d.). In case a firm wishes to distinguish itself as vegan, it should be well-aware of keeping the abovementioned practices excluded from its activities.

A vegan company, placed, for example, in the food industry, produces alternatives to fish, meat, dairy, and eggs, in addition to ingredients such as sweeteners. These products can derive, for example, from plants including vegetables, fruits, whole grains, nuts, seeds and/or legumes. Nonetheless, in the textile industry, such companies seek to find vegan fabric alternatives to those sourced from animals, ranging from silk, fur, feathers, and cotton to leather. There should be a warning regarding the fact that, even though a company is vegan, it does not necessarily mean it is in the top-50 ESG. Thus, the reason to keep both in the thesis.

3.5. Preliminaries of Investments

There are two major currents in the analysis of stock behaviour in the capital market. Not being an agreement among the advocates of each, both fundamental and technical currents have their strengths and weaknesses, both being fallible. Therefore, the most correct way to interpret the signals that the market gives an investor is to reconcile the two analyses before investing.

3.5.1. Investment profile

Silva M. (2022) expressed in one of the most successful investment books in Portugal to date that, with the ever-growing globalization of investments, it has become necessary to adjust the exposure to an equity portfolio to the investor's profile. An investor's profile may have, among other restrictions, religious, environmental, or others. As a result of this occurrence, there is now a need to create specific, thematic indices, to provide a dynamic benchmark, therefore selecting companies which pay high dividends or follow certain internal governance criteria. This need has arisen, as well as the above stated, to provide a benchmark for investors who do not wish to buy shares in companies linked to production of military hardware or those which do not sustain a policy of respect for the environment.

As environmental concerns are currently a daily worry, these guide the investments of many managers, who choose to exclude stocks from their portfolios that are opponents to the well-being of the environment.

Among the vast index offering, it is possible to find some which hold an investment politic based on specific themes, such as religious beliefs, or ecological background. Below (Table 1) showcases 10 different indices, each belonging to a specific topic (Silva M., 2022).

Theme	Index
Sustainability	STOXX Global ESG Environmental Leaders
	STOXX Global ESG Social Leaders
	STOXX Global ESG Governance Leaders
Private Equity	STOXX Europe Private Equity 20
Sports	STOXX Europe Football
	STOXX Global Grand Prix
Rare earth metals	STOXX Global Rare Earth
Exposure to emerging markets (EM)	STOXX Global 1800 EM Exposed
Religion	STOXX Europe Christian
	STOXX Europe Islamic

Table 1. Thematic indices. Adapted from the 2022 version of 'Bolsa: Investir nos Mercados Financeiros'.

3.5.2. Fundamental analysis

As introduced by (Silva, 2022) in the 2022 version of his proclaimed book, a fundamental analysis gathers the so-called real data from the companies, drinking from their financial records the elements which permit to achieve a conclusion as to the true value of the company and, consequently, the fair value of its shares. This evaluation is based on updating the future cash flows of the company to the present moment, i.e., determining the present value of all the money the company can generate annually over a certain period, representative of its future life. This model is frequently known as discounted cash flows, commonly known as DCF. The sum of these cash flows discounted to the present time at an interest rate thought to be what an investor would require assuming the business risk of a given company, calculates the supposed market value of the company. By dividing this value by the number of shares which make up the Capital Stock, the fair price for each share is obtained. Many entities believe fundamental analysis is supported by great quality data as it allows a reliable conclusion to be reached about the real value of a stock (CFA Institute, 2022). However, the complex calculations and assumptions involved make it more time-consuming to prepare and interpret. At the speed at which the market currently trades, it would be unfeasible to base a decision on whether to invest in a particular stock solely based on a fundamental analysis of the financial records of a company. However, there are other market indicators, also based on assumptions that fit the fundamental analysis of the companies and allow a simpler use and a faster and easier calculation to work with. This is the case of market multiples, for example.

3.5.3. Risk measures

The risk of an investment portfolio deserves constant vigilance. For a private investor, the care to be taken is no less than that required for Asset Management companies. These companies, besides obeying to a series of legal rules, permanently monitor the investments made. It is important to consider volatility as a fundamental concept in risk analysis, as suggested by (Silva, 2022). It gives those interested the amplitude of price variations of a security, in case its quotation varies in a very wide value range, or if, on the contrary, it trades in a tight price band. Volatility is usually analysed using the statistical measure of the standard deviation calculated over a set period, usually 1 or 3 years. The higher the standard deviation, the greater the volatility, and therefore the risk, of an asset or investment portfolio. The larger the range of price variation, the more the price of a security varies in each period, therefore the more volatile it is, the less stable and riskier it is.

One of the most common ways to assess whether an investment portfolio was managed considering not only the performance of assets, but also their risk, is the Sharpe ratio. This indicator measures the risk-adjusted performance, and the higher its value, the better the adjustment (thus, the better the manager). A value above 1.5 shows some care in the selection of asset risk. Values below 1 represent that a too high risk was taken for the profitability that the portfolio presents. Besides Sharpe's index, there are other statistics, such as Treynor's Ratio, Jensen's Alpha and M^2 as (Scholz & Wilkens, 2005) suggest. When faced with the need to evaluate the performance of distinct portfolios, the greatest one will be the one that presents the highest values in each of the indicators. It is usual for Asset Management companies and Investment Fund Management companies to present these statistics for each one of the portfolios and investment funds under commercialization (Silva, 2022).

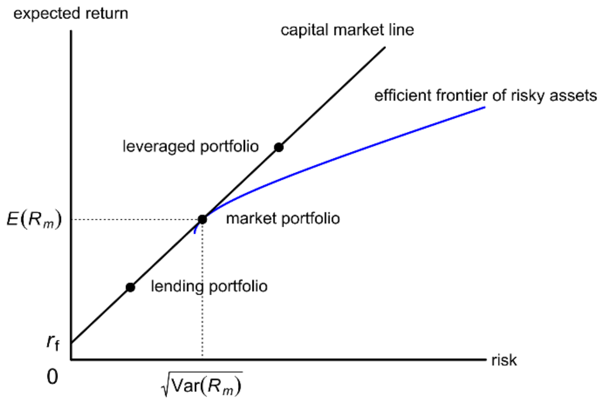


Figure 4. Sharpe Ratio in graphical format. Source: seekingalpha.com.

3.5.4. The Alpha and Beta parameters

Another factor which measures the success of an investor is related to the return generated by the management of an investment portfolio against the return provided by the market benchmark, usually expressed by an index.

A market index indicates the movement of a specific market, revealing whether the market capitalisation of all companies listed, for example, has increased or decreased. In addition, a market index is a series of pure numbers, used for comparing between different index numbers. Index numbers are constructed having in mind a fixed base date and base value. Indices are calculated on an entire market, as well as being available for a particular sector of a market (Mezali, 2013). Some securities vary in a wider range than others and are therefore more volatile, generating a return higher or lower than the market average, the benchmark. These deviations from the average provide managers and investors with the possibility of obtaining returns higher than the index to which they belong. In practice, the profitability of a stock can be measured and explained by two parameters: the alpha and beta parameters.

3.5.4.1. Alpha Parameter

The Alpha of a stock measures the return generated by this stock, which is independent of the market, that is, which is not explained by the performance of the market where it is traded. Thus, if faced with a positive Alpha value, it is a sign the stock contributes an additional return in the remuneration of the investor who takes the specific risk of investing in this security instead of investing in the index of which it is part.

3.5.4.2. Beta Parameter

The Beta parameter measures the sensibility of a stock or portfolio to market variations, considering the variation of the benchmark, which, as we mentioned before, is naturally the reference index. This sensitivity of a security to the variations of the index a known by the name of market risk or systematic risk. Three situations may occur, those being the following:

a. The stock has a β equal to 1 (β=1). In this case, the security (or portfolio of securities) tends to present a variation equal to the variation of the market. For example, if the index varies 5%, the variation of the stock will also be 5%;
b. The stock has a ß greater than 1 (β>1). This fact translates into a greater sensitivity of the stock, which will register variations greater than those of the market. For example, if β-1.5, this means that if the market varies 5%, the stock will have a variation of 7.5% (5% x 1.5); and,
c. The stock has a β less than 1 (β<1). Following the same reasoning, in this case the stock's variation will be less than the market's variation. If β=0.5, the variation of the security will be half of the market variation. In practice, we can say that the more conservative stocks, such as energy, retail and pharmaceutical, among others, normally have a Beta lower than 1, while the more aggressive stocks, such as technology, natural resources, will in principle have a β higher than 1.

$$Return\ of\ a\ stock\ or\ portfolio = \alpha + (\beta * \text{Market Performance}) \qquad (1)$$

The importance of the β parameter is, fundamentally, to know what percentage of the variation of a stock or portfolio of securities, is explained by the variation of the market.

Again, it is usual for Asset Management and Investment Fund Management companies to present the Alpha and Beta values of their portfolios and investment funds.

Another relevant indicator is the R^2. It indicates to what extent the performance of a certain stock or portfolio depended on the evolution of the benchmark index. This value can vary between 0 and 1. If it is 0.5, this means that the performance of a certain stock or portfolio is explained in 50% by the evolution of the benchmark index.

3.5.5. Diversification

It is crucial to introduce the concept of diversification in portfolios, a notion supported by Markowitz in 1952. When a portfolio only holds a short amount of assets, it may be subject to a high degree of risk, as it is represented by a relatively large variance in return. The variance of the return of a certain portfolio can be decreased by adding distinct assets in the portfolio. This process is referred to as diversification. For example, as a beginner investor, someone could invest their entire wealth, or life savings, in one stock. In case this occurred, the investor would be exposed to a company-specific (one specific stock can do better or worse than others operating in the same market) or market risks (for example, the Financial Times Stock Exchange (FTSE) 100, based in London, where the stock is traded). Nevertheless, in case the portfolio was to be expanded, hence including diverse assets or stocks, there would be a reduction to the overall risk, as addressed by (Etchegaray, 2021). Consequently, diversification is crucial when investing, both industry and geography-wise. Although an investor hopes to reduce the risk of an investment by doing so, it is never certain to happen, as one cannot expect the future.

3.5.6. Transaction Costs

When considering investing in stocks, also known as equities and shares, there is a transaction cost associated with buying or selling the asset. Transaction costs are usually presented in basis points (one basis point (bp) being 1/100 of one percent). A transaction cost can be a commission given to an intermediary third party, varying by stock. It typically alters according to the liquidity (easily purchased/sold) the stock holds and the amount to trade.

The trade-off between getting to a better position and the existence of additional transaction costs is a common feature of portfolio construction. The decision should, however, be based on whether the gain from trading covers the transaction costs.

3.5.7. Index Tracking

Lately, there has been a certainty stating low-priced passively managed index funds deliver the highest risk-adjusted returns in every category of mutual funds (Bogle, 1998). Index fund management is a strategy which allocates stock while supplied with critical index tracking skills. In addition, it attempts to reproduce the performance of a certain benchmark index. Typically, index funds do not contain every stock comprising the index. Nonetheless, these are intended to replicate the benchmark index with a relatively small number of stocks, which should be easily managed and controlled in the capital market. A TE which is measured by TE volatility, is also involved in this activity. The TE consists of the sum of the deviations of returns of the replicating portfolio from the benchmark index. The volatility level of the TE should be minimized to the lowest possible value, since it produces the closest possible returns to the benchmark returns (Clarke, Krase, & Statman, 1994; Konno & Yamazaki, 1991).

There are distinct types of TE measures available, such as quadratic, linear, and absolute. The quadratic measure is commonly favored in comparison to the others, being that it holds several desirable statistical properties (Roll, 1992).

3.6. Machine Learning

Even though Machine Learning does not hold a set description, it can be described as "(…) a computational field that strives to create computer programs that repeatedly improve with their own experiences on data." (Vieira, 2020). In 1997, Tom Mitchel described ML in the book entitled 'Machine Learning', as "A computer program (…) said to learn from an experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E." (Mitchel, 1997:2).

The use of ML techniques in financial markets have been successful, in specific when referring to the prediction of financial time-series. Nowadays, researchers are seeking to develop intelligent algorithms which are able to obtain the hidden patterns inherent to the stock market, for example, thus predicting the behaviour of stock prices more efficiently.

The main objective of Machine Learning is to distinguish patterns in a set of data, using those to predict about unobserved data. The result is the ability to predict accurately when presented with new data. When a new model is designed, it is later presented to the training data, hence learning crucial guidelines based on the specific data under study.

The field of ML includes a distinct amount of several other fields, for example mathematics, probability, statistics, among others.

The approaches used by academics in this field can be divided into two main classes, being those the following:

a. Econometric models developed centred on statistical approaches such as the Linear Regression (LR), Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA) models. Even though these models offer an effortless implementation, they hold a non-realistic fact, assuming financial time-series data follow a linear pattern and is stationary.
b. Predictive models which forecast market stock prices based on intelligent algorithms resembling biological processes, further solving nonlinear and complex problems. A few

examples of the beforementioned algorithms are GP, Artificial Neural Networks (ANN) and Support Vector Machines (SVM) (Oliveira, 2016).

In the following section, the most common machine learning algorithm will be presented, often used to predict the stock market, hence giving special emphasis to the Genetic Algorithm model.

3.6.1. Genetic Algorithm

The GA, invented by (Holland, 1975), is a stochastic optimization technique and a search algorithm which is established on the survival of the fittest among string structures (Goldberg, 1989). The development of such algorithm evolved from biology research so as to drive it to a (near) optimal solution.

In most GA optimization applications, each chromosome represents the weight of an individual stock which is optimized to reach a possible portfolio solution. A fitness value is affixed to each chromosome, defining the level of representation (good/bad) of each chromosome. As addressed by (Adeli & Hung, 1995) through the implementation of mutation, crossover values, and natural selection, there is a convergence to a population only containing chromosomes with good fitness (Orito, Yamamoto, & Yamazaki, 2003; Xia, Liu, Wang, & Lai, 2000).

A GA enables the calculus of intensive mathematical problems rapidly, easy to run with realistic constraints, further providing reasonable solutions. The implementation of a Genetic Algorithm with a selection and cross mutation is used to enrich the feature selection for certain classifiers.

A genetic algorithm includes evaluation and evolution stages. Throughout the evaluation phase, various classifiers of a specific type are trained and tested. The genetic algorithm analyses the best configuration to use among said classifiers. The evolution process, previously mentioned, includes phases such as selection, combination, mutation, and completion. Each step holds a specific chance of succeeding which is based on a random number for the specific species. Both selection and combination have a 5 percent chance of success, whereas the mutation has a 10 percent chance of success. The completion phase stabilizes, at last, the population size. As "Genetic algorithms belong to the larger class of evolutionary algorithms (EA), which generate solutions to optimisation problems using techniques inspired by natural evolution, such as inheritance, mutation, selection, and crossover." (Bashir, 2015), it is of great importance to introduce the concepts of fitness function, generation of a population, mutation, selection and lastly, crossover. Below a flowchart can be found, which resembles the operation of a Genetic Algorithm.

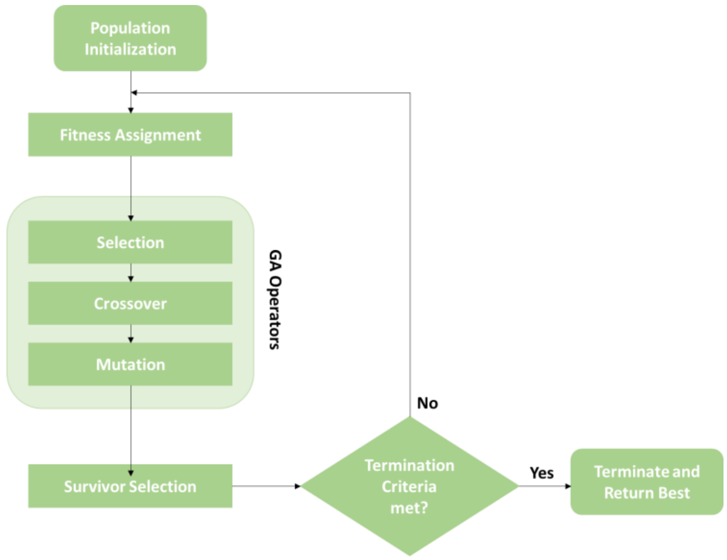


Figure 5. Genetic algorithm flowchart. Adapted from online findings.

3.6.2. Fitness Function

The fitness of a particular individual is considered a measurement of 'ability' of solving the set problem. In optimization, as well as looking for a feasible solution, a searcher is also attempting to find the best solution in the set search space. A good fitness function ensures the GA discovers the best available solution. Regarding portfolio optimization, the fitness function must be a measurement of return and risk.

3.6.3. Generation of a population

"Generating an initial population is a challenge in the implementation of the genetic algorithm." (Etchegaray, 2021). The starting point of this stage is determined by the problem which said individual is attempting to solve and, consequently, the restrictions which come along. It should be noted the generation of an initial population is crucial, as it can establish which areas of search space the algorithm will cover. A common practice for this step is to generate arbitrary arrays which meet all the constraints inflicted by the initial problem.

3.6.4. Selection

During this step, the GA selects the solutions it wishes to crossover, commonly involving the selection of parents considering the probability based on fitness. Nevertheless, diverse approaches can be used, such as the tournament or roulette wheel selections, for example. Selection will ultimately eliminate individuals classified as low-quality, whilst preserving its diversity. The new selection of individuals is made as seen on the following formula equation, where the reproduction probability for each individual is calculated.

$$p_i = \frac{f_i}{\sum_{i=1}^{n} f_i} \tag{2}$$

Regarding the above equation, the following can be stated: f_i is the fitness of the individual i (a fitness function is needed to evaluate the quality of each candidate solution with regard to the task to be performed, as previously mentioned). *n* is the size of the population under analysis. Each time a single chromosome is selected for the new population. This is achieved by generating a random number r from the interval [0, 1]. In case r $\prec p_i$, the first chromosome will be selected, otherwise the i^{th} chromosome such as $p_{i-1} \prec$ r $\leq p_i$ will be (Sefiane & Benbouziane, 2012).

3.6.5. Crossover

The crossover is a function known for bringing two individuals together and combining those into two different individuals, resembling the originals. The main logic behind this concept is the following: in case two distinct individuals are good solutions to a problem, the possible combination of those may discover an even better solution. Many approaches exist, as well as different algorithms for combining individuals. The used crossover is reliant on the constraints of the encoding. It should be added "(…) different crossovers have a different effect on each generation." (Etchegaray, 2021).

3.6.6. Mutation

Lastly, the mutation stage is applied to an individual. It generates small modifications to the individual, in order to generate a new one. As mutations are used to create diversity in a certain population, it is, in addition, a complement of the crossover operator. Mutation provides fine tuning (in which parameters of a given model are adjusted in a precise way to fit with certain observations), thus adding new elements to the current genes in the entire population (Sefiane & Benbouziane, 2012).

The following diagram, contained in Figure 6, can graphically enlighten the phases mentioned throughout the past few paragraphs.

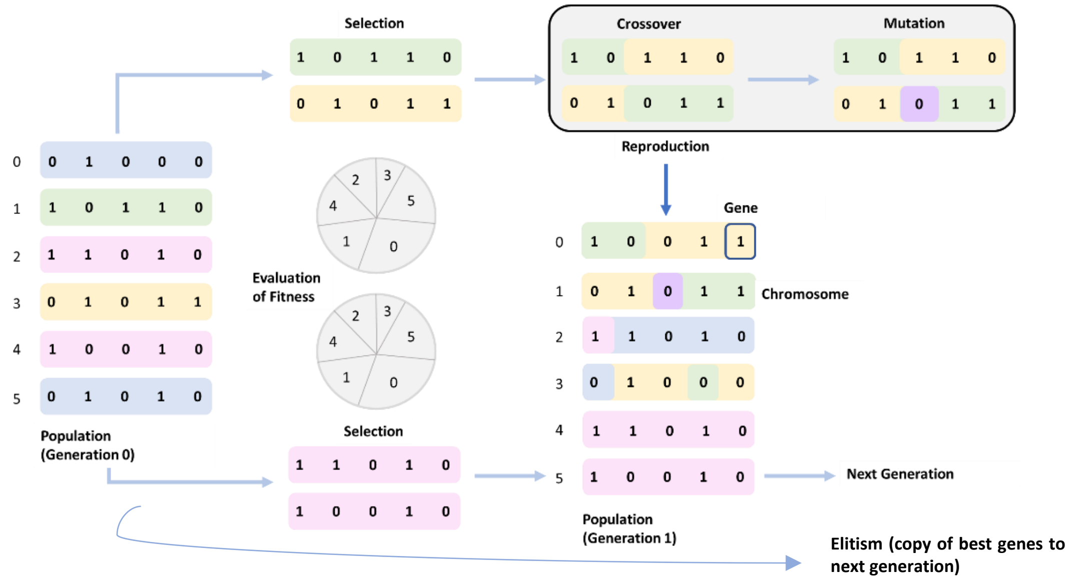


Figure 6. Genetic selection method using genetic algorithm. Adapted from 'Vision-Based Structural FE Model Updating Using Genetic Algorithm' by Park, G.; Hong, K.-N.; Yoon, H in 2021.

3.6.7. Evaluation

In a GA application, evaluation is performed by means of the fitness function which depends on the specific problem and the optimization objective of the GA (Petridis et al, 1998). In achieving the aim of this work, the objective function (fitness function) is modelled to find the solution that scores less on the fitness scale, hence in this application, crossover procedure with the least objective function should lead to better solution. The aim is to choose weights of the portfolio invested in each asset to maximum return and minimum of risk.

4. METHODOLOGY

4.1. DATA

The index obtained hopes to exclude companies engaged in animal exploitation, defense, human rights abuses, fossil fuels extraction and energy production, among other environmentally damaging activities.

For this research, as mentioned, 115 companies were identified as good candidates to be included in the optimization process. However, because of the time-consuming element of the GA implementation, it was decided to only consider 61 assets of the above-mentioned universe. Note the sample size is entirely arbitrary. Therefore, under research were 1) Vegan; and 2) Top-50 ESG ranked companies, considered as of August 2021. Some examples of organizations which could be of use for this research for 1), from seven distinct industries are the following. This information was collected with the help of web searches, in combination with websites such as People for the Ethical Treatment of Animals (PETA).

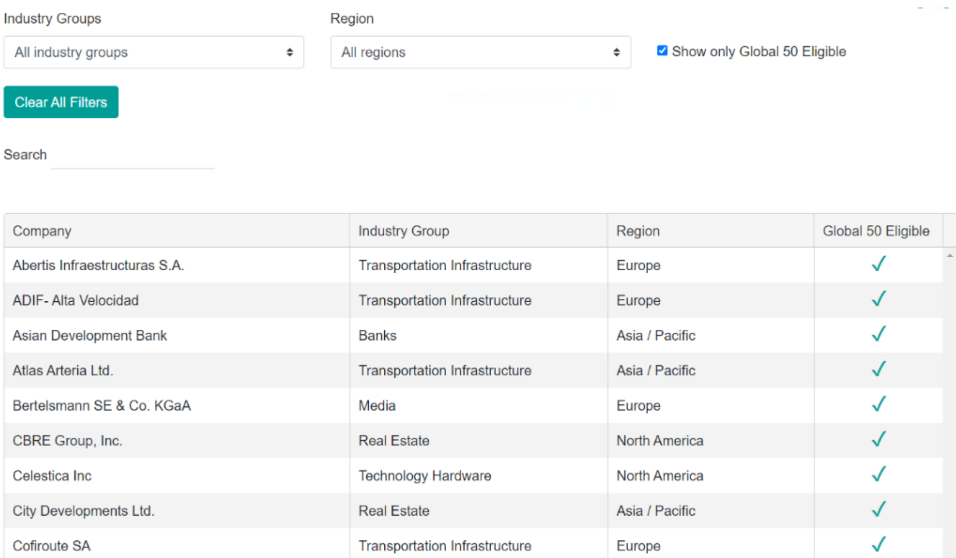
1. Food and Beverage: Beyond Meat; Tattooed Chef; The Very Good Food Company Inc.; Else Nutrition; Maple Leaf Foods; Burcon NutraScience; Laird Superfood; Forum II Merger Corporation; Eat Beyond; Ingredion Incorporated; AAK; Modern Meat; Total Produce PLC; Archer Daniels Midland; Simris Alg; Mission Produce; Natural Order Acquisiton Units; Plant & Co. Brands Ltd; MeaTech; Plantx Life Inc; Oatly; Sol Cuisine; SunOpta; Conagra Brands; B&G Foods.
2. Beauty: Honest Company Inc.; E.L.F. Cosmetics.
3. Packaging & Supplies: Good Natured Products; Bunge Limited.
4. Transportation: Tesla.
5. Wind Power: General Electric; NextEra Energy Partners; Siemens Gamesa Renewable Energy; Vestas Wind Systems.
6. Solar Energy: Enphase Energy; Sunrun; Invesco Solar.
7. Alternative Proteins: Agronomics Limited.

Relating to companies which are part of 2), it was decided to only evaluate a sample out of the top 50 ESG-ranked companies, as this project focuses on top performers under an ESG analysis. These firms can be found on the 'Sustainalytics' website, showing companies from all types of industries and all regions of the world (Figure 7).

It should be addressed the mentioned 'Sustainalytics' website is widely known for rating the sustainability of listed companies built around their ESG performance. According to its official website,

"(…) Risk Ratings measure a company's exposure to industry-specific material ESG risks and how well a company is managing those risks. (…) combines the concepts of management and exposure to arrive at an absolute assessment of ESG risk. We identify five categories of ESG risk severity that could impact a company's enterprise value." (Sustainalytics, n.d.). The mentioned categories can vary from 'Negligible', scoring from 0 to 10, to 'Severe', where the score attained is 40+.

It should be clarified not one of the Vegan companies mentioned previously are included in the Top-50 ESG ranking. Hence, no data will be doubled.



Company	Industry Group	Region	Global 50 Eligible
Abertis Infraestructuras S.A.	Transportation Infrastructure	Europe	✓
ADIF- Alta Velocidad	Transportation Infrastructure	Europe	✓
Asian Development Bank	Banks	Asia / Pacific	✓
Atlas Arteria Ltd.	Transportation Infrastructure	Asia / Pacific	✓
Bertelsmann SE & Co. KGaA	Media	Europe	✓
CBRE Group, Inc.	Real Estate	North America	✓
Celestica Inc	Technology Hardware	North America	✓
City Developments Ltd.	Real Estate	Asia / Pacific	✓
Cofiroute SA	Transportation Infrastructure	Europe	✓

Figure 7. Application of filters on the Sustainalytics website: 'Industry Groups', 'Region' and 'Show only Global 50 Eligible'. Source: https://www.sustainalytics.com/corporate-solutions/sustainability-solutions/top-rated-companies.

4.2. Research Design

The goals proposed required 5 years' worth of historical daily stock market data (sample) and the corresponding date and adjusted close price of specific firms within a population. Yahoo Finance, an online site, was one of the most used websites throughout the development of this thesis, to attain financial information regarding each company under analysis, as suggested in Figure 8.

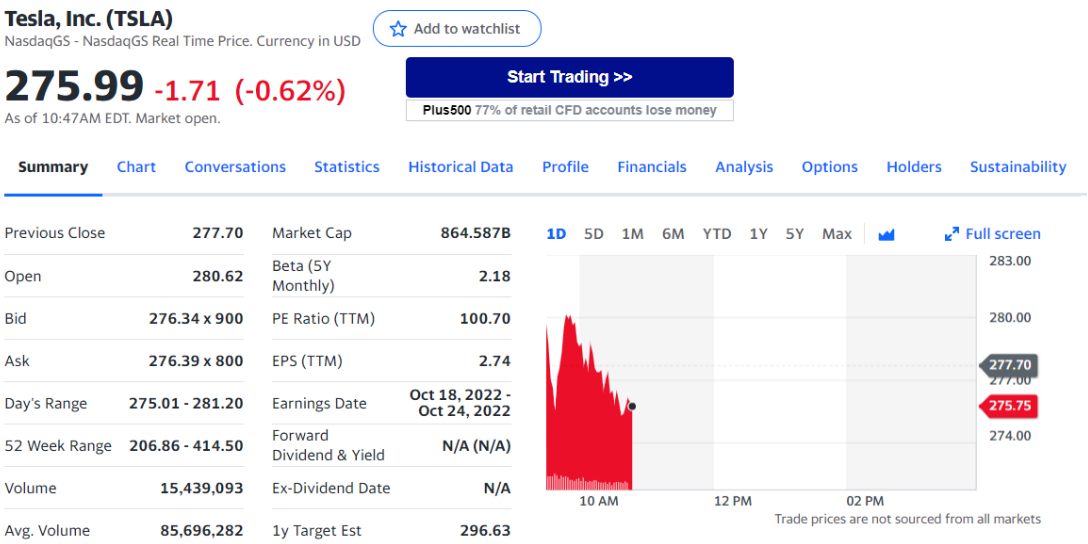
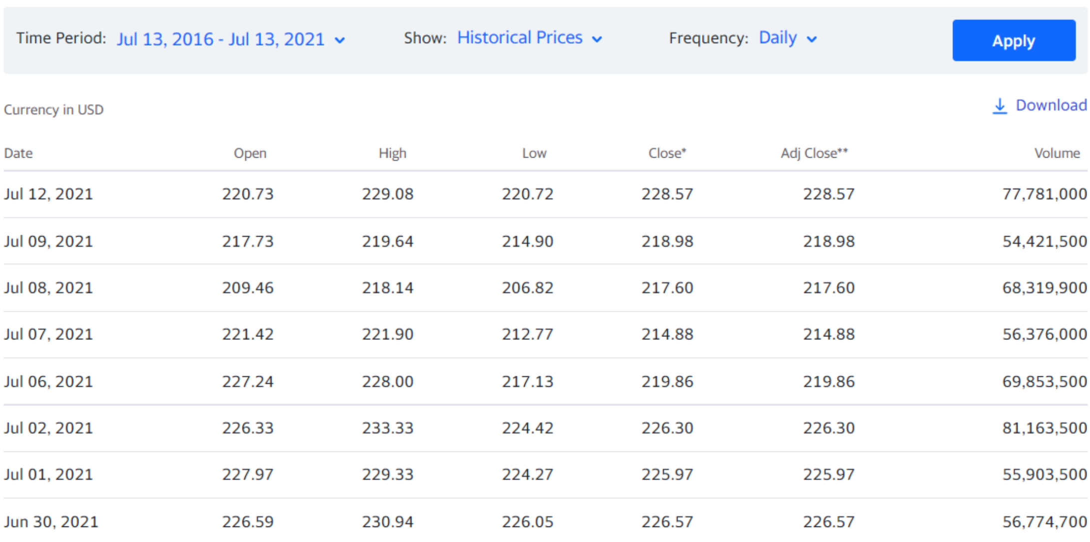


Figure 8. Financial information regarding Tesla, Inc. (one of the companies under study). Information obtained through finance.yahoo.com.

Certain columns of each dataset (Open, High, Low, Close, Volume), as indicated below (Figure 9), were eliminated for the development of the project. There was the intent to overlap the Date (DD-MM-YYYY) of the stocks, with the purpose the final dataset contained information beginning in the same period. This action took place providing every firm went public on distinct dates throughout the years. RStudio and Microsoft Excel were the two software's in use throughout the project.



Time Period: Jul 13, 2016 - Jul 13, 2021 Show: Historical Prices Frequency: Daily Apply

Currency in USD Download

Date	Open	High	Low	Close*	Adj Close**	Volume
Jul 12, 2021	220.73	229.08	220.72	228.57	228.57	77,781,000
Jul 09, 2021	217.73	219.64	214.90	218.98	218.98	54,421,500
Jul 08, 2021	209.46	218.14	206.82	217.60	217.60	68,319,900
Jul 07, 2021	221.42	221.90	212.77	214.88	214.88	56,376,000
Jul 06, 2021	227.24	228.00	217.13	219.86	219.86	69,853,500
Jul 02, 2021	226.33	233.33	224.42	226.30	226.30	81,163,500
Jul 01, 2021	227.97	229.33	224.27	225.97	225.97	55,903,500
Jun 30, 2021	226.59	230.94	226.05	226.57	226.57	56,774,700

Figure 9. Historical Data regarding Tesla. Information to be input for the later GA output. Source: finance.vahoo.com.

The project applied a conclusive research design, as it is expected to help in inferencing. Within this broad design, it was chosen to employ descriptive research, with a multiple-cross sectional design. Quantitative methods of data cleaning and data analysis were implemented. It should be noted that, when it came to the non-existence of an adjusted close value, the decision made was to replace such emptiness with the average of said stock.

This type of research design used statistical tests, advanced analytical techniques, as well as a large sample size.

4.3. Methods for Data Collection instruments and Analysis

The necessary data was collected through documents and records, namely datasets in the csv. format on Yahoo Finance. Once the data was collected, through the download of csv files, and further cleaned and treated, such as the fusion into the same currency (the use of exchange rates of Banco de Portugal was crucial) the beginning of the algorithm implementation took place, followed by a reoptimization using a Heuristic Local Search (HLS), achieved through RStudio. As it was decided to gather 5 years' worth of historical data, the companies which did not hold such data were removed from the project input, as 5 years is believed to be of great length, given that the more information available, the better the algorithm will perform.

With the intention of using a GA to support the optimization of the portfolio, there will be a selection of stocks, from those collected, further indicating the percentage of the overall investment to be divided between each stock.

4.3.1. Portfolio structure

As the selection of *N* financial assets has been gathered (either stock, bonds, funds, ETF, among others), the composition of an optimized portfolio can be achieved. For the specific case to be introduced, only stocks were selected as financial assets to be included in the portfolio. It should be noted this methodology was adapted from that introduced by Malato (2018).

Each one of the selected assets have several historical returns, which are the price relative difference from one period to another. It should be noticed a period is defined in accordance with the needs of each analyst (days, weeks, months, among others). For the specific case below, the calculated return uses daily information, as previously stated.

The return of the ith asset between period *t* and period *t-1* is defined as the following equation shows:

$$p_i(t) = \frac{Price_i(t) - Price_i(t-1)}{Price_i(t-1)} \tag{3}$$

When an investment portfolio is to be built, there is the intention to merge several assets together allocating a fraction (weight) $\mathcal{X}$ of capital to each.

The portfolio return at time *t* is then shown as indicated below:

$$P(t) = \sum_{i=1}^{N} x_i \cdot p_i(t) \tag{4}$$

The main aim of portfolio optimization is to discover the values of the weights which maximize certain performance metrics according to the weight constraints, which are the following:

$$\sum_{i=1}^{N} x_i = 1 \qquad 0 \leq x_i \leq 1 \tag{5}$$

The stated problem can, therefore, be considered of constrained optimization nature.

4.3.1.1. Objective function

In the famous paper written by Markowitz, previously cited, a comprehensive theory regarding portfolio composition was defended. Further studies have since discovered a valuable objective function for portfolio optimization, the Sharpe Ratio, previously introduced, which can be defined as the following equation suggests:

$$Portfolio\ Sharpe\ Ratio = \frac{E[P]}{\sqrt{Var[P]}} \tag{6}$$

As $E[P]$ is the expected return of the portfolio, *Var*[*P*] is the variance related to the portfolio. So as to simplify the following problem, the risk-free return which is incorporated in the original Sharpe ratio formula will not be included for the example to be presented.

The Sharpe ratio can sustain many interpretations. However, it will be analysed, for this context, as a stability metric. If high, its value should mean the average return is higher than the variance, therefore indicating the returns can be quite stable around the mean value. It should be noted the end goal as an investor is, first and foremost, to reach the highest stability of portfolio, not the highest profit.

It should be noted several other objective functions could be used for this specific optimization problem. The Sharpe Ratio will be the chosen one due to its simple nature.

4.3.1.2. Constraints

Constraints are one of the most difficult components of the problem as these usually make it even more difficult to reach a final solution.

In portfolio optimization, the basic constraints are as shown below (7):

$$\sum_{i=1}^{N} x_i = 1 \qquad\qquad 0 \leq x_i \leq 1 \tag{7}$$

The weights (x) must be positive given this specific problem (no possibility of going short) and less than 1. The sum must be equal to 1, in order to cover the entire capital to be invested.

Constrained optimization is an analytical challenge which can be fought in several ways. The introduced method is that of the penalty function.

The penalty function method

Given the fact constraints are great issues of a constrained problem, it is useful to convert this problem to an unconstrained optimization problem.

As a result of the prospects of optimizing a generic *f(x)* function subjected to some constraints, and to make the constraints disappear explicitly, a new function can be created, which penalizes all the points in the function domain which do not satisfy the constraints.

As the present problem is of optimization nature, a positive penalty was added to the original *f(x)* for all the points outside the constraints. Consequently, the global minimum could only be found through the satisfaction of the constraints.

Constructing the penalty function

When the constraint is an inequality in the *g(x)* < 0 form, a penalty function can be built in the form max (0, *g(x)*). Consequently, in case *g(x)* is negative, the max function returns 0; otherwise, it returns the value of *g(x)* itself, increasing the value of the penalty function, preventing the optimization of such function. The greater the value of the penalty function, the further the global minimum is.

Conducive to maintain the function differentiable, it is of use to square it. Thereupon, the inequality constraints provide the following penalty functions:

$$x_i \leq 1 \rightarrow [\max(0, x_i - 1)]^2 \tag{8}$$

$$x_i \geq 0 \rightarrow [\max(0, -x_i)]^2$$

Regarding the sum constraint, as it is an equality constraint, it is simple to describe as a penalty function.

In case there is a generic constraint in the *f(x)-d*=0 form, a quadratic form can be built as the following equation suggests. It is of crucial importance to underline this function holds a very noticeable global minimum when *f(x)=d*.

$$f(x) - d = 0 \rightarrow [\mathrm{f(x)} - \mathrm{d}]^2 \tag{9}$$

Given the above stated, the equality constraint becomes as shown below:

$$\sum_{i=1}^{N} x_i = 1 \rightarrow \left(\sum_{i=1}^{N} x_i - 1\right)^2 \tag{10}$$

In regard to the final function, there is an intention to emphasize the constraints in the optimization procedure. As the undergoing transformation is changing a constrained problem into an unconstrained one, it is valuable to impose the global minimum search to multiply the penalty function by a factor. This is usually multiplied by 10, however, in order to enhance the optimization process, it will be multiplied by 100.

To conclude, the function to be optimized is the Sharpe ratio, however with a minus sign. There is the desire to maximize the Sharpe ratio, nevertheless the penalty function formalism seen so far works only for functions to be minimized. The Sharpe ratio can therefore be multiplied by -1, to transform a maximization problem in a minimization problem.

The final function to be minimized is hence, as suggested below:

$$-Sharpe(\{x_i\}) + 100\left[\left(\sum_{i=1}^{N} x_i - 1\right)^2 + \sum_{i=1}^{N}\left(max(0, x_i - 1)\right)^2 + \sum_{i=1}^{N}(\max(0, -x_i))^2\right] \tag{11}$$

Final remarks regarding GA

GA follow a natural selection law, according to which only the best individuals survive to evolution, as seen above in greater detail. Genetic algorithms are useful tools given the arbitrary portion of their process makes them achieve effective results, even with non-continuous or non-differentiable functions. These algorithms can escape from local minima to the global minimum, which differentiates them from deterministic methods, being the Gradient Descent an example, as Malato (2018) expresses.

5. EXPERIMENTAL RESULTS AND DISCUSSION

The results presented in this section were obtained using the described methodology in section 4, as proposed by several authors, one of those being Malato (2018).

As previously mentioned, all the data was converted into a single structure, carefully joined by the data belonging to each date. As portfolio historical returns are a function of the weights, it should be noted historical data does not vary. The final goal is to, therefore, find an optimal portfolio changing the weights values, by maximizing the Sharpe ratio, in accordance with the constraints, using the data (composed of 61 assets, as mentioned above).

	Adj Close	0HZD_Wendel_Euros.csv	3IC.BE_Icade_SA_Berlin_Euros.csv	59S.F_ScentreGroup_Frankfurt_Euros.csv	3IC.SG_IcadeSA_Stuttgart_Euros.csv	3IC.DU_IcadeSA_Dusseldorf_Euros.csv
407	13/07/2016	88.08490	46.96212	2.211836	47.69978	47.04949
442	14/07/2016	88.49790	47.74169	2.216873	48.17170	47.84474
482	15/07/2016	88.15892	47.61177	2.188540	47.26916	47.72906
623	19/07/2016	87.84181	47.45297	2.180355	47.85274	47.50495
659	20/07/2016	88.65794	47.55402	2.221280	48.71539	47.75075
692	21/07/2016	89.21696	48.34082	2.213095	48.57041	48.56045
731	22/07/2016	89.52250	47.98713	2.200503	49.48383	48.09777
829	25/07/2016	89.82385	48.96880	2.209317	49.49108	49.05206
863	26/07/2016	89.09252	49.15648	2.259687	49.65781	49.20387
897	27/07/2016	90.39140	49.49574	2.233873	49.80280	49.63765
936	28/07/2016	88.81561	49.45964	2.247094	49.41351	49.63765
966	29/07/2016	90.05430	49.48852	2.225058	49.68681	49.56535
44	02/08/2016	88.49793	49.38025	2.218132	49.35262	49.46414
68	03/08/2016	88.33305	48.82444	2.191688	49.08512	48.90024
92	04/08/2016	88.89442	48.61511	2.224428	49.14818	48.74842
124	05/08/2016	90.33450	49.14927	2.216243	49.88253	49.22556
237	08/08/2016	90.29727	49.49574	2.223169	49.54907	49.68103

Figure 10. Data converted into a single structure by the Adjusted Closing date. Source: own research.

5.1. Portfolio Returns

Portfolio returns time series can be defined in function of the weights by the multiplication of each asset return by the weight related to it. By the end of the introduced analysis, the sum of all the weighted time series can be done element by element.

The portfolio returns function was introduced, given its end goal, hence being a function of an array x of weights.

	0HZD_Wendel_Euros.csv	ECN.F_EUR_FrankfurtSE.csv	MJB.SG_MirvacGroup_Stuttgart_EUR.csv	PES.F_Pearsonplc_FrankfurtSE_EUR.csv	REN.AS_RELXPLC_Amsterdam_EUR.csv	SII.F_Wheaton_Precious_Metals_
1	0.004688681	0.052960289	0.025787740	-0.008084557	-0.010012425	
2	-0.003830396	-0.038737967	0.002793494	0.000260228	-0.000631830	
3	-0.003597004	0.020799682	0.003480897	0.001907080	-0.003162666	
4	0.009290837	-0.038522856	0.024983449	0.003979837	0.010786788	
5	0.006305369	0.107285025	-0.011509424	-0.001809521	-0.000313841	
6	0.003424708	0.000299036	-0.002739919	0.003971041	0.001569842	
7	0.003366238	-0.011061338	0.013736253	-0.000601843	0.003761801	
8	-0.008141835	0.024788439	-0.004743124	0.018155286	-0.004372463	
9	0.014579013	0.005014624	-0.014295242	-0.007267936	-0.003450374	
10	-0.017433052	0.002054542	0.008978292	-0.006384632	0.028328526	
11	0.013946795	0.012595206	-0.005476088	-0.101524971	-0.009794829	
12	-0.017282529	0.012438540	0.004129189	0.005530905	-0.008964384	
13	-0.001863072	-0.026285690	-0.023302845	0.014034907	-0.011852777	
14	0.006355073	0.001173680	0.022455717	0.006078988	0.004898714	
15	0.016199881	-0.004982614	-0.005491123	-0.014779969	0.008546954	
16	-0.000412068	0.045361004	0.008972097	0.001226648	-0.009102218	
17	0.006094149	-0.048182321	-0.023255996	-0.004334956	0.014254164	

Figure 11. Asset returns having deleted date column. Source: own research.

5.2. Objective function with penalty

First, the calculation of the Sharpe ratio was performed on the historical weighted portfolio returns. After that, the penalty function was written, followed by the creation of a constraint function which implements all the constraints. At last, the objective function to be optimized was also developed.

```r
sharpe = function(x) {
  port.returns = portfolio_returns(x)

  return (mean(port.returns)/sqrt(var(port.returns)))

}

constraint = function(x) {
  boundary_constr = (sum(x)-1)**2   # "sum x = 1" constraint 1

  for (i in 1:length(x)) {
    boundary_constr = boundary_constr +
      max(c(0,x[i]-1))**2 +  # "x <= 1" constraint 2
      max(c(0,-x[i]))**2     # "x >= 0" constraint 3
  }

  return (boundary_constr)
}

obj = function(x) {
  # End goal is to maximize Sharpe ratio, so it was multiplied by
  # -1 to fit an optimization problem

  return (-sharpe(x)+100*constraint(x))
}
```

Figure 12. Sharpe ratio, constraint, and optimization functions. Source: own research.

5.3. Optimization Process

As for the optimization procedure itself, given it was developed with the support of RStudio software, its general-purpose library entitled ‘GA’ was of crucial use for the following optimization problem. Even though the GA function considered is simple, it is extremely effective in order to solve such maximization problem.

The configuration of the optimization was set to perform 50000 iterations, only stopping in case the maximum fitness did not change for 1500 exact iterations, as these are the stopping criterion set for the GA. Below the remaining arguments are drawn, as conceived by Luca Scrucca.

```
GA settings:
Type                  =  real-valued
Population size       =  50
Number of generations =  50000
Elitism               =  2
Crossover probability =  0.8
Mutation probability  =  0.1
Search domain =
      x1 x2 x3 x4 x5 x6 x7 x8 x9 x10  ...  x60 x61
lower  0  0  0  0  0  0  0  0  0   0         0   0
upper  1  1  1  1  1  1  1  1  1   1         1   1

GA results:
Iterations             = 22412
Fitness function value = 0.07524287
Solution =
             x1         x2         x3         x4          x5         x6         x7         x8
[1,] 0.01211107 0.02325322 0.01881623 0.03333144 0.004233673 0.01229361 0.01030128 0.01588906
> |
```

Figure 13. Genetic Algorithm parameters. Source: own research.

According to Gotshall S. and Rylander B. (n.d.) “(…) for arbitrarily large population sizes the accuracy of the genetic algorithm approaches, but does not reach, 100%. The greater the population size the greater the chance that the initial state of the population will contain a chromosome representing the optimal solution.”. However, given the need of efficiency from this process, the population size cannot be too large as it causes the GA to slow down. Hence, the population size was accepted to be 50. As for elitism, this value was of 2, representing “(…) the number of best fitness individuals to survive at each generation. (...)” (Scrucca, Package ‘GA’, 2021). Finally, in order to set off the GA, the crossover (the chance that two chromosomes exchange some of their parts) and mutation probabilities were of 0.8 and 0.1, respectively. These values mean 80% of the offspring are made by crossover, while 10% of these will be mutated in one generation. It should be underlined both crossover and mutation probabilities must be in the [0, 1] range. Note the purpose of mutation is to prevent the GA from converging to local optima. Nevertheless, this phenomenon occurs very often, (Scrucca, A quick tour of GA, 2021).

5.3.1. Results of the optimization process

The results of the optimization process were stored in a vector. This step made it possible to associate the values to the names of the assets in question.

```
GA | iter = 22392 | Mean = -5.00298940 | Best =  0.07524287
GA | iter = 22393 | Mean = -6.78520748 | Best =  0.07524287
GA | iter = 22394 | Mean = -6.72139974 | Best =  0.07524287
GA | iter = 22395 | Mean = -11.00858371 | Best =   0.07524287
GA | iter = 22396 | Mean = -7.13703553 | Best =  0.07524287
GA | iter = 22397 | Mean = -8.15657341 | Best =  0.07524287
GA | iter = 22398 | Mean = -4.52749136 | Best =  0.07524287
GA | iter = 22399 | Mean = -6.97617914 | Best =  0.07524287
GA | iter = 22400 | Mean = -8.50156772 | Best =  0.07524287
GA | iter = 22401 | Mean = -5.52275669 | Best =  0.07524287
GA | iter = 22402 | Mean = -8.30161535 | Best =  0.07524287
GA | iter = 22403 | Mean = -16.49883341 | Best =   0.07524287
GA | iter = 22404 | Mean = -5.37299687 | Best =  0.07524287
GA | iter = 22405 | Mean = -6.95381157 | Best =  0.07524287
GA | iter = 22406 | Mean = -7.95184209 | Best =  0.07524287
GA | iter = 22407 | Mean = -5.50536313 | Best =  0.07524287
GA | iter = 22408 | Mean = -4.87172689 | Best =  0.07524287
GA | iter = 22409 | Mean = -4.71466808 | Best =  0.07524287
GA | iter = 22410 | Mean = -5.63101952 | Best =  0.07524287
GA | iter = 22411 | Mean = -8.47681421 | Best =  0.07524287
GA | iter = 22412 | Mean = -5.00207229 | Best =  0.07524287
```

Figure 14. Output of the optimization process. Source: own research.

As witnessed above, the optimization process took 22412 iterations (above in red) to optimize the function.

The 'mean' values show the average fitness value of the current population, while the 'Best' values, to the right, indicate the fitness value of the current best individual. As the direction of the evolution is characterized by the fitness function, i.e., the optimization objective, which is the Sharpe ratio of the spread portfolio, allowing measurement of the relationship between expected return and risk. A 'Best' value of 0.07524287, being >0, underlines the return of the investment is greater than its risk-free rate. As the higher the Sharpe ratio, the better, in the present case the risk the portfolio encounters is not being offset well enough by its return. The low value observed indicates a poorer performance of the ratio if there is significant skewness in the returns of the assets class. This implicates the ratio may not be accurately describing "risk-adjusted returns" as the distribution of returns must be normal considering the mentioned ratio.

According to the weights, the allocation should be done as follows, for the first 18 assets (29.5% of the overall dimension) of the portfolio:

```
                                                                 sol
 [1,] "Return_0HZD_Wendel_Euros"                                 "0.0121110677719116"
 [2,] "Return_3IC.BE_Icade_SA_Berlin_Euros"                      "0.0232532173395157"
 [3,] "Return_59S.F_ScentreGroup_Frankfurt_Euros"                "0.0188162326812744"
 [4,] "Return_3IC.SG_IcadeSA_Stuttgart_Euros"                    "0.0333314388990402"
 [5,] "Return_3IC.DU_IcadeSA_Dusseldorf_Euros"                   "0.00423367321491241"
 [6,] "Return_Atlantica_Sustainable_Infrastructure_Frankfurt_Eur" "0.0122936069965363"
 [7,] "Return_ECN.SG_StuttgardSE_EUR"                            "0.0103012770414352"
 [8,] "Return_ECN.F_EUR_FrankfurtSE"                             "0.0158890634775162"
 [9,] "Return_Covivio_Frankfurt_EUR"                             "0.0105347484350204"
[10,] "Return_DNP.F_Frankfurt_EUR"                               "0.00590136647224426"
[11,] "Return_KinnevikAB_FrankfurtSE_EUR"                        "0.0219915360212326"
[12,] "Return_LANDSECURITIESGROUP_PLC_Dusseldorf_EUR"            "0.00524906814098358"
[13,] "Return_LANDSECURITIESGROUP_PLC_Stuttgard_EUR"             "0.00800663232803345"
[14,] "Return_LEG.F_ImmobilienSE_Frankfurt_EUR."                 "0.0102965384721756"
[15,] "Return_LEG_Immobilien_SE_LEG.DEXETRA_EUR"                 "0.0356988161802292"
[16,] "Return_LSU2.BE_LANDSECURITIESGROUP_Berlin_EUR"            "0.0114751756191254"
[17,] "Return_MJB.SG_MirvacGroup_Stuttgart_EUR"                  "0.0235385745763779"
[18,] "Return_PES.F_Pearsonplc_FrankfurtSE_EUR"                  "0.0122621506452 56"
```

Figure 15. Weight allocation stored as a vector for eighteen assets (out of a total of 61). Source: own research.

Figure 15 shows that, for asset 1, which illustrates the allocated weight for WENDEL (0HZD), the percentage to be allocated should be of 1.21%; for asset 2, ICADE SA (3IC.BE), the value should be of around 2.32%, and so on (represented in red, above). The weights for the remaining 43 assets can be seen in the Annex section of this thesis, through Figure A to B.

As seen in Figure 15, each asset certainly does not surpass the 5% mark of weighted asset allocation. This can be seen as a positive occurrence, as asset differentiation is seen as an advantage as previously mentioned in more detail in section 3.5.5. Below (Figure 16), it is possible to identify a scatter-plotted weight allocation of assets.

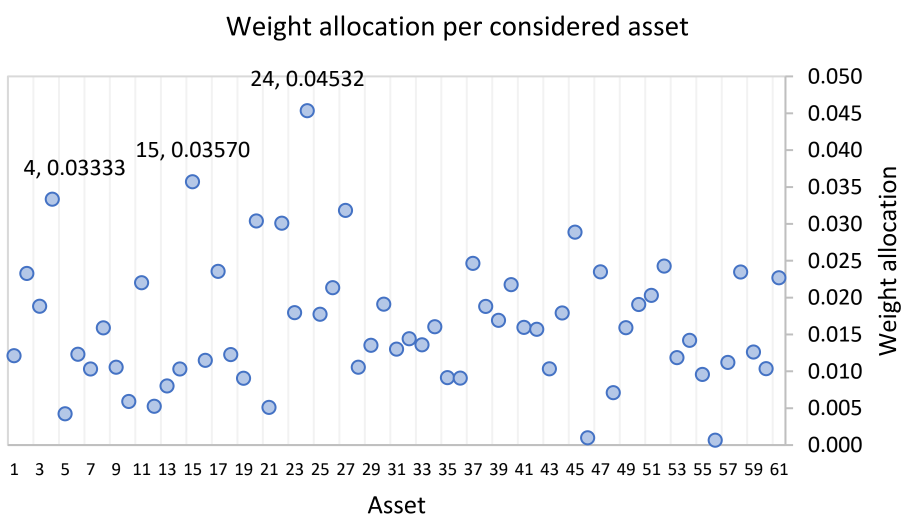


Figure 16. Scatter-plotted weight allocation of assets. Source: own research.

As an example of implementation of the GA output, and allocating an amount of an individual's investment, in the event an investment of 1.000.000 Euros were to be performed, asset '24', 'ThuleGroup_AB' (information contained in Figure A of the Annex), would hold a weight of around 4.532% percent of the pie (45.320 Euros), as can be seen in Figure 16, above. The identified average for the weight allocation is around 1.640%. Additionally, the minimum weight allocated holds a percentage of 0.064%, belonging to asset '56'.

As above seen, each weight is positive and less than 1, as expected. In addition, their sum is of 1.00055. This sum is standard, as numerical methods have a physiological inaccuracy which can be controlled in order to fit the needs of the problem in question.

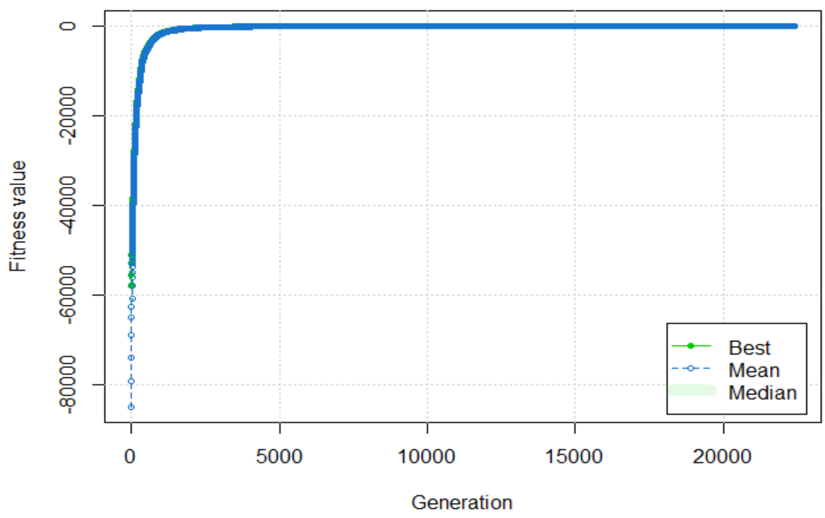


Figure 17. Genetic algorithm plot. Source: own research.

Figure 17 illustrates the behaviour of the fitness value when the number of generations changes, showing that when the number of generations is small, the fitness value is significantly lower (~-80000) according to the examined cases. From the above analysis, it can be established that, when the number of chromosomes per generation is small, it considerably affects the fitness value of the best solution found. However, there is almost no change in the fitness value of the best solution does not change dramatically any further.

As the results above shown confirm the great potential in this approach, the validation of the results was challenging at times because of the problem adopted, as well as the need for the set parameters to suffer a hyperparameter tuning. The decision to adopt the Sharpe ratio as the function to be optimized is considered flawed, given the abovementioned reasonings.

6. CONCLUSIONS AND FUTURE WORK

Genetic algorithms were applied to solve the Green optimal portfolio selection. The method was applied to a sample of 61 assets, regarding vegan and sustainable companies, further obtaining a well-diversified and non-centered asset allocation. The obtained results confirm the possible efficiency of genetic algorithms, given their high-speed convergence towards a better solution. A few functions were presented in the algorithm, for example the penalty function method, to perform portfolio optimization which expects to maximize profits and minimize risks. Some flaws have been identified in regard to the method applied.

The considered method is only one of many possible solutions to solve such an undeveloped topic, taking into account the used data (of free source). In his 2021 thesis, (Etchegaray, 2021) established Multi-Objective Evolutionary Algorithm based on Decomposition (MOEA/D) has "(…) proven to be effective at finding portfolios". Given the success of his work, this EA is considered a viable solution to the problem.

The method involved in this thesis is flexible enough to be used with other objective functions. As previously mentioned, the Sharpe ratio was not fully successful, due to the obtained 'Best' value. As future reference, it should be noted that, when evaluating individual investments, the Sharpe ratio is undeniably more useful. Nevertheless, the Sharpe ratio should be substituted in the future, in this specific example, for the Sortino (portfolio's return subtracting the risk-free rate, dividing that amount by the asset's downside deviation) or the Calmar (average annual rate of return over the past three years divided by the fund's maximum drawdown over that same time period) ratios, in the hopes of a better overall performance.

The possibility of improvement using a deterministic optimization algorithm is to be underlined, following the application of the genetic algorithm, to reach a certain precision, as explained by Malato (2018).

As this research underlines the power of genetic algorithms in portfolio optimization, it also showcases several of its weaknesses. As a GA is commonly sensitive to the initial conditions, the efficiency of such algorithm "(..) relates to the coding of the algorithm, the design of the evolutionary operators and the parameter settings for evolution." (Mosayebi & Sodhi, 2020). As so, the performance of this Green Index is very much related to the tuning of such parameters, which is a very time-consuming process, as well as considered to be a trial-and-error one.

It is of greater importance to underline that the performed optimization was executed having no out-of-sample cross validation, given it is only to be considered of illustrative intent, at this moment. It should not be executed in case those interested wish to perform live trading. For future reference, this should be performed to obtain a usable and real-time functioning optimization of the portfolio.

As future work, it should be of interest to consider developing a Green portfolio, this time taking into consideration several types of financial assets, such as ETF, bonds, and derivatives, among others.

At last, it is beneficial to those interested to consider developing, in addition to a practical analysis, a continual fundamental one, as evolutionary algorithms are not able to hold critical information regarding the evolution of daily markets, as external events can change the course of the investment world.

Genetic algorithms have long provided global solutions to optimization problems, and innovative developments. However, extensions of such GA (such as the quantum-inspired genetic algorithms, or QIGA) are gaining ground. In the future, these are assumed to be refined and explored to improve computational feasibility and accuracy, speeding up information processing tasks, therefore hoping to solve some of the downsides of GA (Ibarrondo, Gatti, & Sanz, 2022).

7. BIBLIOGRAPHY REFERENCES

Adeli, H., & Hung, S. L. (1995). *Machine Learning - Neural Networks, Genetic Algorithms, and Fuzzy Systems.* New York: John Wiley & Sons, Inc.

Adyatama, A. (2019 , December 2). *Optimization with Genetic Algorithm.* Retrieved from RPubs: https://rpubs.com/Argaadya/550805

Alcoforado, J. F. (2016). *Empirical Evidence of the relation between ESG performance and Financial Returns.* Lisbon: Nova School of Business and Economics.

Allaire, L. (2019). *Promoting Socially Responsible Investing To High Net Worth Individuals .* Lisbon: Nova School of Business and Economics .

Almeida, J. F. (2016). *Genetic Algorithms applied to Asset & Liability Management.* Lisboa: Universidade Nova de Lisboa.

Andersen, K., & Kuhn, K. (Directors). (2014). *Cowspiracy* [Motion Picture].

Barclays. (n.d.). *Diversifying your investments - why does it matter?* Retrieved from Barclays: https://www.barclays.co.uk/smart-investor/new-to-investing/reducing-unnecessary-risk/diversifying-your-investments/

Bart. (2012, August 1). *Genetic algorithms: a simple R example*. Retrieved from R-bloggers: https://www.r-bloggers.com/2012/08/genetic-algorithms-a-simple-r-example/

Bashir, L. (2015). *Using Evolutionary Algorithm to Solve Amazing Problem by Two Ways: Coordinates and Locations.* Baghdad: World Scientific News.

Bernarda, M. S. (2021). *Portfolio optimization: from Markowitz to Machine Learning.* Lisboa: Universidade Nova de Lisboa.

Beyond Investing. (n.d.). *Imagine a world with no exploitation of animals or destruction of the environment*. Retrieved from Beyond Investing: https://veganetf.com/

Bowman, J. (2022, January 5). *Investing in Vegan Stocks.* Retrieved from The Motley Fool: https://www.fool.com/investing/stock-market/market-sectors/consumer-staples/vegan-stocks/

Casas, P. (2019, January 15). *Feature Selection using Genetic Algorithms in R*. Retrieved from Data Science Heroes: https://blog.datascienceheroes.com/feature-selection-using-genetic-algorithms-in-r/

CFA Institute. (2022, January 28). *Fundamental Analysis*. Retrieved from https://corporatefinanceinstitute.com/resources/knowledge/trading-investing/fundamental-analysis/

CFA Institute. (n.d.). *ESG Investing and Analysis*. Retrieved from CFA Institute: https://www.cfainstitute.org/en/research/esg-investing

Chan, E. (2009). *Forecasting the S&P 500 index using time series analysis and simulation methods.* Massachusetts: Massachusetts Institute of Technology .

Clapham, E. (2021, August 1). *ParsecFinancial.* Retrieved from Equity Portfolio Construction – The Method To The Madness: https://parsecfinancial.com/equity-portfolio-construction-the-method-to-the-madness/

Confess, L. T. (2020, January 8). *Let The Data Confess*. Retrieved from Feature Selection Using Genetic Algorithm: https://www.letthedataconfess.com/blog/2020/01/08/feature-selection-using-genetic-algorithm-2/

Cudmore, D. (2021, May 20). *Vegan Stocks to Invest In: Public Companies*. Retrieved from VegFAQs: https://vegfaqs.com/vegan-companies-to-invest-in/

Dancho, M., & Vaughan, D. (2022, May 31). Package 'timetk'. *A Tool Kit for Working with Time Series in R*.

Dash, S. (2020). *An Overview of Time Series Forecasting Models Part 1: Classical Time Series Forecasting Models.*

DD. (2018, May 21). *Portfolio Optimization in R*. Retrieved from Coding Finance: https://www.codingfinance.com/post/2018-05-31-portfolio-opt-in-r/

Duarte, A. F. (2016). *Using a Genetic Algorithm with Options Data to Forecast Stocks.* Lisbon: Instituto Superior Técnico.

Duuren, E., Plantinga, A., & Scholtens, B. (2016). ESG Integration and the Investment Management Process: Fundamental Investing Reinvented. *J Bus Ethics*, pp. 525-533.

Elton, E., Gruber, M., & Busse, J. (2002). *Are Investors Rational? Choices Among Index Funds.*

Etchegaray, A. A. (2021). *Portfolio Optimization using Evolutionary Algorithms.* Lisbon: NOVA Information Management School.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 383–417.

Fekri, M., & Barazandeh, B. (2019, February 17). Designing an Optimal Portfolio for Iran's Stock Market with Genetic Algorithm using Neural Network Prediction of Risk and Return Stocks. *Designing an Optimal Portfolio for Iran's Stock Market with Genetic Algorithm using Neural Network Prediction of Risk and Return Stocks*.

Ferreira, M., Matos, P., Starks, L., & Cremers, M. (2015). *Indexing and Active Fund Management: International Evidence.*

Food and Agriculture Organization of the United Nations. (2006). *Livestock's Long Shadow: environmental issues and options.* Rome.

Francesco Corielli, M. M. (2006). Factor based index tracking. *Journal of Banking & Finance*.

Gaivoronski, A. A., Krylov, S., & van der Wijst, N. (2005). Optimal portfolio selection and dynamic benchmark tracking. *European Journal of Operational Research*, 115-131.

Galvão, B. G. (2017). *A multi-population hybrid Genetic Programming System.* Lisboa: Universidade Nova de Lisboa.

Gemst, M. V. (2020). *Forecasting Stock Index Volatility - a comparison of models .* Lisbon: Nova School of Business and Economics .

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization & Machine Learning.* Alabama: Addison-Wesley Publishing Company, Inc.

Gotshall, S., & Rylander, B. (sd). Optimal Population Size and the Genetic Algorithm.

H.Hakansson, N. (1974). Convergence to isoelastic utility and policy in multiperiod portfolio choice. *Journal of Financial Economics*, 201-224.

Hakansson, N. H. (1970). Optimal Investment and Consumption Strategies Under Risk for a Class of Utility Functions. *Econometrica*, 587–607.

Haugen, R. A., & Baker, N. L. (1990). Dedicated stock portfolios. *The Journal of Portfolio Management* , 17-22.

Hiroshi Konno, H. Y. (1991). Mean-Absolute Deviation Portfolio Optimization Model and Its Applications to Tokyo Stock Market. *Management Science*, 519-531.

Hlavaty, T. (2018, September). Portfolio Optimization Methods, Their Application and Evaluation. *Portfolio Optimization Methods, Their Application and Evaluation*.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems.* Ann Arbor: University of Michigan Press.

Hyner, C. (2015). *A Leading Cause of Everything: One Industry That Is Destroying Our Planet and Our Ability to Thrive on It.* Georgetown Environmental Law Review.

Ibarrondo, R., Gatti, G., & Sanz, M. (2022, July 19). Quantum vs classical genetic algorithms: A numerical comparison shows faster convergence. *University of the Basque Country*.

IG. (2020, February 4). *Going vegan: how to invest in veganism*. Retrieved from IG: https://www.ig.com/uk/trading-strategies/going-vegan-how-to-invest-in-veganism-200131#Choose-whether-to-focus-on-stocks-VCTs-or-commodities

J.P.Morgan. (n.d.). *Sustainable investing*. Retrieved from J.P.Morgan: https://am.jpmorgan.com/us/en/asset-management/institutional/investment-strategies/sustainable-investing/

Jeurissen, R., & van den Berg, J. (2005). Index tracking using a hybrid genetic algorithm. *ICSC Congress on Computational Intelligence Methods and Applications*.

Koza, J. R. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection.* MIT: Cambridge.

Kynge, J. (2017). *The ethical investment boom.* Financial Times.

Lai, K. H., Siow, W. J., Kaw, A. A., & Ong, P. (2019, December). *Portfolio optimization using genetic algorithm and harmony search algorithm with varying operators and parameter values.* Retrieved from ResearchGate: https://www.researchgate.net/publication/337796429_Portfolio_optimization_using_genetic_algorithm_and_harmony_search_algorithm_with_varying_operators_and_parameter_values

Leonardo Vanneschi, M. C. (2013). A New Implementation of Geometric Semantic GP and Its Application to Problems in Pharmacokinetics. *EuroGP 2013: Genetic Programming* (pp. 205-216). European Conference on Genetic Programming.

Leonardo Vanneschi, S. S. (2014). Geometric Semantic Genetic Programming for Real Life Applications. In J. H. Rick Riolo, *Genetic Programming Theory and Practice XI* (pp. 191-201). New York: Springer .

Malato, G. (2018, November 17). *Portfolio optimization in R using a Genetic Algorithm*. Retrieved from Medium: https://medium.com/the-trading-scientist/portfolio-optimization-in-r-using-a-genetic-algorithm-8726ec985b6f

March & McLennan Companies. (2014). *Sustainability in the Corporate World Research Report.* Toronto: Mercer LLC.

McMichael, A., Campbell-Lendrum, D., Corvalán, C., Ebi, K., Githeko, A., Scheraga, J., & Woodward, A. (2003). *Climate Change and Human Health.* Geneva: World Health Organization.

Merton, R. C. (1990). The financial system and economic performance. *Journal of Financial Services*, 263-300.

Meticulous Research. (2020, October 15). *Plant Based Food Market worth $74.4 billion by 2027*. Retrieved from Meticulous Research: https://www.meticulousresearch.com/pressrelease/53/plant-based-food-market-2027

Mezali, H. (2013). *Methods for Solving Problems.* London: Brunel University.

Mitchell, T. M. (1997). *Machine Learning.* Pittsburgh: McGraw-Hill.

Mohammed, A. (2019). *Making decisions to save our planet now, for future generations.*

Mosayebi, M., & Sodhi, M. S. (2020, July). Tuning genetic algorithm parameters using design of experiments. *GECCO '20: Genetic and Evolutionary Computation Conference*.

Mossin, J. (1969). Security Pricing and Investment Criteria in Competitive Markets. *The American Economic Review*, 749-746.

MSCI. (n.d.). *ESG Ratings*. Retrieved from MSCI: https://www.msci.com/our-solutions/esg-investing/esg-ratings

Nigel Meade, G. R. (1990). Developing and Maintaining an Equity Index Fund. *The Journal of the Operational Research Society*, 599-607.

Nilgun A. Canakgoz, J. E. (2009). Mixed-integer programming approaches for index tracking and enhanced indexation. *European Journal of Operational Research*, 384-399.

Novicio, T. (2021, April 23). *10 Best Vegan Stocks to Buy Now*. Retrieved from yahoo!finance: https://finance.yahoo.com/news/10-best-vegan-stocks-buy-141155110.html

Nurasiah, I., Nugraha, N., Disman, D., Rozmita, D., & Effendi, K. (2020). Analysis of Economic Growth, Oil Stocks and SIN Stocks in United States. *International Journal of Energy Economics and Policy*, 58-63.

Oh, K., Kim, T., & Min, S. (2005). Using genetic algorithm to support portfolio optimization for index fund management. *Expert Systems with Applications*, 371-379.

Oliveira, A. D. (2016). *Forecasting Stock Markets Using Machine Learning.* Lisbon: Nova Information Management School .

Orito, Y., Inoguchi, M., & Yamamoto, H. (2008, March). Index Fund Optimization Using a Genetic Algorithm and a Heuristic Local Search. *Denki Gakkai Ronbunshi*, pp. 407-415.

Orito, Y., Takeda, M., & Yamamoto, H. (2009). *Index Fund Optimization Using Genetic Algorithm and Scatter Diagram Based on Coefficients of Determination.* Retrieved from Semantic Scholar: https://www.semanticscholar.org/paper/Index-Fund-Optimization-Using-Genetic-Algorithm-and-Orito-Takeda/0cf52af5125b0b1c1f45621b5d434aca96e9119d

Orito, Y., Yamamoto, H., & Yamakazi, G. (2003, 06 01). *Index fund selections with genetic algorithms and heuristic classifications.* Retrieved from Semantic Scholar: https://www.semanticscholar.org/paper/Index-fund-selections-with-genetic-algorithms-and-Orito-Yamamoto/09287d8e45a2f10a6c6e19ee401d01c8610bc826

Packt. (n.d.). *Packt*>. Retrieved from Genetic algorithms structure: https://subscription.packtpub.com/book/big-data-and-business-intelligence/9781788472173/8/ch08lvl1sec60/genetic-algorithms-structure

Palha, C. F. (2022). *Simulation and study of an artificial ecosystem based on a cocoa plantation in Ghana, Africa.* Lisboa: Universidade Nova de Lisboa.

Park, G., Hong, K.-N., & Yoon, H. (2021). *Vision-Based Structural FE Model Updating Using Genetic Algorithm.* MDPI. Retrieved from MDPI.

Perold, A. F. (1984). Large-Scale Portfolio Optimization. *Management Science*, 1143-1160.

PETA. (n.d.). *Cruelty Free*. Retrieved from PETA: https://crueltyfree.peta.org/

Physicians Committee for Responsible Medicine. (n.d.). *A Vegan Diet: Eating for the Environment*. Retrieved from Physicians Committee for Responsible Medicine: https://www.pcrm.org/good-nutrition/vegan-diet-environment

Portugal, B. d. (n.d.). *Taxas de Câmbio*. Retrieved from Banco de Portugal: https://www.bportugal.pt/taxas-cambio

QONTIGO. (n.d.). *QONTIGO Axioma | Dax | Stoxx*. Retrieved from STOXX® Global ESG Leaders: https://www.stoxx.com/index-details?symbol=SXWESGP

Rafi, T. (2021, February 10). *Why Corporate Strategies Should Be Focused On Sustainability.* Retrieved from Forbes: https://www.forbes.com/sites/forbesbusinesscouncil/2021/02/10/why-corporate-strategies-should-be-focused-on-sustainability/?sh=ab5931c7e9f2

Rayer, Q. (2018, November 29). *Why Ethical Investing Matters.* Retrieved from Justice and Peace Scotland: https://www.justiceandpeacescotland.org.uk/Blog/itemid/2651/amid/676/why-ethical-investing-matters

RCM Sustainability White Paper. (2011). *Sustainability: opportunity or opportunity cost? Applying ESG factors to a portfolio does not negatively impact performance and may enhance it.*

Ritchie, H., & Roser, M. (2020). *Emissions by sector*. Retrieved from Our World in Data: https://ourworldindata.org/emissions-by-sector

Roel Jansen, R. v. (2002). Optimal Benchmark Tracking with Small Portfolios. *he Journal of Portfolio Management Winter*, 33-39.

Roger G. Clarke, S. K. (1994). Tracking Errors, Regret, and Tactical Asset Allocation. *The Journal of Portfolio Management*, 16-24.

Roll, R. (1992). Industrial Structure and the Comparative Behavior of International Stock Market Indices. *The Journal of FInance*, 3-41.

Rudd, A., Stefek, D., & Grinold, R. C. (1989). Global factors. *The Journal of Portfolio Management*, 79-88.

Saeed, A. (2018, January). *Using Genetic Algorithm for Optimizing Recurrent Neural Networks*. Retrieved from KD nuggets: https://www.kdnuggets.com/2018/01/genetic-algorithm-optimizing-recurrent-neural-network.html

Scholz, H., & Wilkens, M. (2005). *A Jigsaw Puzzle of Basic Risk-adjusted Performance Measures.* The Journal of Performance Measurement.

Scrucca, L. (2021, October 14). *A quick tour of GA.* Retrieved from cran.r-project: https://cran.r-project.org/web/packages/GA/vignettes/GA.html

Scrucca, L. (2021). *Package 'GA'.* CRAN.

Sefiane, S., & Benbouziane, M. (2012, October 12). Portfolio Selection Using Genetic Algorithm. *Portfolio Selection Using Genetic Algorithm*.

Shiva, V. (2017). It's not an investment if it's destroying the planet. (350.org, Interviewer)

Silva, M. G. (2022). *Bolsa: Investir nos Mercados Financeiros.* Bookout.

Sinha, P., Chandwani, A., & Sinha, T. (2013, July 10). Algorithm of construction of Optimum Portfolio of stocks using Genetic Algorithm. *Algorithm of construction of Optimum Portfolio of stocks using Genetic Algorithm*.

Society, T. V. (n.d.). *Definition of veganism*. Retrieved from The Vegan Society, One world. Many lives. Our choice.: https://www.vegansociety.com/go-vegan/definition-veganism

Sustainalytics. (n.d.). *Top-Rated ESG Companies*. Retrieved from Sustainalytics: https://www.sustainalytics.com/corporate-solutions/sustainability-solutions/top-rated-companies

Tilman, D., & Clark, M. (2014). *Global diets link environmental sustainability and human health.* Nature.

Varsei, M., Shams, N., Fahimnia, B., & Yazdanpanah, A. (2013). A Heuristic Approach to the Index tracking problem: A Case Study of the Tehran Exchange Price Index. *Asian Academy of Management Journal*, pp. 19-344.

Vasiani, V. D., Handari, B. D., & Hertono, G. F. (2017). Stock portfolio optimization using priority index and genetic algorithm . *Stock portfolio optimization using priority index and genetic algorithm* .

Vegpreneur. (n.d.). *Vegan and Sustainable Stocks + Companies you can invest in*. Retrieved from Vegpreneur: https://www.vegpreneur.org/vegan-stocks

Vieira, T. A. (2020). *Forecasting sovereign bonds markets using machine learning: forecasting the portuguese government bond using machine learning approach.* Lisbon: NOVA Information Management School.

Wilson, M. (2003, March/April). *Corporate Sustainability: What is it and where does it come from?* Retrieved from Ivey Business Journal, improving the practice of management: https://iveybusinessjournal.com/publication/corporate-sustainability-what-is-it-and-where-does-it-come-from/

Xia, Y. L. (2000). A Model for Portfolio Selection with Order of Expected Returns. *Computers & Operations Research*, 409-422.

Yu, L., Zhang, S., & Zhou, X. (2006). *A Downside Risk Analysis based on Financial Index Tracking Models.*

Yukiko Orito, H. Y. (2003). Index fund selections with genetic algorithms and heuristic classifications. *Computers & Industrial Engineering*, 97-109.

8. ANNEX

```
[19,] "Return_RDED.F_RELXPLC_FrankfurtSE_EUR"                        "0.00905358791351318"
[20,] "Return_REN.AS_RELXPLC_Amsterdam_EUR"                          "0.0303801000118256"
[21,] "Return_SII.F_Wheaton_Precious_Metals_Corp._Frankfurt_EUR"     "0.00508899986743927"
[22,] "Return_SII.SGWheatonPreciousMetalsCorp._Stuttgart_EUR."       "0.0300786793231964"
[23,] "Return_ThuleGroupAB(publ)_Frankfurt_EUR"                      "0.0179404616355896"
[24,] "Return_TUO.MU_THULE_GROUP_AB_(PUBL)_Munich_EUR"               "0.0453216880559921"
[25,] "Return_VNA.DE_Vonovia_SE_XETRA_EUR"                           "0.0177274942398071"
[26,] "Return_WIS.F_WendelSE_Frankfurt_EUR"                          "0.0213228613138199"
[27,] "Return_OR3W.IL_ThuleGroUP_SEK_to_EUR"                         "0.0318166166543961"
[28,] "Return_1BR1.SG_Unibail_Rodamco_Westfield_Stuttgart_EUR"       "0.0105482339859009"
[29,] "Return_3283.T_Nippon_Prologis_JPY_Yen_EUR"                    "0.0135199129581451"
[30,] "Return_Atlantica_Sustainable_Infrasctruture_NASDAQ_USD_to_EUR" "0.0190826654434204"
[31,] "Return_Atlas_Arteria_OTC_USD_to_EUR"                          "0.0129893571138382"
[32,] "Return_CBRE_US_Dollars_to_EUR"                                "0.014407753944397"
[33,] "Return_CNRAF_Vicinity_Centres_OTC_US_dollars_to_EUR"          "0.0135887712240219"
[34,] "Return_Dai_Nippon_Printing_JPY_to_EUR"                        "0.0160214006900787"
[35,] "Return_Dexus_AUS_dollars_to_EUR"                              "0.00912810862064362"
[36,] "Return_Dexus_OTC_USDollars_to_EUR"                            "0.00906111299991608"
[37,] "Return_DNPCF_OTC_USDollars_to_EU"                             "0.0246240794658661"
[38,] "Return_Electrocomponents_plc_(ECM.L)_GBP_to_EUR"              "0.0187953859567642"
```

Figure A. Weight allocation – Assets 19 to 38.

```
[39,] "Return_Investor_AB_USdollars_to_EUR"                              "0.0169085413217545"
[40,] "Return_InvestorAB_Stockholm_Swedishcoin_to_EUR"                   "0.0217321515083313"
[41,] "Return_Jones _Lang_LaSalle_Incorporated_NYSE_USdollars_to_EUR"    "0.0159432739019394"
[42,] "Return_LAND.L_Land_Securities_GroupPounds_to_EUR"                 "0.0156933963298798"
[43,] "Return_LEGIF_ImmobilienSE_OTC_USdollars_to_EUR"                   "0.0103163123130798"
[44,] "Return_LSGOF_LandSecuritiesGroup_OTC_USdollars_to_EUR"            "0.0179014801979065"
[45,] "Return_MGR.AX_MirvacGroup(MGR.AX)_Australian_dollars_to_EUR"      "0.0288650542497635"
[46,] "Return_MRVGF_MirvacGroup_OTC_USDollars_to_EUR"                    "0.000978812575340271"
[47,] "Return_Pearson_PLC_NYSE_Dollar_to_Eur"                            "0.0234691053628922"
[48,] "Return_PSON.L_Pearson_plc_London_Exchange_GBPound_to_EUR"         "0.00709278881549835"
[49,] "Return_PSORFPearson_plc_OTC_USdollars_to_EUR"                     "0.0158983767032623"
[50,] "Return_RELX_PLC_NYSE_USdollars_EUR"                               "0.0190402865409851"
[51,] "Return_RELX_PLC_LondonSE_GBPound_EUR"                             "0.0202796906232834"
[52,] "Return_RLXXF_PLC_OTC_USDollars_EUR"                               "0.0242619961500168"
[53,] "Return_SCG.AX_ScentreGroup_AustDollars_EUR"                       "0.0118465572595596"
[54,] "Return_STGPF_ScentreGroup_OTC_USdollars_EUR"                      "0.0141841322183609"
[55,] "Return_TCL-A.TO_TranscontinentalInc_CanadianDollars_EUR"          "0.00956393778324127"
[56,] "Return_TCLAF_TranscontinentalInc._USdollars_EUR"                  "0.000640109181404114"
[57,] "Return_THULE.ST_ThuleGroupAB_SKE_EUR."                            "0.0111848413944244"
[58,] "Return_VAKRANGEE.NS_VakrangeeLimited_NSE_INRcoin_EUR"             "0.023464247584343"
[59,] "Return_VCX.AX_VicinityCentres_AUDdollars_EUR"                     "0.0126003623008728"
[60,] "Return_WPM.TO_WheatonPreciousMetalsCorp._CAD_EUR"                 "0.0103398263454437"
[61,] "Return_WPM_WheatonPreciousMetals_Corp._NYSE_USdollars_EUR"        "0.0226633548736572"
```

Figure B. Weight allocation – Assets 39 to 61.