An Evolutionary Computing Approach to Financial Portfolio Management Based on Growth Stocks & Sector/Industry Distribution

André Nicolau Fernandes

Thesis to obtain the Master of Science Degree in Information Systems and Computer Engineering

Supervisors: Prof. Rui Fuentecilla Maia Ferreira Neves
Prof. Nuno Cavaco Gomes Horta

Examination Committee

Chairperson: Prof. Alberto Manuel Rodrigues da Silva
Supervisor: Prof. Rui Fuentecilla Maia Ferreira Neves
Member of the Committee: Prof. Luís Manuel Silveira Russo

May 2016
Resumo

Este trabalho apresenta uma proposta de um sistema com base na Inteligência Artificial, com o objectivo de criar e gerir portfólios financeiros. O sistema incorpora informação financeira com algoritmos evolucionários afim de atingir melhor desempenho na escolha de ações para compor o portfólio. O principal objectivo é construir uma estratégia que capture as ações em crescimento mais promissoras do mercado. Para conseguir atingir este objectivo, o sistema desenhado incorpora a abordagem fundamental e a divisão do mercado por sectores/indústrias, usando rácios financeiros e indicadores específicos dos sectores. Várias estratégias de investimento foram testadas e posteriormente validades, utilizando como medidas de avaliação o retorno obtido, a variância e o risco extra. As simulações foram realizadas com dados do índice S&P 500 do período compreendido entre Janeiro de 2011 e Dezembro de 2014. Restrições reais são usadas para definir um ambiente semelhante ao enfrentado por gestores de portfólios reais. Os resultados demonstraram que combinar computação evolucionária com rácios financeiros é uma boa solução para encontrar as empresas mais promissoras. A divisão do mercado em sectores também foi provado ser uma opção rentável. As simulações obtiveram retornos acima do mercado, apresentando variâncias ligeiramente inferiores. Os melhores resultados sugerem que as ações mais promissoras são as que apresentam maiores returns on equity (ROE), rates of revenue (RR) e price to earnings growth (PEG) mais elevadas, e adicionalmente, baixos debt ratios (DR).
Abstract

In this work we propose an artificial intelligence based system to compose and manage financial portfolios. Financial knowledge and evolutionary algorithms are incorporated in the system to achieve a better performance when choosing the stocks to compose the portfolio. The main goal is to design a strategy which captures the most promising growing companies on the market. To achieve this goal, the system designed incorporate a fundamental approach and a market division by sectors/industries, using financial ratios and sector specific indicators. Several investing strategies were tested, and further validated using as measures the return, variance and extra risk. The simulations were conducted using data from the S&P 500 index for the periods January 2011 until December 2014. Real constraints are used to define an environment similar to the ones faced by real portfolio managers. The results demonstrated that combining evolutionary computation with financial ratios is a good solution to select the most promising companies. The division of the market by sectors also proved to be profitable. Simulations obtained returns above the market with slightly lower variances. The best results suggested that the most promising stocks are the ones presenting higher returns on equity (ROE), higher rate of revenue (RR) and price to earnings growth (PEG), and additionally, low debt ratios (DR).
Acknowledgements

First, I would like to express my gratitude to my supervisors, professor Rui Neves and professor Nuno Horta, for their support, patience, advice, understanding and constant feedback and guidance during the development of this thesis. Specially to my advisor Rui Neves, for the invaluable transmitted knowledge on the financial market domain; for his constant presence and availability, without which it would not be possible to conclude this work.

Finally, my gratitude toward my family, friends and colleagues, who accompanied me on this journey.
to my family and friends,
E Pluribus Unum.
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GICS Hierarchy Diagram</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>Revenue from each sector of the S&amp;P 500 index</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>Revenue from sector Health Care</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>Revenue from sector Energy</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>Example of a Balance Sheet</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>Example of an Income Statement</td>
<td>12</td>
</tr>
<tr>
<td>7</td>
<td>Graph generated using the Markowitz Modern Portfolio Theory. Image obtained</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>from Markowitz's book [1].</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Evolutionary Algorithm General Behaviour</td>
<td>19</td>
</tr>
<tr>
<td>9</td>
<td>System's Layered Architecture</td>
<td>29</td>
</tr>
<tr>
<td>10</td>
<td>System Data Flow</td>
<td>30</td>
</tr>
<tr>
<td>11</td>
<td>System Data Flow</td>
<td>31</td>
</tr>
<tr>
<td>12</td>
<td>Edgar 2.0 Flow Diagram</td>
<td>32</td>
</tr>
<tr>
<td>13</td>
<td>Price Record</td>
<td>33</td>
</tr>
<tr>
<td>14</td>
<td>Directories disposition after the Excel Converter execution</td>
<td>34</td>
</tr>
<tr>
<td>15</td>
<td>Normalize Edgar Flow Diagram</td>
<td>35</td>
</tr>
<tr>
<td>16</td>
<td>Directories disposition after the 1st phase of the Normalized Edgar</td>
<td>36</td>
</tr>
<tr>
<td>17</td>
<td>Directories disposition in the end of the Normalize Edgar process</td>
<td>37</td>
</tr>
<tr>
<td>18</td>
<td>Statements Standardizer class diagram</td>
<td>38</td>
</tr>
<tr>
<td>19</td>
<td>Statement Interpreter Flow Diagram</td>
<td>39</td>
</tr>
<tr>
<td>20</td>
<td>Final tree for data storing</td>
<td>40</td>
</tr>
<tr>
<td>21</td>
<td>Example of financial statement in the final format</td>
<td>41</td>
</tr>
<tr>
<td>22</td>
<td>Fundamental Ratios Module class diagram</td>
<td>43</td>
</tr>
<tr>
<td>23</td>
<td>Sector Object Illustration</td>
<td>44</td>
</tr>
<tr>
<td>24</td>
<td>Company Object Illustration</td>
<td>45</td>
</tr>
<tr>
<td>25</td>
<td>Fundamental Ratios Module data flow diagram</td>
<td>46</td>
</tr>
<tr>
<td>26</td>
<td>Picking Module Data Flow Overview</td>
<td>50</td>
</tr>
<tr>
<td>27</td>
<td>Picking Genetic Algorithm Chromosome</td>
<td>51</td>
</tr>
<tr>
<td>28</td>
<td>Generic crossover example</td>
<td>53</td>
</tr>
<tr>
<td>29</td>
<td>Crossover methodology used in this work</td>
<td>54</td>
</tr>
<tr>
<td>30</td>
<td>Mutation methodology used in this work</td>
<td>55</td>
</tr>
<tr>
<td>31</td>
<td>Minimum jump to entry the portfolio - illustration</td>
<td>57</td>
</tr>
<tr>
<td>32</td>
<td>Company’s Rate Calculation Procedure</td>
<td>59</td>
</tr>
<tr>
<td>33</td>
<td>Portfolio Simulator class diagram</td>
<td>60</td>
</tr>
<tr>
<td>34</td>
<td>Portfolio Simulator flow diagram</td>
<td>61</td>
</tr>
<tr>
<td>35</td>
<td>Different Rankings produced by the <code>Ranking</code> class</td>
<td>64</td>
</tr>
<tr>
<td>36</td>
<td>Strategy A investment example</td>
<td>66</td>
</tr>
<tr>
<td>37</td>
<td>Strategy B investment example</td>
<td>67</td>
</tr>
<tr>
<td>38</td>
<td>Strategy C investment example</td>
<td>69</td>
</tr>
<tr>
<td>39</td>
<td>Graphic for revenue of all sectors</td>
<td>70</td>
</tr>
<tr>
<td>40</td>
<td>Graphic for revenue growth of all sectors</td>
<td>71</td>
</tr>
<tr>
<td>41</td>
<td>Graphic for Earnings Per Share of all sectors</td>
<td>72</td>
</tr>
<tr>
<td>42</td>
<td>Graphic for Earnings Per Share growth of all sectors</td>
<td>73</td>
</tr>
<tr>
<td>43</td>
<td>ROI graphs for two investments</td>
<td>75</td>
</tr>
<tr>
<td>44</td>
<td>Variance graphs for two investments</td>
<td>77</td>
</tr>
<tr>
<td>45</td>
<td>Illustration for the transformation from their current values to gain</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>percentage</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>Simulation environment</td>
<td>81</td>
</tr>
<tr>
<td>47</td>
<td>Genetic Algorithm configuration</td>
<td>81</td>
</tr>
<tr>
<td>48</td>
<td>Results Overview Table</td>
<td>82</td>
</tr>
<tr>
<td>49</td>
<td>Best sector ordering for each combination Version-Strategy</td>
<td>83</td>
</tr>
<tr>
<td>50</td>
<td>Best strategy for each version (left) and best version (right).</td>
<td>84</td>
</tr>
<tr>
<td>51</td>
<td>Case Study I - Validations results comparison</td>
<td>85</td>
</tr>
</tbody>
</table>
List of Acronyms

LSE  London Stock Exchange
NYSE  New York Stock Exchange
USD  United States Dollar
DJI  Down Jones Index
SIC  Standard Industrial Classification
NAICS  North American Industry Classification System
GICS  Global Industry Classification Standard
GARP  Growth At a Reasonable Price
GDP  Gross Domestic Product
SCI  Stanley Capital International
GAAP  Generally Accepted Accounting Principles
BS  Buy and Sell
IPO  Initial Public Offering
ES  Exponential Smoothness
HPM  Hybrid Prediction Module
EBIT  Earnings Before Interest and Taxes
CF  Cash Flow
IS  Income Statement
BS  Balance Sheet
EA  Evolutionary Algorithm
GA  Genetic Algorithm
SGP  Standard Genetic Programming
SOEA  Single Objective Evolutionary Algorithm
MOEA  Multi Objective Evolutionary Algorithm
MOGP  Multi Objective Genetic Programming
STGP  Strongly Typed Genetic Programming
MELG  Multiple-Scenario Evolutionary in the last Generation
MEVO  Multiple-Scenario Evolution
MVCCPO  Mean-Variance Cardinality Portfolio Optimization
GNP-RA  Genetic Network Programming with Rule Accumulation
HPM  Hybrid Prediction Module
RST  Rough Sets
ANN  Artificial Neural Network
GNN  Genetic Neural Network
RNN  Recurrent Neural Network
GFS  Genetic Fuzzy System
CGFS Clustering-Genetic Fuzzy System
ROI  Return on Investment
ROE  Return on Equity
ROA  Return on Assets
ROCE Return on Capital Employed
CR   Current Ratio
QR   Quick Ratio
DR   Debt Ratio
DER  Debt Equity Ratio
EPS  Earnings Per Share
PER  Price to Earnings
PEG  Price to Earnings Growth
PER-F Price to Earnings Future
P/BV Price to Book Value
DY   Dividend Yield Ratio
DPR  Dividend Payout Ratio
NIGR Net Income Growth Rate
RGR  Revenue Growth Rate
RR   Revenue Rate
RSI  Relative Strength Index
ROC  Rate of Change
SMA  Simple Moving Average
HMA  Hull Moving Average
VMA  Variable Moving Average
OBV  On Balance Volume
MACD Moving Average Convergence Divergence
RPD  Relative Percentage Deviation
ROR  Rate of Returns
PVI  Positive Volume Index
GPI  Growth Potential Index

**ARMR**  Autoregressive Moving Reference

**IR**  Information Ratio

**SR**  Sortino Ratio
1 Introduction

Stock markets have been around for centuries. The London Stock Exchange (LSE) and the New York Stock Exchange (NYSE) are examples of large and old stock exchange, been founded in 1773 and 1792 respectively. Although, it was in the actual century that the most remarkable changes begun. Over the last 50-60 years, the world suffered the Third Industrial Revolution, the so called Digital Revolution, marked by an exponential technology evolution in every possible fields. Of course, in this Information and Digital Age, the financial and economic world was not forgotten or neglected. On the contrary, the advances in the computer science were well applied in this subject. With the introduction of the internet and the possibility of worldwide knowledge in real time (globalization), the exchange markets had to change their ways of operating, introducing the remote stock trading. Also, nowadays, stock information is an easy thing to store and analyse. As result of the easy access to trading prices, historical prices and volume of transactions, new investing strategies and philosophies appeared.

Until then, investors based their stock choices on the companies’ fundamentals, like their profit, sales and net income. Though in the current days it still represents the most common strategy, many investors and analysts give preference to the analysis of prices and trends, technique known as Technical Analysis. Technical analysts believe that by looking into historical data related with stocks and the overall market, they will be able to forecast the future at a certain degree, without studying the micro and macro economy. Both philosophies have supporters and criticisms. Motivated by all this, the stock exchange become a popular research field among computer science people. With all the information available, it starts to be difficult to manage and chose portfolios without the computers help. It may be impossible in practice to manage financial portfolios with a large amount of stocks without a computer. Computer scientists thought “what if we could create an automatic system to choose the best stocks and manage financial portfolios?”.

Mainly on the Artificial Intelligence field, specialist did a exhaustive research in order to combine both areas. With the extreme evolution in computation, solving large problems with dozens of variables in real time is a reality, opening the doors to tools such as evolutionary algorithms (EA), fuzzy logic, artificial neural networks (ANN). These are techniques based on the natural biological evolution. As Darwin stated, the fittest survives and seed the next generation. After several generations of natural selection and reproduction, the population converge on an optimal set of individuals.

In this project we created and tested a system that uses an EA, applied to the stock exchange. More concisely, to solve the problem of selecting the best stocks among the thousands available and manage the resulting portfolio, the system applies an evolutionary algorithm, mixing the economic and the artificial intelligence worlds. The algorithm uses several financial ratios, calculated with the company’s fundamentals (Fundamental Analysis). The objective is to select companies exhibiting an enthusiastic growth potential, supported this selections on the industry and sector in which the company is inserted, instead of choosing from the overall market. The main contributions made in this paper are: (1) Combine fundamental analysis of both companies and industries/sectors with Evolutionary Algorithms in order to develop an adaptive approach for investing; (2) investing decisions based on the industry/sector, investing only in the best companies from the best sectors instead of looking to the overall market; (3) explore indicators based on sector analysis, calculating fundamental ratios based on the companies inserted in each sector/industry; (4) explore new financial ratios based on fundamental analysis to determine if a company is overvalued, supported on current fundamental values and future projections of profits.

Resuming, in this document, is described the implementation and testing of a new solution aiming the automatic construction of financial portfolios with high growth companies based on fundamental analysis, evolutionary computing and industry distribution.

1.1 Motivation

As mentioned before, each day technology evolves and new ways to beat the market rise from different branches. Investors start to put their believes on computer science, especially on intelligent systems that are able to manage dozens of stocks at the same time. As a consequence of this demand, the motivation is to develop an innovative system, using unexplored techniques and methodologies, ant that is able to beat the stock market index and previous solutions, providing high returns to the investor.
1.1.1 General Goals

- To develop a critical position regarding the domain of the work, in this case, the financial markets. Gather conditions to clearly understand the problem in question and propose a rational and coherent solution.
- To understand the economic world, more precisely the stock market. What affects it, and how investors can efficiently take advantage from it.
- Better understanding of the possible applications that technology presents in the financial environment, more concretely speaking, the artificial intelligence branch.

1.1.2 Concrete Goals

- Thoroughly comprehend fundamental analysis, why is it still the most common analysis and from which new angle can it be explored.
- Study the artificial techniques called evolutionary computation such as Genetic Algorithms, Neural Networks, Fuzzy Systems, Machine Learning, Rough Sets, and other, understanding how they can be used to construct highly profitable financial portfolios.
- Develop a system using Genetic Algorithms combined with fundamental analysis and market analysis by industry/sector that accurately select profitable companies and entry/exit points.
- Understand the valuation methods for this kind of systems.

1.2 Problem Statement

Design a intelligent system that accurately analyses the financial market and select the companies exhibiting the strongest grows in each financial sector or industry. The solution should be robust, being able to produce similar results on real markets when compared to the ones obtained on the training.

1.3 Methodology

To solve the stated problem, this project consists on the construction of a new computational system which combines financial and artificial intelligence concepts. The developed system primary uses the financial information concerting each industry to rank them accordingly. Next, with a fundamental analysis approach, the companies that present greatest growth in each sector or industry are picked through the execution of a genetic algorithm. Thorough the entire period, the genetic algorithm decides dynamically which stocks compound the financial portfolio and how much money should e invested. To conclude, the system is validated using specifically design metrics that captures the system’s performance.

Three innovations can be considered in this work: (1) the use of fundamental analysis to choose the companies presenting high grow, including two original fundamental ratios created for this work; (2) instead of picking the best stocks from the overall market, pick stocks considering their sectors, distributing the portfolio's slots dynamically; (3) use specific indicators for each sector, deciding whether to buy or sell based on the industry performance and rank.

1.4 Document Structure

This project document is structured as given next:

- Chapter 2 addresses all theory and concepts needed to understand the presented work. Also, an exhaustive review of the already existing solutions is provided, along with relevant conclusions.
- Chapter 3 presents documentation regarding our solution. The architecture, implementation and internal procedures of the system can be observed here.
- Chapter 4 describes the valuation methods that applied to measure the performance and quality of the system, along with the results obtained.
Chapter 5 finishes the document. In this chapter are provided the final conclusions, the system limitations and possible improvements, and a review of the achievements.
2 Concepts and Literature Review

In this section the background information and fundamental concepts that are relevant and necessary for the reader to understand the presented work are addressed. The section is divided in six sub-sections. The first five sub-sections are concerned with theoretical concepts, and the sixth was kept for the analysis of existing solutions and their methodology regarding the stock picking strategies and portfolio management.

2.1 Market Analysis

When it comes to investing and picking the best assets within the market to include in a portfolio, an elaborated market analysis is mandatory if investors want to maximize their returns and mitigate the risk of losing. There are several ways to evaluate the market, depending on the investor philosophy and position, however usually it comes down to two different schools of thought when we are talking about analysing securities. Fundamental Analysis and Technical Analysis.

2.1.1 Fundamental Analysis

Fundamental Analysis of a business or security, involves analysing its financial statements and strength, using historical and present data to derive a stock’s current fair value and forecast future values [2]. Fundamental Analysis tries to combine and analyse all the factors that can affect a stock price in a totally logic and emotionless way [3], looking to the global macroeconomic conditions, the industry and, most important, focusing on a rigorous evaluation of the company’s fundamentals. This analysis requires a good analyst to be useful and trustworthy, since it involves looking for several aspects as revenues, expenses, assets, liabilities, corporative actions and other aspects of a company. The actual price of a security can be calculated through the sum of all future cash flows generated by this security [4]. Breaking the generated cash flows into a dividend return $D_1$ and a capital gain return $R$, an analyst can estimate the intrinsic value of a security with the formula $P_0 = \frac{D_1 + R}{1 + R}$. When fundamental analysts compare this value with the current security’s market price, if the market price is higher, its possible that the stock is overvalued and the investor should consider selling the securities, otherwise the investor believes that the security worth more and its undervalued, so consequently its a potential buy signal.

2.1.2 Qualitative Analysis

Through Qualitative Analysis, a analyst seeks valuing a company not only by the numbers (the traditional financial criteria, Fundamental Analysis), but also looking to the general qualities of a company. Hence, qualitative analysts use as criteria and contribution to their evaluation factors as:

- **Management and internal coordination**: The person behind the company and the way he manages the company is the backbone of any successful company. Who is running the company can be a very relevant factor for the overall performance of a company.

- **Company’s Philosophy**: In what style the company operate? Looking for the company’s historical and past actions is possible to discern the company’s style, and then the analyst must decide if he agrees with this philosophy and trusts it. A company that operates following the same philosophy for twenty years, probably has a successful and profitable philosophy.

- **Business Model**: How does de company make money is a vital part to decide whether to invest or not. If the company’s activities will not be profitable, analysts reconsider the worth of an investment.

- **Market Competition and Market Share**: A qualitative analysis also looks to the characteristics of the industry. A not growing industry is a possible bad industry for investing.

- **Brand Name**: A big and valuable brand name reflects the success of a company. It takes years of development and marketing to achieve a brand name so strong that investors could invest on a company just by its name. As an example, Coca Cola.
Most investors rely their investments on quantitative analysis as fundamental or technical analysis, but supplementing the analysis with qualitative factors can give a fair advantage to the investors, although it might be hard to find this information in some cases.

2.1.3 Technical Analysis

On the other side of the coin, investors can support their investments with technical analysis. Even though fundamental analysis and technical analysis can be applied together, it is not very common to see a analyst converging the two strategies. The reason is simple, technical analysts believe and take as granted that is possible to get the appropriate investment signals about the market future performance only looking through the present data and historical data. Technical analysis follows three key principles [4] [5]: market action accounts for everything, prices move in trends, history repeats itself. This means that technical analysts presuppose the irrelevance of fundamental analysis of either the company and economic situation. Statistics generated by past market activity and looking at the past charts of prices is enough to make inference about the future, for the reason that price evolution follows identifying market trends that repeats over time.

“I believe the future is only the past again, entered through another gate”

Sir Arthur Wing Pinero, 1893

Furthermore, technical analysis study the people involved in the market, the human behaviour, presupposing that the human nature will react the way it did in the past facing similar situations, causing the history recurrence. Besides the criticism about this philosophy, Lo and MacKinlay [6] have shown that is possible to use past prices and charts to forecast future prices and returns to some degree, studying the price evolution of single stocks or the market capitalization as a whole. To analyse the market and decide whether to pick and buy a stock using technical analysis, investors has at their disposal hundreds of indicators and chart patterns. A technical indicator can be expressed by a mathematical formula, normally applied to stock’s prices and volume. Plotting the results and comparing with the price evolution for a given stock, analyst pretend to deduct if the stock is over or undervalued. However, no indicator is infallible! The technical analysis process requires interpretation, being more subjective than formulaic.

2.1.4 Investing Strategies

A Value Investing is the manner and philosophy first introduced by Benjamin Graham and David Dodd, which rests on three key characteristics [7] [8]: the fact that the prices of financial market suffer significant movements, the fact that market prices of financial assets do have a fundamental value, which can be, within a reasonable accuracy, calculated by a disciplined analyst, and last, apply a strategy of buying securities only when their prices are notably below the calculated intrinsic value of the security. Graham referred this gap as “the margin of safety”. That being said, this investment philosophy is based on a strong fundamental analysis. Investor seek for companies with strong fundamentals that are being traded for a low price (undervalued stocks). Following Graham and Dodd master recipe, analysts select some securities for valuation, estimate their intrinsic values, calculate the appropriate margin of safety for each security, decide how much of each security to buy, decide when to sell securities. Investors following this philosophy believe that, in a long term, the market will react (and overreact), correcting the stock’s price, which will lead to a major return of capital to the investors that were able to choose the companies with the biggest margin of safety and potential to increase its share price. Further it is presented in this work the analysis of the financial statements and some economic indicators used on this approach to evaluate a stock price.

B Growth Investing is the other end of the rope. Although growth investors share a few interests with value investors regarding fundamental indicators, namely the ones about the company’s profit margins such as earnings growth or revenues, growth investors are not concerned with the calculation of the stock’s intrinsic value or its market current price being over or undervalued. Growth investors look for stocks that present a high projected growth rate. Even if the stock is a bit overvalued, if the perspective growth rate is 10% to 15% per year on a 5-year range, it will catch the investor’s eye. For all that, a
growth stock can be defined as a stock with expectations to grow at an above average rate compared to the overall market. Knowing the theory is a thing, but finding this growth stocks is not as easy as it may seem. As a start, there is no automatic formula, it will always depend on self interpretation and judgement. Leaving aside this fact, growth analysts can proceed from general guidelines:

(a) **Historical Earnings** - Based on the annual revenue, did the company grow? The philosophy here is, if a company verifies a certain rate grow for a period of 5 or 10 years, it is likely that it will continue to verify this grow for the next 5 or 10 years.

(b) **Forward Earnings** - Growth analysts look for companies that not only performed well on the latest years, but that have a projected growth rate for the following years of 10% to 15%. Normally, to be consider a growth stock, the stock price should be able to double ub five years. The difficult here is that this are just estimations.

(c) **Management** - The way the company’s leaders are managing the costs and revenues (qualitative analysis of the company) is an important factor when it comes to decide which securities the investor wants in his portfolio. An high growth in sales does not always represent growth gains in earnings. If the earnings are not keeping up the high annual revenue, it may be a signal of a poor managing that could lead the company to debt. A pre-tax profit margin exceeding the last five years pre-tax profit margin average is a good indicator of good management. Besides that, it is also important to analyse the efficiency of the business. An efficient use of the assets is reflected on a stable or steadily increasing return of equity in the last 5 years.

As stated before, there are hundreds of economic and technical indicators, and some of them are very useful in this area. Further in this paper, a detailed analysis of some of those indicators is provided.

**C Growth At a Reasonable Price (GARP) Investing** is the merging of the two great investing philosophies analysed above. Warren Buffet, known as one of the greatest supporters of value investing, said that growth and value investing are not contrasting approaches, having called ignorants to investment managers who do not support that theory. In fact, Peter Lynch pioneered and is one of the biggest supporters of this hybrid investing strategy commonly referred to as GARP strategy. As stated, this strategy is the combination of both value and growth investing. Value investors want companies that are selling below their intrinsic value. Growth investors want companies that present a high perspective of growth. GARP investors care about companies with good growth prospects, that is, companies presenting positive revenues for the past years and positive projections, but are sceptical about excessively high growth estimations. Besides, investors who follow this approach do not feel comfortable buying securities selling highly above their intrinsic value. Consequently, GARP investors are very fond of the Price/Earnings to Growth ratio (PEG). PEG ratio determine a stock’s value, taking into account the company earnings growth. PEG ratio rounding 0,5 are the numbers GARPers are looking for, which means that the company is not overvalued and its growth projections are good enough.

### 2.1.5 Categories of Companies

Conjointly with the factors already mentioned, investors and analysts should complement their research about the company by analysing the category in which the company is inserted. Depending on the investor’s philosophy, companies can represent a more attractive shot considering their category. In tune with Lynch and Rothchild, there are six major categories of companies, being this classification based on the company’s growth rate, capitalization and economic behaviour [9]. For each, the main characteristics are described:

**A Slow Grower** As the name suggests, these are growing companies that present a low growth rate, hardly above the gross domestic product (GDP). Old and big companies are the ones more likely to fall in this category, as they have already reached the maturity phase, and are not expected to present big growth. The advantage in investing on this type of companies is that normally, companies like this report remarkably consistent numbers regarding revenues and sales, growing year after year, which can represent a good investment strategy for a stable payout, as this companies generally bring out a good dividend payout ratio.
B **Stalwart** Stalwart company is a term made popular by Linch, who used as an example Coca-Cola. That being said, stalwart companies are large companies (multi-billion dollars companies), growing slightly faster than slow growers. Investors can expect slow growth prospects, but steady, which can be a very appealing business. With a annual gain of approximately 10% to 12% and no chance of bankruptcy overnight given their large capitalization and durable competitive advantage, even in recession times this companies provide their investors with fairly good returns.

C **Cyclical** The cyclical companies are highly dependent and vulnerable to the overall economy. As the economy goes up, cyclical companies’ stock prices go up, but when the economy is in bad shape, the revenues decline as well as the stock’s price. These are companies, for example, in the airline, auto or tire industries. The principle is quite simple, these companies’ business is based on discretionary items that consumers can afford to buy more when up on a thriving economy, causing the fast growth, up rising revenues and profits, and consequently, the upgrade registered on the stocks’ price. However, being this dispensable items, when the economy is not doing well, those expenses are the first things consumers will cut, provoking the opposite effect. In big crisis, this companies may even go bankrupt.

D **Fast Grower** A growth company is any firm generating positive cash flows at faster rates than the overall economy. It is considerer as a fast grower company when growing rates are above 20% per year. Usually, these are small and aggressive new enterprises, that opt to reinvest its own retained earnings and profits in order to expand the business fast. From an investor perspective, these can represent good investments, as the potential for increase value is huge, but they come with a higher risk and volatility in the stock prices than the other types of enterprises.

E **Turnaround** Some companies go through very hard times, that can even led them to bankruptcy, though they were great and stable companies once. The poorly performing can be related to several aspects, as bad news for the company’s business, competitive changes on the industry that may have caused the company’s services to be perceived by consumers, bad management or management scandals of corruption, etc. However, when these companies manage to identify the problematic factor that caused the poor performance and solve them, it is probable a quick recovery of the stock price. This effect gave the name to this type of companies, and an investor may profit from it if he anticipates this turnaround.

F **Asset Play** An asset play company is a company believed to be incorrectly-valued, having stocks that are undervalued as its combined asset value is higher than its market capitalization, this is, the current stock price does not reflect the real value of the company’s assets. These companies hold valuable assets that worth more than what the company made public, offering assets to the market at low price, making an attractive buy.

### 2.1.6 Sectors

In this section, the concept of sector and industry classifications is explained. Over the years, in financial research, the industry classification of companies is essential. The main objective/motivation is to design a classification system capable of grouping similar entities. This similarity can be based on different metrics, such as market performance, production process or product-output. Usually is based on the last two.

There are several motives to endorse classification systems when upon an academic research: (1) the most logical one, comparison purposes. The majority of researches that use industry classification, use it to compare entities. It may not be a very valuable thing to compare two very distinguish entities. (2) the second good reason is related to performance measurement. In financial research, is important to ask “Is this company being superior to his similar?” Historically, some sectors are more competitive. In some cases only the best company from the given sector endures over the decades. (3) Segment valuation. A simple way to evaluate the industry is to verify if the companies operating obtain good results. Given the importance and influence of these classifications in financial and accounting publication, a vital characteristic of the system is homogeneity, this is, the capability of the system to assign the same classification to analogous companies.

Next, the three main classification systems are presented. On this work, special focus is given to the Global Industry Classification Standard (GICS).
1. **Standard Industrial Classification (SIC)**
   This system is the first of its kind, created in 1939 by the U.S Office of Management and Budget. It distributes companies over 10 divisions containing 81 major groups, typically based on sales information. The largest segment wins. Due to its age, several problems appeared such as the new industries. Although, it was extensively used in research, maybe there was no better alternative and due to the easy access.

2. **North American Industry Classification System (NAICS)**
   The successor of the SIC system in 1999. Follows the main rules and structure, although providing three major upgrades: classifies companies accordingly to their production process; fast response on situations of emerging industries; provides a common classification system for business in Canada and Mexico.

3. **Global Industry Classification Standard (GICS)**
   The most recent system, developed by the Standard&Poor’s (S&P) and Morgan Stanley Capital International (MSCI). It is a system developed by Wall Street analysts for Wall Street analysts. Like SIC and NAICS, it's a well-established hierarchical industrial classification. GICS can be seen as a four layered pyramid. From the upper layer to the bottom, GICS consists of 10 sectors, 24 industry groups, 67 industries and 156 sub-industries (see [1]). Differently from the other approaches, companies are evaluated quantitatively and qualitatively. This means that GICS classification is mostly based on revenues, but also use earnings, market perception and analysts perception.
   Investors can count on key features such as:
   - **Universal** industries classifications, covering over 95% of the world’s equity market.
   - **Reliable** structures, reflecting the current markets.
   - **Comprehensive** hierarchy, ranging from the most general sector to the most specific industry.
   - **Evolutionary** system, every year classifications are reviewed.

   ![GICS Hierarchy Diagram](image)

   Figure 1: GICS Hierarchy Diagram

   Not many previous works studied the performance of the different systems. Although, a few concluded that GIGS provides a better distribution of firms [10], and that GICS outperforms the rivals on homogeneity, either for large or smaller firms [11].

   Reports of each sector’s performance are published every quarter. In the figure 2 the reader can observe the performance graph of each sector for the last 5 years. Notice that although some curves follow a similar trend, there are clear evidences that some sectors outperform the others.
In figures 3 and 4 are represented the graphs for the S&P 500 sector health care and for the S&P 500 sector energy. It is a straightforward conclusion that an investment in the health care sector is a much better option than the energy sector.
That being said, to emphasize the already stated innovation in this work, the objective is to take advantage of these graphs and choose the best sectors, not investing in the companies from the worst performing sectors.

2.1.7 Financial Statements

Financial statements are the reports provided by companies from time to time, normally quarterly or annually, used for reporting their financial status and corporate activity. This documents adhere to generally accepted accounting principles (GAAP) and are meant for both the company and general public. Financial statements for business usually include three documents: income statements, balance sheet and cash flows. Those present all information about the financial performance and position changes. The documents provide means for analysts and value investors to do the quantitative (fundamental) analysis of the company.

A Balance Sheet

The balance sheet is an accountant’s snapshot of the company’s accounting value on a particular date [12]. Provides information of what the firm owns and how it is financed. Through the two sides that constitute the balance sheet, investors have an idea of the company owns and owes, and the shareholders’ investments.

The three segments are: the assets on the left side and the liabilities and stockholders’ equity on the right side. The accounting definition that describes the balance is

$$Assets \equiv Liabilities + Stockholders’equity$$  \hspace{1cm} (1)

This definition must always hold. The left side, the assets side, is ordered by the amount of time it would take for the company to convert them in cash. The right side, the liabilities and the stockholders’ equity side, is ordered by the payment priority. The balance sheet provides to the analyst three important aspects: accounting liquidity, debt versus equity, and value versus cost.

Accounting Liquidity - Some assets are not tangible, meaning that is not easy to convert them in cash if needed. When analysing the balance sheet is important to define the assets that can be quickly converted to cash. Current assets are the most liquid and fixed assets are the least liquid.
Debt versus Equity - Generally, creditors have always the first claim on the company’s cashflow. This happens because companies that borrow money make contractual obligations to repay the money plus interest over a period, and if not, creditors may sue the company, which can result in bankruptcy. When analyzing the potential and performance of a company through the balance sheet, it is crucial to look into the stockholders’ equity.

\[ \text{Assets} - \text{Liabilities} \equiv \text{Stockholders' equity} \] (2)

Value versus Cost Investors should have a clear idea about the information that can be found on the balance sheet. The book value of the company, even though it says “value”, under the GAAP terminology, are based on cost. Many investors and managers want to know the value, and this is not found on the balance sheet.

An example of a balance sheet can be observed in 5.

<table>
<thead>
<tr>
<th>ANF COMPANY INC.</th>
<th>LIABILITY AND SHAREHOLDERS' EQUITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>BALANCE SHEET</td>
<td></td>
</tr>
<tr>
<td>DECEMBER 31, 2014</td>
<td></td>
</tr>
<tr>
<td><strong>ASSETS</strong></td>
<td><strong>Current Liabilities</strong></td>
</tr>
<tr>
<td><strong>Current Assets</strong></td>
<td></td>
</tr>
<tr>
<td>Cash</td>
<td>$ 100,000</td>
</tr>
<tr>
<td>Accounts Receivable</td>
<td>20,000</td>
</tr>
<tr>
<td>Inventory</td>
<td>15,000</td>
</tr>
<tr>
<td>Prepaid Expense</td>
<td>4,000</td>
</tr>
<tr>
<td>Investments</td>
<td>10,000</td>
</tr>
<tr>
<td>Total Current assets</td>
<td>149,000</td>
</tr>
<tr>
<td><strong>Property and equipment</strong></td>
<td></td>
</tr>
<tr>
<td>Land</td>
<td>24,000</td>
</tr>
<tr>
<td>Buildings and improvements</td>
<td>250,000</td>
</tr>
<tr>
<td>Equipment</td>
<td>50,000</td>
</tr>
<tr>
<td>Less accumulated depreciation</td>
<td>(5,000)</td>
</tr>
<tr>
<td><strong>Other assets</strong></td>
<td></td>
</tr>
<tr>
<td>Intangible asset</td>
<td>4,000</td>
</tr>
<tr>
<td>Less accumulated amortization</td>
<td>(200)</td>
</tr>
<tr>
<td><strong>Total assets</strong></td>
<td>$472,100</td>
</tr>
<tr>
<td><strong>Liabilities and Shareholders' Equity</strong></td>
<td><strong>Total Liabilities</strong></td>
</tr>
<tr>
<td><strong>Current Liabilities</strong></td>
<td></td>
</tr>
<tr>
<td>Accounts payable</td>
<td>$30,000</td>
</tr>
<tr>
<td>Notes payable</td>
<td>10,000</td>
</tr>
<tr>
<td>Accrued expenses</td>
<td>5,000</td>
</tr>
<tr>
<td>Deferred revenue</td>
<td>2,000</td>
</tr>
<tr>
<td><strong>Total current liabilities</strong></td>
<td>47,000</td>
</tr>
<tr>
<td><strong>Total liabilities</strong></td>
<td>247,000</td>
</tr>
<tr>
<td><strong>Shareholders' Equity</strong></td>
<td></td>
</tr>
<tr>
<td>Common stock</td>
<td>10,000</td>
</tr>
<tr>
<td>Additional paid-in capital</td>
<td>20,000</td>
</tr>
<tr>
<td>Retained earnings</td>
<td>197,100</td>
</tr>
<tr>
<td>Treasury stock</td>
<td>(2,000)</td>
</tr>
<tr>
<td><strong>Total liabilities and shareholders' equity</strong></td>
<td>$472,100</td>
</tr>
</tbody>
</table>

Figure 5: Example of a Balance Sheet

B Income Statement
The income statement measures performance over a specific accounting period. The accounting definition of income is

\[ \text{Revenue} - \text{Expenses} \equiv \text{Income} \] (3)

It can be seen as the recording of activities performed between two balance sheets, as it gives a summary of how the revenues and expenses, net profit or loss, incurred over a trimester or a year. As explained by Ross [12], usually it includes three sectors: The operations section, which shows the revenues and expenses from principal operations. Also, in this sector, analyst can consult the earnings before interest and taxes (EBIT), a important fundamental indicator of the company. The non operating section, that includes all financial costs, such as interest expense and taxes. Finally, the third sector, the net income, frequently expressed as earnings per share. An example of a income statement is available in 6.
Considered by many the most important statement of the financial statements, as it represents the actual cash flow of the company, this is, all the cash inflows and outflows during a given period. Inflows can be generated by the operations of the company, such as sells, and the result of external investments. The outflows represent the pay for business activities, such as suppliers, and investments during the period. Cash flow statement is important because is not the same as net working capital. Income statements may not reflect the changes in cash, which is what values a company. From equation 1:

$$CF(Assets) = CF(Liabilities) + CF(Stockholders' equity)$$  \hspace{1cm} (4)

Cash flow is defined for three different major sources: Operation cash flow, investing cash flow and financing cash flow.

**Operating Cash flow** is the cash flow generated by the company’s services or sales, and reflects tax payments.

**Investing Cash flow** is the cash flow involving the changes in current and fixed assets, such as acquisitions or disposition of assets.

**Financing Cash flow** is the cash flow that arises from external activities such as issuing cash dividends and more stocks, sales of equity securities, or redemption of long-term debts and capital stock.

### 2.1.8 Indicators

**A Financial**

Financial ratios are indicators that expose a company’s performance and financial condition. Its part of
the quantitative analysis, as most formulas for this ratios use values provided by the financial statements. The principal use for them is for comparison purposes, this is, company’s should calculate the ratios for the current year and then compare the results with the ones from previous years. Analyst can also use this ratios to compare the company’s fundamentals to those of other companies on the same industry. There are dozens of ratios, next are described some of the most important and common financial ratios used by investors, divided in four categories: profitability, liquidity, leverage and market value ratios.

(a) Profitability Ratios

**Return on Equity (ROE)** shows how profitable a company is by comparing the net income with the money shareholders have invested on the company, this is, reveals the percentage of profit that the corporation earned. The higher the ratio, the higher the returns for the shareholders.

\[
Return \text{ on Equity (ROE)} = \frac{Net \text{ Income}}{Average \text{ Shareholders' Equity}} \tag{5}
\]

**Return on Assets (ROA)** shows how profitable a company is by comparing the net income with the company’s total assets. Is reveals if the management is employing the assets well in order to make money, this is, the higher the ROA, the better capacity to generate money.

\[
Return \text{ on Assets (ROA)} = \frac{Net \text{ Income}}{Total \text{ Assets}} \tag{6}
\]

**Return on Capital Employed (ROCE)** complements the ROE ratio, this is, evaluates the capacity to generate returns from all capital available. As in the ROE, it uses the shareholders’ equity, adding the company’s liabilities. This way, analyst have a better profitability indicator.

\[
Return \text{ on Capital Employed (ROCE)} = \frac{Net \text{ Income}}{Total \text{ Assets} + Debt \text{ Liabilities}} \tag{7}
\]

**Net Profit and Pretax Profit** show the company’s profit margin. Net profit is usually the most mentioned ratio to discuss the company’s profitability. It shows how much money can the company keep from the earnings made in sales.

\[
Net \text{ Profit} = \frac{Net \text{ Income}}{Revenue(NetSales)} \tag{8}
\]

The pretax profit represent the earnings before tax. Analysts often prefer this ratio as they can analyse the tax-management techniques adopted by the company in order to maximize the returns.

\[
Pre - tax \text{ Profit} = \frac{Pre - tax \text{ Income}}{Revenue(NetSales)} \tag{9}
\]

(b) Liquidity Ratios

**Current Ratio (CR)** shows the ability of a company to cover current liabilities. It is the most popular ratio to evaluate the firm’s liquidity. An high value indicates that the assets of the company can easily cover the liabilities, this is, the short-term assets are readily to pay off the short-term liabilities. A positive high CR reduces the risk of the investment, however a lower CR may not necessarily be a bad indicator, it could mean that the company is using their assets to grow the business.

\[
Current \text{ Ratio (CR)} = \frac{Current \text{ Assets}}{Current \text{ Liabilities}} \tag{10}
\]

**Quick Ratio (QR)** represent the ability of a company to cover current liabilities with its most liquid assets. CR has a drawback which is: total assets include the inventory. The problem is that inventories may include things hard to liquidate and with uncertain liquidation values. To know the exact amount available to pay the short term obligations, the QR was created, which basically refines the CR, by only considering cash, accounts receivable and notes receivable.

\[
Quick \text{ Ratio (QR)} = \frac{Current \text{ Assets} – Inventories}{Current \text{ Liabilities}} \tag{11}
\]

13
(c) **Leverage Ratios**

**Debt Ratio (DR)** shows the dependency on leverage. The lower the DR, the less leverage the company is using, and the less risk the company have. A high percentage indicates an high debt-load, which means the company is currently more financed by money owed to others.

\[
Debt Ratio (DR) = \frac{Total \ Liabilities}{Total \ Assets} \tag{12}
\]

**Debt Equity Ratio (DER)** is similar to the DR, the lower the percentage, the less leverage used and less risky. Although, debt equity ratio compares the company’s liabilities to the shareholders’ equity, showing if the company is being more financed trough debts or is being financed most by the stakeholders.

\[
Debt to Equity Ratio (DER) = \frac{Total \ Liabilities}{Shareholders' \ Equity} \tag{13}
\]

(d) **Market Value Ratios**

**Earnings per Share (EPS)** can serve as a profitability indicato, since it reports the portion of company’s profit allocated to each outstanding share of common stock, this is, represents an stock’s performance. The number of shares outstanding can change over the year, for example each six months. For that reason, when calculating the EPS, is good practice to use a weighted average number of shares outstanding over the reporting term.

\[
Earnings \ per \ Share \ (EPS) = \frac{Net \ Income - \ Dividends \ on \ preferred \ stock}{Average \ Outstanding \ Shares} \tag{14}
\]

**Price to Earnings (PER)** shows how many times a stock is trading per each dollar of EPS. Its the most reported and used ratio by investors. Given that it uses the EPS ratio, a low PER is preferred, as it indicates the number of years required to retrieve the investment. Although, a high PER can also be good news. High PER ratios may indicate positive expectations about the payout or substancial growths on earnings compared to the overall market.

\[
Price \ to \ Earnings \ (PER) = \frac{Stock \ Price}{Earnings \ per \ Share \ (EPS)} \tag{15}
\]

**Price to Earnings Growth (PEG)** is a refinement of the PER. Differently from the PER, PEG takes into account the growth perspectives of the stock’s earnings. Supported by the PEG ratio, investors and analyst are given an idea of the stock’s valuation, this is, if it is over or undervalued, relative to the growth perspective. Being the unity the neutral value, less than one means that the stock’s price is being undervalued, and the oposite for values with higher PEG ratios.

\[
Price \ to \ Earnings \ Growth \ (PEG) = \frac{Price \ to \ Earnings}{Earnings \ per \ Share \ (EPS)} \tag{16}
\]

**Price to Cash Flow (P/CF)** Shows the relation between the amount of cash flow that the company generates and the stock’s price. It is similar to the PER, being prefered by many to evaluate the company’s stock as cash flows are not easily manipulated, giving a more truthful value.

\[
Price \ to \ Cash \ Flow \ (P/CF) = \frac{Stock \ Price \ per \ Share}{Cash \ Flow \ per \ Share} \tag{17}
\]

**Price to Book Value (P/BV)** expresses how much shareholders pay for the net assets of a company. It relates the the market value of the company to the value of a company’s assets on the balance sheet statement. This ratio gives to invstors an indication about the under or overvaluation of the stock, showing how many times its intrisic value a company’s stock is trading.

\[
Price \ to \ Book \ Value \ (P/BV) = \frac{Stock \ Price \ per \ Share}{Shareholders's \ Equity \ per \ Share} \tag{18}
\]
**Dividend Yield Ratio (DY)** relates dividends with the share price. Represents the annual expected return in dividends, relative to its share price. Depending on the investing strategy, investors prefer low or high percentages for the DY ratio. Growth investors have little interest in dividends, as they prefer the company to invest in the company growth, expecting to further capture large capital gains.

\[
Dividend \text{ Yield (DY)} = \frac{\text{Annual Dividends per Share}}{\text{Share Price}}
\]  

**Dividend Payout Ratio (DPR)** shows the percentage of the earnings (revenues or net income) distributed as cash dividends. A good DPR indicates that the earnings cover the dividend payment to the shareholders.

\[
Dividend \text{ Payout Ratio (DPR)} = \frac{\text{Dividends per Share}}{\text{Earnings per Share}}
\]

### B Market

(a) **Trend Indicators**

Trend indicators measure the main direction of the market. A market trend is the moving direction of the market during a certain period. Trends are classified regarding their time frame: secular trends for long periods, primary trends for medium and secondary trends lasting short times. These indicators are not meant to lead a price of a security, as a trend has to be establish first to be measured and this is a time cost process. Despite that fact, trends are them most important and stronger indicator in technical analysis, since stock markets are, for most of the time, trending.

(b) **Market Breadth Indicators**

Breadth Indicator determine the strength of a trend or of the market condition. Once the trend is identified, is useful to know if the trend is well supported. To achieve this, analysts compare the volume of rising stocks and the volume of falling stocks. In a good up trend (bullish trend), the volume and number of rising stocks should be dominant.

(c) **Contrarian Indicators**

Track down the investing behaviour of different kind of investment groups, this is, tries to catch the attitude among traders and investors, to measure the overall bullish or bearish approach. Are useful to confirm price action. Contrarian investors believe that when everybody is going on the same way, there will be no more purchasing power, meaning that the market is at a peak.

(d) **Oscillators**

Designed to lead price movements and are oscillating around a certain value, these market indicators are leading indicators. They oscillate between two extreme values, indicating if the market is deemed to be overbought or oversold. When upon a strong bullish trend, if the oscillator goes to the lower extreme, it means that the market is oversold, and is time to buy. The other way around is not necessarily true, which constitutes a enormous disadvantage to this market indicator.

### C Technical Indicators

Contrary to financial ratios, technical indicators do not care about any fundamental part of the company. These indicators are part of technical analysis, this means that technical indicators use information retrieved from the charts, such as stock’s open price, close price, high price, low price, volume of transactions, etc. Observing the results in a plot diagram gives a perspective of the stock’s price evolution and trend. Although some indicators have a more successful historical performance, there is no guarantee that the signals will work. Investors should not base their investments in only one indicator. Instead, it is good policy to combine several indicators. Matching fundamental ratios, there are dozens of technical indicators at investors disposal. Here is presented one of the most common indicators:

(a) **Relative Strength Index (RSI)**

Momentum oscillator indicator, informs the investor if the stock is overbought or oversold. Its value is always in the interval \([0 – 100]\), floating on the basis of average returns and average losses.
across a defined period of time. When the signal is up to 70, it indicates a strong overbought stock. A strong oversold is indicated by values below 30.

\[
RSI = 100 - \frac{100}{1 + RS}, \quad \text{where } RS = \frac{Average \ Gains(Up \ closes)}{Average \ Losses(Down \ closes)}
\]  

(21)

2.2 Portfolio Theory

As already mentioned, in a more broad sense, investing is all about picking the best assets to include in a portfolio. When it comes to finance, a portfolio consists in a collection or combination of financial assets, the securities, such as stocks, futures, bounds or cash. The process of selecting a portfolio may be divided into two stages \[2\], the stage of observation, which lead to beliefs about the future performance of the available securities, and the stage of choosing the relevant securities based on the most relevant and strong beliefs. To construct the portfolio, the investor’s comprehension about the degree of risk is crucial. Investors must have a well designed set of goals and be willing to accept the risk derived from the chosen portfolio allocation.

2.2.1 Portfolio Management

Some investors are more inclined to accept risk than others. Normally, there is a direct proportion between the expected returns and the risk. The higher the risk, the higher the returns. Although, the probability of incurring in a loss is also higher. The management of a financial portfolio depends upon the investor. In general, portfolio investment strategies can be divided in two types, aggressive investment strategies and conservative investment strategies. Aggressive portfolios are for those who seek the highest possible return in a long time horizon, people able to take wide variations (high risk tolerance). Conservative portfolios are for those who prefer the safety (risk averse) to high returns.

2.2.1.1 Modern Portfolio Theory

In the year 1952, Harry Markowitz introduced a theory, called the modern portfolio theory, where he explain how investors can use the diversification rationally in order to achieve higher returns \[\text{[3]}\]. In his theory, Markowitz assumes that the normal human being is not very found of risk, using this fact to construct the “expected returns - variance of returns” rule. When facing two investment possibilities with same expected return, investors will always choose the one with less risk, i.e, variance of returns. Below are presented the mathematical formulas developed in his work:

The discounted anticipated return of a portfolio is given by

\[
R = \sum_{i=1}^{N} R_i X_i
\]  

(22)

Where \(N\) is the number of securities, \(R_i\) is the discount return of the \(i^{th}\) security and \(X_i\) is the relative amount invested on the \(i^{th}\) security. Since \(X_i \geq 0\) and \(\sum_{i=1}^{N} X_i = 1\), \(R\) is a weighted average of \(R_i\).

The objective here is to maximize \(R\), being a possible approach the follow: let \(X_i\) be 1 for the \(i^{th}\) security with max \(R_i\), this is, having a portfolio with only one security. Although this approach may seem a good and profitable idea, it goes right against the “diversify your portfolio” rule. To diversify and maximize, Markowitz created the expected returns-variance of returns (E-V) rule. \(R\) can be seen as a weighted sum:

\[
R = \alpha_1 R_1 + \alpha_2 R_2 + \cdots + \alpha_n R_n
\]

The expected value, \(E\), of a weighted sum is the weigh sum of the expected values.

\[
E(R) = \alpha_1 E(R_1) + \alpha_2 E(R_2) + \cdots + \alpha_n E(R_n)
\]  

(23)

The variance, \(V\) of a weighted sum is given by:

\[
V(R) = \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j \sigma_{ij}
\]  

(24)
To conclude, given the expected value of $R_i$ ($\mu_i$), the covariance of $R_i$ and $R_j$ ($\sigma_{ij}$), total yield of the portfolio $\sum_{i=1}^{N} R_iX_i$ and $\sum_{i=1}^{N} X_i = 1$, the expected return $E$ and the variance $V$ is

$$E = \sum_{i=1}^{N} X_i \mu_i \quad ; \quad V = \sum_{i=1}^{N} \sum_{j=1}^{N} X_iX_j \sigma_{ij}$$

(25)

For different securities, investor has a choice of various combinations of $E$ and $V$ depending on the values for $X_1, X_2, \ldots, X_N$. With the Y-axis projecting the variance (or risk) and the X-axis projecting the expected return, its possible to plot a graph such as the one on the figure and select the desired portfolio. From all the attainable $E,V$ combinations, investors can select a few called the efficient combinations, also known as the Pareto front, represented in the image by the curve pointed as “efficient $E,V$ combinations”.

Figure 7: Graph generated using the Markowitz Modern Portfolio Theory. Image obtained from Markowitz’ book [1]

Since this theory was published back in 1952, many critiques were made. In 50 years the world suffered big changes, specially in technology. With the new science areas and new researches within economics, financial and the stock market world, some assumptions made by Markowitz begin to be criticized and considered wrong. The biggest critics lie on the Markowitz’ assumptions regarding the human behaviour. Assumptions like “investors act rationally”, “investors have similar view of what risk is”, “investors look at investments over equal time periods” are obviously wrong. Investors are not that rational, their behaviour can harshly be predicted. Also, many type of investments can be made. A day trader will not have the same horizon and risk idea as a long term investor. Also, Markowitz’ model does not take in consideration transactions costs, limitations on the investment size, politics, sell restrictions etc.
To conclude, the modern portfolio theory is, nonetheless, the greatest and most widely theory used when it comes to risk evaluation of an investment. Despite all criticism, from 1952 until today, Markowitz’ model still is the leading work in his area.

2.2.2 Diversification and Risk

When investing with portfolios, the objective is to maximize the expected returns and minimize the risk associated. The portfolio’s risk can be split in two parts: Systematic risk and non-systematic risk. The systematic risk is the risk that comes from the economic and political uncertain. This risk affects the overall market, and it is constant, i.e., it is a risk inherent to every sector and company and cannot be mitigated. The non-systematic risk is specific to each company, exclusively dependent on the strategic behaviour of the company. Its based on the financial structure of the company, the financial statements, the news affecting the company, the competition in the sector, etc. Contrary to the systematic risk, the non-systematic risk can be controlled.

The most popular and used technique to minimize the portfolio’s risk is diversification. Using a metaphor, the idea of diversification is not to put all the money under the same pillow. In financial terms, diversification means to spread the risk. Securities will perform differently, so by mixing a variety of securities, investors achieve a greater stability. In case some of them suffer a great decline, the entire portfolio will resist the impact and reduce the financial losses.

Notwithstanding, the risk will never be null. Even with a perfect diversification technique that shrunk down to zero the non-systematic risk, there is always the systematic risk, which cannot be totally eliminated. Thought there are several risk-management techniques, such as mixing securities from different countries to reduce the impact of political and economic decisions, it capital that investors understand that these techniques only reduce or prevent loss, never guarantee results or that we avoid a losing investment.

2.3 Evolutionary Computing

Since its creation, computer science has been trying to develop automated problem solvers, i.e., automated algorithms. Nature’s solutions, with its “natural problem solvers”, inspired people to work in order to copy those mechanisms for every engineering fields. In this work, the interest goes to the breach of artificial intelligence. Evolutionary computing is based on the natural evolutionary process, copying biological procedures and methodologies to achieve an optimal solution for a given problem. There are several types of problems which can be solved using evolutionary computing, the most popular and explored are the optimization problems. Optimization problems are those where the desired output is known, the model to evaluate the inputs is known, and the task is to find the input(s) leading to the desired output. Constructing and managing a financial portfolio is a good example (and part of this work) of an optimization problem. Generally, evolutionary computing has associated meta-heuristic optimization algorithms, known as evolutionary algorithms, which basically are algorithms created with the optimization problem in mind.

2.3.1 Evolutionary Algorithms

Evolutionary algorithms are mostly based on the Darwin’s theory and principles about evolution, the natural selection theory. The idea of natural selection is simple: given a population of individuals on a certain environment, the strongest and fittest survive and seed the next generation, theoretically well adapted to the environment. Having this in consideration, this process can be seen as a cycle, moving from state to state until a candidate fulfils the requirements (a solution). There are two key forces on these systems: variation operators to create a population diversity, and selection must act as a force pushing quality. The general diagram of an evolutionary algorithm is given in figure

As the reader can observe from the above figure, a evolutionary algorithm follows a collection of simple rules: generate a random population (the initial population); select the fittest individuals from the population and apply a set of operators to refine them; evaluate the population to analyse if the desirable criteria was matched. Case a solution had been found, the algorithm terminates, else the algorithm repeats with the new refined population. Genetic algorithms are the most widely known type of evolutionary algorithm. Below the steps are carefully explained.
1. **Definition of Individuals**
   The first step is the representation of the individuals, also called chromosome. This chromosomes are nothing but a collection of variables. The encoding of the chromosomes is called genotype, and the best solution is obtained by decoding the genotype after the termination of the algorithm. This genotype contains a chromosome with the best variables for the problem in hand.

2. **Evaluation Function**
   The evaluation function represent the requirements desired. Given a chromosome, this function measures its quality in solving the problem. Then, the better chromosomes are picked to generate the new population. At each iteration the mean is supposed to be better, until the ideal criteria is matched.

3. **Population**
   The population is the group of possible solutions, a set of genotypes. This population changes at each iteration, as result of the evolution. The process of picking the fittest genotype or the parent selection is at population level.

4. **Parent Selection**
   The selection process is to determine the genotypes that will seed the next generation. Based on their quality, the best are the next parents. This step is responsible for the quality improvements. Notice that sometimes, low-quality individuals are selected, although the probability is small.

5. **Genetic Operators**
   This phase of the evolutionary algorithm is for the creation of new individuals from the old generation. There are two possible variation operators: Mutation and crossover.

   **Mutation** is a unitary operation. In this process, a genotype is chosen and a new one (slightly) different is created, called the offspring. It is based on random changes on the values of the variables (genes) of the chromosomes.

   **Crossover** is a binary operation, also called recombination. In this operation, two parent genotypes are chosen, and their genes are merged into one or two offspring individuals. It is also a random process, as the parts that are combined depend on random drawings. The motivation for this process is simple. By mating two chromosomes with good qualities, the offspring may end with
only the good qualities from both parents. Nonetheless, there is no guarantee that the offspring will be of better quality.

6. Survivor Selection

After the offspring conclusion, the replacement begins. Based on their quality, the individuals are distinguished and selected to constitute the new population. This selection is similar to the parent selection, but in a different phase of the evolutionary cycle. The strongest (fittest) individuals survive natural selection. If any of these individuals match the desired characteristics, the algorithm finishes. Otherwise, the cycle restart with the new population.

2.4 Existing Solutions

Given the unpredictability of the market in the real world, problems are not linear. In the majority of cases, there is not a perfect solution, instead problems have a set of solutions that combined constitute a optimal possible solution.

Based on this problem, the artificial intelligence gained relevance in the area of forecasting financial markets and optimization of funds in a portfolio. Some of the many solutions developed over the last decades are presented here, with special relevance to the ones using evolutionary algorithms. Nonetheless, some other solutions using different artificial intelligence techniques are given a brief overview, such as neural networks (ANNs), fuzzy logic systems and mixes of both with genetic algorithms.

2.4.1 Genetic Algorithms Related Work

Yan & Clack in 2007 [14], based on the assumption that historical facts that caused crashes and market environments with great volatility repeat over time, proposed a genetic solution to achieve robustness when faced to an out-of-sample very volatile environment. In their work, performance is tested across several environments, such as bearish and bullish market, and considering three possible evolution scenarios for the population, (1) Standard GP (SGP), (2) Multiple-Scenario evaluation in the last generation (MELG) and (3) Multiple scenario evolution (MEVO). This brings advantages as it tests individuals in multiple possible scenarios, calculating a degree of volatility for each one. The fitness function chosen by the authors is a unique-objective approach, the Sharpe Ratio, which is calculated with the ROI and the Risk Free Rate. The system performed better with the unique-objective approach when opposed to one using a multi-objective approach, so the authors opted for the first. The Moving Average Convergence Divergence (MACD) was also used to add date in order to compare it with the three GP solutions for better result evaluation. The results showed that all three GP systems, compared to the portfolio and overall market index, presented better results when it comes to volatility. When it comes to robustness, MELG performed the best, also presenting greater returns in comparison to the portfolio index.

Hassan & Clack in 2009 [15], pioneered testing of multiple objective genetic programming (MOGP) solutions’ robustness in out-of-sample environments. The approach was simple, based on the data available of the stock, the algorithm has to rank its future performance, even if it can not predict the precise future price. Because the authors pretended to test unseen environments, problematic concerns appeared, regarding the gathering of new data and the decision of when should the system retrain. The genetic algorithm used was the SPEA2, using two known techniques hoping to achieve better robustness, the Mating Restriction and Diversity Preservation technique. To test the robustness a K-means clustering algorithm was used, being the best solutions the ones that obtained the same rank during training and validation, i.e the more robust solution. Other metrics were considered, and the results regarding performance for the out-of-sample environment were compared with the ones obtained with a Buy-and-Hold strategy, evaluating the quality of the solution, how preservation of solutions order change and the distribution of solutions on the front. The proposed MOGP performed the best, supporting that restraining the mating to niche characteristics, provoking speciation, does benefit robustness.

Kaucic in 2012 [16], developed a solution based on the technical analysis, this is, an evolutionary system supported on several technical indicators. In 2010 Kaucic [17] had already proposed an investment strategy using evolutionary computation combined with machine learning techniques and technical rules to generate trading systems, obtaining positive results when tested on simulation markets for the three main types,
bull, bear and no-trend market. In his work of 2012, the author deploys a solution composed of three different modules. The investment module, denotes the long-plus-short strategy for the portfolio. The technical module, where several technical indicators are applied to detect promising stock, such as ROC, RSI, SMA, HMA, VMA, MACD and OBV. The last module is the evolutionary learning module, where a genetic algorithm is used to select the best combination of technical signals, using the reward-to-risk ratio (ratio between the reward and risk of the portfolio) as performance measure. Using three distinct ways of calculating the risk of the portfolio, three evolved portfolios where created, Information Ratio (IR), Sortino Ratio (SR) and Omega ratio. All three evolved portfolios performed significantly better than the DJI index, being that Omega produced the best results, 26% annualized mean returns. Comparing the three portfolios with the Sharpe Ratio, Omega won again.

Esfahanipour & Mousavi in 2011 [18], implemented a model called Buy and Sell (B&S) using GP to generate trading rules: buy or sell. The idea is to generate one of three possible signals: buy, sell, no trade. The innovation is the insertion of a third possibility, the “no trade” signal. Using historical data for training, the authors managed to extract these rules from tree data structures. These trees are constructed carefully in order to let the assumptions of GP stainless. If both sides of the tree return true/false, the output is buy/sell. Otherwise the signal is “no trade”. Regarding performance, the model outperform the hold and buy strategy, although surpassed in extremely bullish markets.

Huang, Chang, Kuo, Lin, Hsieh, & Chang in 2012 [19], studied the selection of assets based on fundamentals of initial public offerings (IPO). Initially, the model ranks the stocks using fundamental indicators which calculate the profitability of the company, the liquidity, leverage, efficiency, etc. The genetic search model implemented seeks the simultaneous optimization of the input features, sorting indicator and weights assigned to each fundamental indicator. In this work, the average return of all portfolio is used to evaluate the performance. The results across several cross-validation scenarios proved that the approach provides the investor above average returns.

2.4.2 Portfolio Related Work

Soleimani, Golmakani, & Salimi in 2009 [20], proposed a model based on Markowitz which takes in consideration cardinality constraints, transaction lots and regarding sector capitalization. Basically, the authors extended the Markowitz’s model to include a set of equations that represent several constraints, being that these equations must be verified at all time, triggering mechanisms otherwise. The interest in this solution goes mainly toward the sector capitalization. The authors used an unusual portfolio constraint, the sector capitalization, which means that the stock picking is not only based on its value and projections, but also on the sector where it is inserted. Using the genetic algorithms, the weights of each stock are calculated and are related with the sector capitalization. The results where compared with the results in LINGO, an algorithm that can generate global optima, using constraints such as the limit number of stocks per portfolio. The results showed average return and variation of 11% and 8.3% respectively. The differences between the proposed GA and the LINGO solution are below 3%, which is a very good performance. For large scale problems, the results were poorer, although satisfactory, presenting average returns and variation of 7% and 15% respectively.

Anagnostopoulos & Mamanis in 2011 [21], similarly to the previous work, conducted an evaluation of five different MOEA considering the cardinality constrained portfolio problem. They named their model as mean-variance cardinality portfolio optimization (MVCCPO), and justified the approach with the objective of avoiding extremal situations when choosing the assets, as an example, avoid excessively investing in a single asset. In their study the focus was to verify if any advantages can be extracted from the bi-objective cardinality and test it against the existing solutions of MOEAs on the MVCCPO model, being the first to prove their effectiveness. To achieve that goal, the authors tested five MOEAs: SPEA2, NSGA-II, PESA, NPGA2 and e-MOEA. Also, for comparison purpose, a single objective evolutionary algorithm (SOEA) was included in the study. As performance measures, the chosen approach calculates a linear normalization of the preferences generated bye the algorithms. The results showed that the SPEA2 performed best in 75% of the tests made, although it had the greatest computational cost. As conclusion, all MOEAs performed better than the SOEA. SPEA2, NSGA-II and e-MOEA were the best techniques, but having to choose one, SPEA2 overcomes the others.

Sadjadi, Gharakhani, & Safari in 2011 [22], again looked into optimization for cardinality constrained
portfolios. As previous works, the authors extended the Markowitz’s model to support some obligatory constraints, which results on a more realistic model, but of course, harder to solve. The constraints applied are the usual constraints for this type of approach (verified in several solutions across the state of the art for this problem): limit the number of holding assets, consider the return for each asset as uncertain, limit with upper and under bonds the investments in each asset, and no short selling allowed. Having all these constraints, a GA-based solution procedure was developed to find a optimal, or at least good solutions, that verifies those constraints. In the GA, the selection mechanism used was the roulette wheel and the uniform selection, ensuring the diversity and improving the quality of the final solution. As performance measure, they use relative percentage deviation (RPD), which calculates the proximity between the optimal solution and the one generated by the GA. Results showed that there is not significant difference between nominal and robust methods regarding returns. Also that, given four different sets of values for the constraints, when the constraints hold with probabilities of 99%, there is no much difference on the returns obtained.

Lwin, Qu, & Kendall in 2014 [23], used a learning-guided approach to solve the constrained portfolio optimization problem. The authors approach have similarities with the previous work, like the extension of the Markowitz’s model to support constraints. The innovation on this work is in the evolutionary algorithm. To start the simulation instead of randomly generating new candidates, a learning-guided solution is used. In practice this means that an external archive is kept throughout all iterations of the EA, storing a second population featuring some promising non-dominated solutions. The new individuals are composed with the assets stored on the solutions in the external archive. The authors considered in this work four MOEAs: NSGA-II, SPEA2, PESA-II and PAES. The performance evaluation compared the proposed model (MODE-wAwL) with the four MOEAs. Results showed the superiority of the model proposed in every aspects tested, it found solutions near the Pareto-front, with good distribution and more efficiently.

Silva, Neves, & Horta in 2014 [24], proposed a MOEA solution using both fundamental and technical indicators. Using a several of fundamental ratios as DR, ROE, PM, PER, RG, and others, the authors constructed a portfolio with the most promising assets. Further, based on technical indicators, decisions were made concerning the market choose, the timing to entry and exit, stop loss triggers, and some portfolio typical constraints. Using a MOEA, the results obtained revealed that is a better investment technique than just following the index. They also conclude that the more valuable genes were the ROE and rate growth of net income. Other two models were tested, differing on the genes weights, which lead to conclude that the most important genes are the ROE and profit margin. This is a relevant previous work as it follows some ideas that will be part of the proposed solution in this thesis.

Mousavi, Esfahanipour, & Zarandi in 2014 [25], introduced an approach using multi-tree genetic programming to construct and manage a portfolio, this is, which assets to buy, when and how much. Similarly to the previous work of Esfahanipour & Mousavi, in this work the rules are trinary, this is, the signal encoded on the rules can be buy, sell or “no trade”. For comparison purposes, the authors use three other approaches: a GA solution, a GP solution and the Buy and Hold strategy. The metrics are the Rate of Return (ROR) and the conditional Sharpe Ratio (CSharp). Two different experiments were conducted, the first in the Tehran Stock Exchange and the second on the Toronto Stock Exchange. The proposed multi-tree GP overcomes the GP existing solutions. Once again, the advantages of the trinary rules system are reflected in good rates of return and high risk adjusted returns.

2.4.3 Neural Networks, Fuzzy Systems, Others

Yeh, Lien, & Tsai in 2010 [26], described a system using evolutionary computation with neural networks (GNN). The system used the reinforced learning to develop the buy-sell decision directly. The authors express some concerns about the robustness of the system, this is, even if a certain individual performed very well in the training period, it may not do so on the testing. To try to avoid this situation, the reproduction probability of the individual is only based on his profit during training. The best survive. If the model produced individuals which performed as good in testing as in training, the model is said to achieve the generalization at the model level. Using GNN with this purpose, combined with evolutionary computation and technical indicators such as RSI, PVI, PVC, MVI, etc, showed promising results. However, when compared with the Buy and Hold strategy, the GNN performed significantly poorer, which compromise the classification of the proposed solution. Results also proved that using long or short term indicators does not have significant impact on the performance.
Cheng, Chen, & Wei in 2010 [27], proposed an hybrid model. The model used rough sets (RST) mixed with GAs, with the objective of stock price forecasting. The main idea is to use the RST algorithm to extract rules from the technical indicator data set, proceeding to refine the rules with the GA. The advantages stated by the authors refer that with a RS algorithm rules can be extracted from complex stock data, presented as simple decision rules of natural language and without needing statistical information. In their work the authors described the system architecture along with the several modules developed. In a short overview, the system consists of three major phases, with two processes each: the pre-process phase, where the aliasing of data is performed using CPDA and MEPA methods; the rule generating and refining module, where the RST and GA take action; and the performance evaluation module. Results suggest that the hybrid model performs superiorly against RST or GA algorithms alone and the Buy and Hold strategy. Also, the model performs the best when the market is extreme, this is, highly bullish or bearish.

Manahov, Hudson, & Linsley in 2014 [28], performed a study comparing small and large stocks using Strongly Typed Genetic Programming (STGP), a specific technique build on GP. The advantage aimed when using STGP is the fact that this algorithm reduces the search space by defining a type for each constant and each variable. The use of the GP proceeds normally, similarly to the solutions discussed above. The main interest for me in this work is the comparison between small-cap and large-cap stocks. The results are based on five different measures of forecasting accuracy, MAPE, MAE, RMSE, MaxAE and MaxAPE. All five showed that the proposed model is superior to the random walk forecast, either in-sample or out-of-sample experiments for all three markets. Besides this results, the authors alert to the fact that these measures do not imply profitability. Although, the performance results were positive, being that the small-cap stocks reported the highest hit ratio, 58.4%. This suggests that the returns gained from small-cap stocks are higher than the returns from large-cap stocks.

Mabu, Hirasawa, Obayashi, & Kuremoto in 2013 [29], developed a system using a mechanism of rule-based Genetic Network Programming with rule accumulation (GNP-RA). In a GNP solution, in each iteration of the algorithm, executes judgement and processing nodes, going from one to another sequentially depending on the output. This passing from node to node disable the decision making based one just a single action, instead it goes to the next node. The proposed solution, GNP-RA, extracts and stores the best rules obtained in each iteration. Rules extracted more than one time, are considered reliable and used further. The proposed solution is also mixed with reinforcement learning techniques to enhance the search and extraction of rules. Results showed that the model obtained the best average profit compared to the Buy and Hold strategy and to the model GNP-Sarsa. Authors concluded that this combinations of techniques work very well. Besides, GNP-RA avoid large losses by exiting before the down trends. thought it cannot profit so much on up trends.

Rather, Agarwal, & Sastry in 2014 [30], implemented a hybrid model for prediction of stock returns, using a recurrent neural network (RNN). The proposed model, called hybrid prediction module (HPM), tries to combine predictions obtained through different approaches, linear models and non-linear models. The authors objective is to develop a system robust and able to predict the non linear patterns, such as jumps in data. To achieve this goal, the HPM model combines three prediction models: exponential smoothness (ES), the linear model; autoregressive moving reference (ARMR) ; and the RNN predictions. To test predictive performance, the metrics chosen were MAE and MSE, using also the correlations between the targeted returns and the actual returns. The linear models performed the worst. RNN models produce satisfactory predictions and presented good correlations, although, it could not predict the sudden jumps. The proposed model performed the best, prediction error was the lowest and the correlation the highest achieved. That being said, authors stated that this performance is highly due to the RNN.

Hadavandi, Shavandi, & Ghanbari in 2010 [31], used a different approach of the artificial intelligent domain, the genetic fuzzy systems (GFS), combined with artificial neural networks (ANN) to forecast stock prices. This approach is said to be the first of his type, no literature existed using GFS and GAs. A hybrid method called clustering-genetic fuzzy systems (CGFS) is presented. This model can be partitioned in three major steps: the variable selection stage; the division of data stage, to reduce complexity; and the GFS stage, for the stock price prediction. The idea behind the GFS stage is to transform the fuzzy rules generated in chromosomes that can further be optimized using GAs. To measure the forecasting accuracy, the MAPE statistic evaluation was used. The proposed approach obtained better accuracy forecasting the prices for all three cases tested.

Östermark in 2012 [32], proposed a solution to incorporate asset growth potential and some safety switches to prevent losses in bear markets. The work of Östermark has special relevance as this thesis’ goals are
concerned with growth stocks. The main idea presented here is to create a system that recognizes growth potential and minimize the risk of buying this kind of assets with fuzzy safety switches. Assets with high growth potential, usually come with a high risk associated. The safety switches would reduce the portfolio overall risk. The authors proved that systems with this characteristics can effectively outperform the existing solutions. The Growth Potential Index (GPI) and the switches identified favourable entry/exit points.

In tables 1, 2 and 3 is exposed a summary of the related work discussed above, where the principal aspects of the solutions are presented.

2.5 Chapter Conclusion

In this chapter was presented a review of the necessary concepts for the reader to understand the work done. These concepts approach the two worlds combined in this work, the financial and the software world. Starts with the explanation of the several ways to analyse the market and investing strategies: fundamental analysis, the study of companies’ financial statements, this is, their revenues, debts, assets and others; qualitative analysis, which looks to factors as management, company’s philosophy and business models; and technical analysis, the strategy using charts to predict the market trends and movements. Regarding the investing strategies, it was explained how investors can value investing and growth investing. Next, the sector distribution theory was described. How companies are distributed by sectors and how/why is it possible to use this distribution to make high profits.

Several fundamental ratios used to forecast the market are exposed in this chapter, providing some highlights about how this ratios are calculated and can be employed in the creation of investing strategies and decisions. The reader can also find in this chapter the description of Markowitz’s Modern Portfolio Theory, which is a widely used model to calculate the risk associated to an investment.

The chapter continues to the evolutionary computing section, where the evolutionary and genetic algorithms are explained.

The chapter is closed with an extensive exposition of several previous works, which are briefly summarized in the tables 1, 2, and 3. From the present existing solutions is possible to conclude that genetic algorithm and economics are not strange to each other. Over the last years a lot of researchers tried to employ evolutionary strategies to predict the stock market. Many different optimization techniques were explored, like genetic algorithms, neural networks, fuzzy systems and vector machine learning. Although a fair comparison is a difficult task, it is possible to conclude that using evolutionary techniques the results surpass the competitor methodologies. These previous works results come as a motivation for this work, as the results are very promising.
<table>
<thead>
<tr>
<th>Reference</th>
<th>Period</th>
<th>Markets</th>
<th>Algorithms</th>
<th>Fitness Functions</th>
<th>Portfolio Analysis</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hassan, Clack (2009)</td>
<td>May 1999 to Dec 2005</td>
<td>FTSE100</td>
<td>SPEA2</td>
<td>ROI</td>
<td>Markowitz’s Model</td>
<td>The model presented improved robustness in out-of-sample environments, presenting very positive returns</td>
</tr>
<tr>
<td>Kaucic (2012)</td>
<td>25 Jan 2006 to 19 Jul 2011</td>
<td>DJI</td>
<td>GA</td>
<td>reward-to-risk</td>
<td>Technical Analysis</td>
<td>The three evolved portfolios proposed largely surpassed the DJI index</td>
</tr>
<tr>
<td>Esfahanipour, Mousavi (2011)</td>
<td>22 Aug 2004 to 21 Aug 2008</td>
<td>TSE</td>
<td>GP</td>
<td>Risk Adjusted Return using Sharpe Ratio</td>
<td>Technical Analysis</td>
<td>Outperformed the Buy and Hold model. 10.85%</td>
</tr>
<tr>
<td>Huang, Chang, Kuo, Lin, Hsieh, Chang (2012)</td>
<td>2005 to 2011</td>
<td>SFI</td>
<td>GA applied to search model</td>
<td>Mean Fundamental Analysis</td>
<td>The model showed above average for first day returns</td>
<td></td>
</tr>
<tr>
<td>Gorgulho, Neves, Horta (2011)</td>
<td>6 Jan 2003 to 6 Jan 2009</td>
<td>DJI</td>
<td>GA</td>
<td>ROI</td>
<td>Technical trading module</td>
<td>60%</td>
</tr>
</tbody>
</table>

Table 1: Genetic Algorithms Related Work
<table>
<thead>
<tr>
<th>Reference</th>
<th>Period</th>
<th>Markets</th>
<th>Algorithms</th>
<th>Fitness Functions</th>
<th>Portfolio Analysis</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soleimani, Golmakani, Salimi</td>
<td>NA</td>
<td>Hang Seng, DAX100, FTSE100, S&amp;P500, NASDAQ</td>
<td>GA, LINGO</td>
<td>RAR</td>
<td>Markowitz’s Model</td>
<td>11% for small scale problems, 7% for large scale problems, with 3.5% of risk</td>
</tr>
<tr>
<td>Anagnostopoulos, Mamanis</td>
<td>NA</td>
<td>Hang Seng, DAX100, FTSE100, S&amp;P500, NASDAQ</td>
<td>MOEAs (SPEA2, NSGA-II, PESA, NPGA2, e-MOEA)</td>
<td>Mean, Variance</td>
<td>Markowitz’s Model</td>
<td>SPEA2 performed better among the five MOEAs and the SOEA tested</td>
</tr>
<tr>
<td>Sadjadi, Gharakhani, Safari</td>
<td>NA</td>
<td>Hang Seng, DAX100, FTSE100, S&amp;P100, Nikkei225</td>
<td>GA</td>
<td>Mean, Variance, RPD</td>
<td>Markowitz’s Model</td>
<td>No significantly changes in returns by portfolios with different constraints when these hold with probability of 99%</td>
</tr>
<tr>
<td>Lwin, Qu, Kendall</td>
<td>Mar 1992 to Sep 1997</td>
<td>NA</td>
<td>MODEwAwL, NSGA-II, SPEA, PESA-II, PAES</td>
<td>Inverted Generational Distance, Generational Distance, Diversity, Hypervolume</td>
<td>Markowitz’s Model</td>
<td>Best performance obtained with MODEwAwL</td>
</tr>
<tr>
<td>Silva, Neves, Horta</td>
<td>17 Jun 2010 to 7 Feb 2014</td>
<td>S&amp;P500</td>
<td>MOEAs</td>
<td>Mean, Variance</td>
<td>Fundamental and Technical Analysis</td>
<td>The proposed models obtained best results than the S&amp;P500 index</td>
</tr>
<tr>
<td>Mousavi, Esfahani, Zarandi</td>
<td>20 Mar 2009 to 21 Sep 2011</td>
<td>TEFIX30, Tehran Stock Exchange, Toronto Stock Exchange</td>
<td>Multi-tree GP</td>
<td>Cumulative risk adjusted</td>
<td></td>
<td>The model overcomes the existing GP and GA solutions, as well as the B&amp;H strategy.</td>
</tr>
</tbody>
</table>

Table 2: Portfolio Solutions Related Work
<table>
<thead>
<tr>
<th>Reference</th>
<th>Period</th>
<th>Markets</th>
<th>Algorithms</th>
<th>Fitness Functions</th>
<th>Portfolio Analysis</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yeh, Lien, Tsai (2010)</td>
<td>5 Aug 1989 to 12 July 2004</td>
<td>TAIEX</td>
<td>GNN</td>
<td>profit</td>
<td>Technical Analysis</td>
<td>Annual profitability of 8.3%, although outperformed by the Buy and Hold strategy</td>
</tr>
<tr>
<td>Cheng, Chen and Wei (2010)</td>
<td>23 Jun 1999 to 11 May 2000</td>
<td>TAIEX</td>
<td>GA, RST</td>
<td>percentage of correct observations by the chromosome</td>
<td>Technical Indicators</td>
<td>Better performance than RST or GA alone and Buy and Hold strategy</td>
</tr>
<tr>
<td>Manahov, Hudson, Linsley (2014)</td>
<td>2t May 1979 to 16 Nov 2012</td>
<td>Russel</td>
<td>STGP</td>
<td>Total Profit</td>
<td>Technical Analysis</td>
<td>The model produced higher returns than random walk forecast and the Buy and Hold strategy. Better performance for small-cap stocks</td>
</tr>
<tr>
<td>Mabu, Hirasawa, Obayashi, Kuremoto (2013)</td>
<td>4 Jan 2006 to 30 Dec 2009</td>
<td>Tokyo Exchange</td>
<td>GNP-RA</td>
<td></td>
<td></td>
<td>Produced higher profits than B&amp;H strategy</td>
</tr>
<tr>
<td>Rather, Agarwal, Sastry (2014)</td>
<td>2 Jan 2007 to 22 Mar 2010</td>
<td>NSE</td>
<td>todo</td>
<td>Mean, Standard Deviation</td>
<td></td>
<td>The HPM presented low prediction error and high correlation between the expected returns and the returns obtained</td>
</tr>
<tr>
<td>Hadavandi, Shavandi, Ghanbari (2010)</td>
<td>17 Sep 2002 to 17 Mar 2005 to</td>
<td>Stocks of IBM, Dell, British Airlines and Ryanair</td>
<td>GFS, GA</td>
<td>Mean Square Error</td>
<td></td>
<td>Improved prediction of prices in all cases tested</td>
</tr>
<tr>
<td>Östermark (2012)</td>
<td>Dec 1998 to Dec 2008</td>
<td>NA</td>
<td>GHA</td>
<td>GPI</td>
<td></td>
<td>The model outperformed the B&amp;H strategy. The GPI improved the performance of the portfolio</td>
</tr>
</tbody>
</table>

Table 3: Neural Networks, Fuzzy systems and other solutions Related Work
3 Solution’s Architecture

This chapter describes a system that automatically creates an investment strategy using evolutionary computation. This section starts with an overview of the system (section 3.1), followed by the architecture of the solution (section 3.2) and data flow (section 3.3), proceeding to a more detailed view of each module (section 3.4 to 3.7), and finally a brief conclusion of the chapter (section 3.8). Besides, this chapter provides a description of the new ideas introduced and implemented in this work.

3.1 System’s Overview

The proposed system attends two major problems. (1) The selection of promising growth stocks; (2) make the decision whether to buy or sell over time. The approach to solve the given problems was found on the branch of artificial intelligence, the so called Evolutionary Algorithms (EA). Two different operations are contemplated. The first is in charge of the filtration of stocks, electing the companies that exhibit a greater growth potential in each sector on a daily basis. The second decides when to buy/sell and the percentage of each asset on the portfolio. Both operations are based on fundamental analysis and optimized using an EA. That being said, this project proposes four new methods focused on maximizing returns:

1. Combine fundamental analysis of both companies and industries with Evolutionary Algorithms.
2. Stock selection based on sectors. Pick the best companies from each sector, instead of looking only to the overall market.
3. Explore several indicators based on sector analysis.
4. Explore new indicators based on fundamental analysis.

3.2 Architecture

In order to achieve the goals of this work, the system must handle several aspects and perform a set of steps to construct a portfolio that conforms to the expectations. The called steps are represented as modules, each of them aggregating actions. Each of these modules have a distinct objective and functional responsibility, creating together the system. Next, the modules specification is presented, along with several diagrams representing each module and interaction between them.

From the image 9, the reader can observe that the proposed system is made of six major modules, distributed in three layers:

**User Module**: Module oriented to the interaction with the user. Through this module, user can specify some constraints and parameters needed.

**Picking Module**: The most important module of the system. This module is responsible for the companies selection. Based on their fundamentals, this module has the objective of ranking the companies considering their growth potential and their sector.

**Portfolio Simulator Module**: In charge with the simulation of a real portfolio, both creation and management, including buying and selling operations.

**Data Processing Module**: Perform all data collection and storage. The collection regards the stock data prices, the company’s fundamental statements and the sector’s indicators.

**Fundamental Ratios Module**: Specify the set of fundamental ratios used to select the most promising companies.

**Technical Module**: Specify the technical indicators used to analyse the market charts.
3.3 Data Flow

Regarding the information transfer from module to module within the system, seven main steps are proposed (see [10]). One complete execution of the model should act like this:

1. The user provides the system with his preferences and constraints, the budget and investing strategies for example.

2. Next, the system starts operating, calculating a set of fundamental ratios with the data gathered from companies in S&P 500 to understand their profitability, liquidity, debt, etc. The sectors performance is also calculated and stored during this phase.

3. Having the fundamental information, the so called Picking GA starts its execution, generating several chromosomes (individuals) with different weights for each fundamental ratio obtained in the previous step.

4. Each individual is submitted to a simulation using the portfolio simulator for a quality check. During the training period there is an evolution process. The survivals, this is, the stocks presenting higher returns, are selected to constitute the new generations.

5. Once the training period is over, the individuals are sorted by performance. The best individuals are selected for a real life simulation (testing period).

6. The system runs again the portfolio simulator with an out-of-sample period to measure the performance of the individuals in an unseen environment.

7. The results are delivered to the user.
3.4 Data Processing Module

This module is responsible for every action regarding data. Inside this module, are considered two main actions: (1) the download of historical data available from the S&P 500 and gather the financial statements of the companies over the last years. In this work, all tests are performed with S&P 500’s data, although the system is implemented so that other markets may be used in a most simple way. (2) The process of cleaning and storing the data. The available data is not formatted as wished. In order to use it, cleaning and transforming mechanisms were applied until the data is stored in the desired format.

3.4.1 Implementation and Functionality

The Data Processing Module is composed by four pieces of software. (1) Edgar 2.0, (2) Excel Converter, (3) Normalize Edgar and (4) Statements Standardizer. In the figure [11] are represented the inputs and outputs of each software, along with the interaction between them.
1. **Edgar 2.0**: The download robots used to extract all necessary files from the web.

2. **Excel Converter**: Software in charge of converting all excel files to a unique format, in this case, .xlsx.

3. **Normalize Edgar**: Software responsible for cleaning and sorting the excel files, selecting only the important sheets.

4. **Statements Standardizer**: Last sub-module, which produces the financial statements with the desired format, making it equal to all companies.

In the next sections, each sub-module is described and explained in detail, giving the reader full understanding of their functionality and how they are implemented.

### 3.4.1.1 Edgar 2.0

*Edgar 2.0* is the name of the first software sub-module. Its function is to download the data from the web. This sub-module was coded using *Python* and consists of two different programs. (1) *Financial Statements Robot* and (2) *Stock Prices Robot.*
• **Financial Statements Robot:** As the name suggests, this first robot was used to download all the financial statements available from the S&P 500’s companies, ranged between 2010 and 2015. These files are free to the public through the American system called “Edgar”, available at [www.sec.gov](http://www.sec.gov). “Edgar” is the name given to the system where companies are required to publish their fundamentals every trimester for the last decades, although the downloadable files only go back to 2010. In the figure [12] is represented the flow of the algorithm. The process can be described in three steps:

1. Specify the companies’ tickers (unique group of letters representing a particular company) along with their unique numeric code, in two separated text files.
2. After receiving the targeted securities, the developer needs to indicate the starting data period, in this case, 2009 (which means, download the files from 2009 until today’s date).
3. Finally the download process is executed. This process comprehends both download and storage functionalities. The robot automatically runs all the url paths for each company and saves as Excel files (.xls or .xlsx, depending on the Excel version) all the available documents for the company on the given period. In the end of the process, the result is a directory containing a sub-directory for each company, which contains six sub-directories for each year, and finally the excel files (i.e Google’s statements for 2010, `Financial_Reports_by_Edgar2.0\GOOG\2010\000119312510108704.xls`).

![Figure 12: Edgar 2.0 Flow Diagram](image)

• **Stock Prices Robot:** The Stock Prices Robot is a simpler robot, responsible for collecting the historical stocks’ prices. Similar to the first robot, the data retrieved is from the S&P 500’s companies, ranged between 2010 and 2015. Nonetheless, the program is prepared to operate with any company and any period of time. The data is downloaded trough the [Yahoo Finance](https://finance.yahoo.com) website, following three identical steps:

1. Specify the companies’ tickers in a text file.
2. The developer specifies both start and end date, in this case, 2009 and 2015 respectively.
3. The download process is executed. In this case, the result from the storage functionality is a csv file for each company, with the desired financial data (i.e `Stock_Prices_by_Edgar2.0\GOOG.csv`). Each file contains daily records accordingly with the configuration represented on figure [13].
Where:
- Date: The corresponding day, using the format “yyyy-mm-dd”.
- Open: The opening price at which the stock was traded at a specific date.
- High: The highest price at which the stock was traded at a specific date.
- Low: The lowest price at which the stock was traded at a specific date.
- Close: The closing price at which the stock was traded at a specific date.
- Volume: The number of shares traded for the given security during a specific date.
- Adj. Close: The adjusted closing price for the given security at a specific date.

3.4.1.2 Excel Converter 2.0

Once the financial statements are stored, to build a software able to process all the files in an easy and smooth way, it was necessary to convert all the Excel files to a single Excel version. As stated before, some companies had their statements as files .xls, and others used the .xlsx format. The difference in the extensions is caused by the version of the Excel used, this is, if the company is using Excel 2007, the files are saved as .xls. On the other hand, if the version used was the 2013, files come with the extension .xlsx. Anyway, the objective was having all the files using the same version. For this work, we chose Excel 2013. The Excel Converter 2.0 is the second sub-module created for data processing. It is a single file program, coded using the programming language Visual Basic. The choice of the language was motivated by the many primitive functions to deal with Microsoft Office that are already available to the developer. The execution process can be described in two simple steps:

1. Iterate throughout the companies’ directories and open each excel file, one at a time.
2. If the current open document is in the old format, .xls, call the operation “Save As” and store the document as a new file, with the same name and the desired extension, .xlsx. Otherwise, it closes the file and proceed to the next one.

At this time, the data stored consists in thousands of Excel files, 2013 version, distributed by many directories, following the format stated above, this is, Financial_Reports_Formated\GOOG\2010\000119312510108704. xls). The directories’ tree is represented in figure [4].
3.4.1.3 Normalize Edgar

The sub-module *Normalize Edgar* is the third one concerned with data. The purpose of this sub-module is to facilitate the further use of the financial statements. Despite the fact that all files are already distributed by company and year, as described in section 2, three different statements are published every trimester: (1) Income Statement, (2) Balance Sheet and (3) Cash Flow. At this point, the stored data consists of Excel files for each trimester, each of them containing a large number of pages (sheets). To gain access to the desired statements, a filtering and sorting process was executed. This sub-module is decomposed in five classes, implemented using *Java* language: (1) *Excel to CSV*, (2) *Converter*, (3) *Organizer*, (4) *Sorter*, (5) *File Categorizer*. In figure 14 a diagram is provided for the reader to understand the logical path and interaction between these classes. Along with the diagram, an individual description for each class is provided.

Figure 14: Directories disposition after the Excel Converter execution
Excel to CSV: As the reader can observe from the diagram, the Excel to CSV is responsible for the first phase of this sub-module. The functionality of this class is to iterate through the several directories, create a Converter CSV object and proceed to invoke the method convertFiles for each one of the Excel files.

Converter CSV: This class is intended to separate the sheets of the Excel files. As stated before, each Excel file contains a large amount of pages. 90% of them are irrelevant for this work. This second phase aims at creating one csv file for each Excel sheet. Instead of having one Excel document for each trimester, directories now contain a couple of hundred csv files, as the reader can see in figure 16.
Figure 16: Directories disposition after the 1st phase of the Normalized Edgar

- **Organizer CSV**: Similar to the Excel to CSV program, this class iterates the companies directories, now containing csv files instead of the old Excel files. For each file, the execution follows two steps: (1) check the file type and (2) copy the file to the right directory. Both steps are responsibility of the Sorter CSV class, explained next.

- **Sorter CSV**: This class has two main procedures: (1) Find the type of file it receives, (2) store it on the right directory. To achieve the first goal, an instance of File Categorizer is used. Using the files’ name its possible to know if the file corresponds to either an Income Statement, a Balance Sheet, a Cash Flow or if it is an irrelevant file (for this work of course). Having the type of file, this class invokes its sort method to store the file on the right company, year and type.

- **File Categorizer**: Auxiliary class used by the Sorter CSV class to categorize files. Given a file’s name, it looks for key words to assign the type. Four types are considered: (1) Income Statement, (2) Balance Sheet, (3) Cash Flow and (4) Irrelevant.

At this point, the data stored is cleaned and sorted, this is, every company have three files, representing the different financial statements, for each trimester. The figure 17 shows the directories tree of the stored data.
3.4.1.4 Statements Standardizer

To complete this module, one last problem needs to be taken care of. Despite the fact that “Edgar” is a platform of national dimension, supported and managed by the American government, there is no template or rules for the companies to follow. There are no further obligations besides presenting their reports for the trimester. Granting all this, a major problem appeared. Each company has its own format for the financial statements, its own names for the fields constituting the statements. This means that, for example, when in need to know the total revenue that a given company declared for a specific trimester, the software needs to know that company X names it as “Revenue” but company Y chose to call it “Sales”. To solve this (big) problem, a sub-module named Statements Standardizer was developed. The goal of this sub-module is to “absorb” all csv files for each company and produce three new files, one for each financial statement, written in the format used by Google on their financial platform, Google Finance. This sub-module is the most important sub-module of the Data Processing Module, being also the more complex to explain. Next, some diagrams are presented to illustrate the functionality of this module, along with a description for each class.

Figure 17: Directories disposition in the end of the Normalize Edgar process
• **Statement Interpreter**: The main class of this sub-module is the “Statement Interpreter”, responsible for managing the files until they are on the correct format. In figure 19, the reader can observe the flow created when this class is executed. In the first phase of the algorithm, for each new company, the three statements are initialized. During the process, these three objects are continuously altered, being exported as csv files in the end. Next, one by one, the files related to the company are imported and analysed. Initially, the program seeks for the semesters covered in the document, as well as the release date to the general public. To conclude the first phase, the currency in use is also saved. Having the trimesters and currency, the second phase of the algorithm begins. In this phase, called the Interpret phase, each line of the document is processed. Using the Statement Categorizer, the search process is executed and hopefully it returns the field’s name used on the Google Finance’s template. Having encountered a name, the algorithm proceeds to insert the values on the respective statement object. The insertion process comprehends two steps: (1) Format the currency. All data is meant to be in millions, so depending on the currency used by the company, some calculus may be needed. (2) Insert the formatted value on the statement. Once all files were read, the finish() method is invoked for each statement object and the three output files are printed.
• **Statement Categorizer**: Equivalently to the File Categorizer from the Normalize Edgar sub-module, this is an auxiliary class used to determine which field from the Google Finance template corresponds
to the one present in the file. As stated, each company names their fields has they wish, so in order to classify every field of all companies, each field has assigned a list of possible synonyms. Given a random company’s field, this class looks up for the respective Google’s field on the lists concerned with the specific statement (to avoid irrelevant searches, each statement has their own lookup function). Notice that this is not a 100% match program. Some companies do not provide the necessary fields to complete Google’s templates and others may have go unnoticed, as some companies do not respect any kind of rules or patterns, changing the fields’ names every trimester.

- **Statement**: Abstract class implementing the common aspects to all the statements, more specifically, information regarding the currency in use, the trimesters and year to which data is related, the window used (three, six or nine months) and the general headers of the files.

- **Income Statement, Balance Sheet, Cash Flow**: This three classes are the three extensions of the Statement class, representing the three financial statements. All three follow a similar structure, differing only on the “fields” that each on holds. Each field is represented by an array, which is constantly updated with values until all slots are filled. Once it is complete, the finish method produce the output csv file with the desired format.

Having done the previous operations, the Data Processing Module is concluded. The data required for this work is now duly standardized, stored and ready to use for the next modules. In the figure 20 the reader can observe the final structure of the data directory. Notice that despite the fact that Statements Standardizer’s output are three files for each company, like it is represented in figure 20 for the next modules, these files were unified into one single file per company, for easier access. The final format is showed in figure 21.

![Figure 20: Final tree for data storing](image)
<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>In Millions of USD (except for per share items)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Company A - Balance Sheet</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Cash &amp; Equivalents</td>
<td>2450</td>
<td>2519</td>
<td>2330</td>
<td>2675</td>
<td>2742</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Short Term Investments</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Cash and Short Term Investments</td>
<td>2450</td>
<td>2519</td>
<td>2330</td>
<td>2675</td>
<td>2742</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Accounts Receivable - Trade</td>
<td>874</td>
<td>916</td>
<td>875</td>
<td>899</td>
<td>849</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Receivables - Other</td>
<td>-420</td>
<td>-468</td>
<td>-468</td>
<td>-467</td>
<td>-419</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Total Receivables</td>
<td>454</td>
<td>446</td>
<td>407</td>
<td>452</td>
<td>430</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Total Inventory</td>
<td>1040</td>
<td>1042</td>
<td>1054</td>
<td>1066</td>
<td>1088</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Total Current Assets</td>
<td>4712</td>
<td>4618</td>
<td>4584</td>
<td>4983</td>
<td>5073</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Property/Plant/Equipment</td>
<td>1163</td>
<td>1147</td>
<td>1139</td>
<td>1134</td>
<td>1129</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Goodwill</td>
<td>3071</td>
<td>3066</td>
<td>2995</td>
<td>3047</td>
<td>3017</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Net Intangibles</td>
<td>1069</td>
<td>995</td>
<td>945</td>
<td>916</td>
<td>859</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Long Term Investments</td>
<td>128</td>
<td>120</td>
<td>124</td>
<td>139</td>
<td>129</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Total Assets</td>
<td>10653</td>
<td>10587</td>
<td>10278</td>
<td>10686</td>
<td>10638</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Long Term Debt</td>
<td>2111</td>
<td>2106</td>
<td>2701</td>
<td>2699</td>
<td>2695</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Total Liabilities</td>
<td>5302</td>
<td>5279</td>
<td>5488</td>
<td>5597</td>
<td>5191</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Retained Earnings (Accumulated Deficit)</td>
<td>-41</td>
<td>-145</td>
<td>-173</td>
<td>91</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Total Equity</td>
<td>5351</td>
<td>5308</td>
<td>4790</td>
<td>5289</td>
<td>5447</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Total Liabilities &amp; Shareholders’ Equity</td>
<td>10653</td>
<td>10587</td>
<td>10278</td>
<td>10686</td>
<td>10638</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Total Common Shares Outstanding</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Company A - Cash Flow</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Cash from Operating Activities</td>
<td>245</td>
<td>315</td>
<td>215</td>
<td>377</td>
<td>194</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Cash from Investing Activities</td>
<td>-72</td>
<td>-71</td>
<td>-54</td>
<td>-51</td>
<td>-47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Net Issuance (Retirement) of Stock</td>
<td>-79</td>
<td>-140</td>
<td>-681</td>
<td>0</td>
<td>-100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Net Issuance (Retirement) of Debt</td>
<td>-</td>
<td>-</td>
<td>-5</td>
<td>0</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Cash from Financing Activities</td>
<td>-63</td>
<td>-165</td>
<td>-332</td>
<td>6</td>
<td>-68</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Net Change in Cash</td>
<td>99</td>
<td>69</td>
<td>-189</td>
<td>345</td>
<td>67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Company A - Income Statement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Revenue</td>
<td>1380</td>
<td>1424</td>
<td>1335</td>
<td>1395</td>
<td>1366</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>Other Revenue</td>
<td>300</td>
<td>308</td>
<td>317</td>
<td>323</td>
<td>313</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>Total Revenue</td>
<td>1680</td>
<td>1732</td>
<td>1652</td>
<td>1718</td>
<td>1679</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>Cost Revenue</td>
<td>637</td>
<td>674</td>
<td>625</td>
<td>640</td>
<td>626</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>Gross Profit</td>
<td>1043</td>
<td>1058</td>
<td>1027</td>
<td>1078</td>
<td>1053</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>General Expenses</td>
<td>484</td>
<td>497</td>
<td>449</td>
<td>450</td>
<td>488</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>Research &amp; Development</td>
<td>179</td>
<td>181</td>
<td>171</td>
<td>173</td>
<td>177</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>Depreciation/Amortization</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>Interest Expense - Net Operating</td>
<td>-25</td>
<td>-25</td>
<td>-27</td>
<td>-30</td>
<td>-29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>Other Operating Expenses</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>-3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>Total Operating Expenses</td>
<td>1463</td>
<td>1519</td>
<td>1416</td>
<td>1433</td>
<td>1461</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>Operating Income</td>
<td>217</td>
<td>213</td>
<td>236</td>
<td>285</td>
<td>218</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>Interest Income - Net Non Operating</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>Gain on Sale of Assets</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>Income Before Tax</td>
<td>195</td>
<td>198</td>
<td>212</td>
<td>254</td>
<td>191</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>Income After Tax</td>
<td>179</td>
<td>166</td>
<td>168</td>
<td>724</td>
<td>195</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>Net Income</td>
<td>179</td>
<td>166</td>
<td>168</td>
<td>724</td>
<td>195</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>Diluted Weighted Average Shares</td>
<td>3.52E-04</td>
<td>3.49E-04</td>
<td>3.43E-04</td>
<td>3.45E-04</td>
<td>3.38E-04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>Diluted EPS Excluding Extraordinary Items</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>Diluted EPS Including Extraordinary Items</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>Dividends per Share</td>
<td>0.22</td>
<td>0</td>
<td>0.12</td>
<td>0.12</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>Diluted Normalized EPS</td>
<td>0.51</td>
<td>0.48</td>
<td>0.49</td>
<td>0.61</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 21: Example of financial statement in the final format
3.5 Fundamental Ratios Module

Although the GA module is considered to be the most important module of the system, without the proper ratios, it is worthless. The picking of stocks based on fundamental ratios has been object of several studies over the years, being once again used in this solution. This module specifies several ratios that can be used in the GA to evaluate quantitatively a certain company. The objective is to determine the ratios that better suit the model, this is, that more accurately find highly growing companies. As stated already in the previous chapter, these ratios consist in mathematical formulas applied to the companies’ results. Of course, there are hundreds of different ratios, being one of the major decisions made in this work choose the right ratios to combine and maximize the results.

In this work, several ratios were used. As mentioned, there is no right ratio, a better winning ratio. The probability of having a false signal when using just one ratio e very high. Having said that, for this work, taking into account the final objective of finding growing companies, 12 ratios were selected and 2 new ratios were created. These ratios can be distributed in four categories:

- **Profitability Ratios:** Return On Equity, Return On Assets, Profit Margin, Revenue Rate, Revenue Growth Rate, Net Income Growth Rate
- **Liquidity Ratios:** Current Ratio
- **Leverage Ratios:** Debt Ratio, Debt/Equity Ratio
- **Market Value Ratios:** Earnings Per Share, Price to Earnings, Price to Earnings Growth, Price to Earnings Future, Price to Book Value

After its calculus, to each ratio, given its value, a score is assign. This is meant to normalize all values and, this way, simplify the mix of the several ratios, ranking each company with a normalized data set. Five distinct scores were used:

- Excellent: 5
- Good: 4
- Normal: 3
- Bad: 2
- Very Bad: 1

Each ratio will be addressed in the following section.

3.5.1 Implementation and Functionality

The *Fundamental Ratio Module’s* main objective is to calculate all the ratios for the entire period on which the software is running, avoiding further access to external files. In the figure [22] the reader can observe the structure of this module. In this work we invest on the S&P 500, which means there are 500 possible companies to buy. It would be of great computational cost to calculate the 14 ratios for each company for each day of simulation, as it would require many external files loading and reading. To solve this problem, at the beginning of the simulation, all files created by the *Statements Standardizer* module are loaded, the ratios calculated and saved in serialized objects. This way, throughout the simulation, to access the values, a simple get() method can be used, which is extremely faster than accessing external files. Below, each class is explained and all process is clarified.
Figure 22: Fundamental Ratios Module class diagram

- **Ratios Manager:** This is the main class of this module. The *Ratios Manager* class execution consists in two main activities: (1) complete the companies’ map and (2) complete the sector’s map. These two maps can be seen has two big boxes, each of them containing many smaller boxes representing companies in the first case, and sectors in the second case. That is the abstract idea, the manager creates boxes and fills them with the information extracted from the external files, making them accessible for the rest of the software execution in a simple way.

  1. For the first case, two steps are consider: (1) loading the files, this is, the Income Statement, the Balance Sheet and the Cash Flow regarding the company being address at each moment; (2) store the “box”, represented by the *Company* class, which is produced by the *Ratios Calculator* class. The functionality of the calculator class is explained on the next item.

  2. For the second case, the contemplated steps are similar: (1) loading the files concerning the sectors, this is, the information regarding the revenues and earnings per share of each sector for the last six years; (2) create a “box” for each sector, represented by the *Sector* class, load it with the respective data, and store it. The sector management is entirely concern of the *Ratios Manager* class.

- **Ratios Calculator:** This class is responsible for the true meaningful functionality of this module, which is, calculate the fundamental ratios for each company and stored them in accessible *Java* objects. To achieve this goal, the class follows three steps (for each company). First, it creates a new *Company* object. Second, it runs the files related to the company on a search function which is in charge of extracting all the relevant fields to used in this project. Having the relevant fields available, the third step begins: calculate each ratio and store the result on the *Company* object. Each ratio is calculated with a simple mathematical function that accesses the necessary fields for its calculus through the object representing the specific ratio. To illustrate this functionality, an example is provided. Supposing the
program is trying to calculate the Return on Assets for a given company. First, the total assets and
the net income are extracted. Second, with the Return On Assets class, the ratio is calculated. Finally,
the returned value is stored on the company and made accessible to the manager.

- **Values Normalizer**: Auxiliary class that normalizes all ratios results. Each fundamental ratio has
  associated a list of value range identifiers, translating the ratios’ values to a value between 0 and 5.

- **Sector**: Class used to define a sector of the S&P 500 index. Each sector is characterized only by its
  unique name (i.e. Health Care). The functionality of this class is very straightforward: has four lists
  filled up with quarterly figures, representing the revenue, the revenue growth, the earnings per share
  and the earnings per share growth. Those values are used to rank the sectors on a quarterly basis. The
  selection of which list to use when ranking the sectors is related with the strategy employed to invest,
  and will be explained further in this document, on the section regarding the portfolio simulator. In [23]
  the reader can see an illustration of the sector object.

![Sector Object Illustration](image)

- **Company**: Class used to define a company. Each company is characterized by a name (i.e. GOOG)
  and a sector (i.e. Health Care). The functionality of the Company class is very related with the
  functionality of the Sector class. The difference is that instead of four lists, the Company class has
  a list for each fundamental ratio and one for the shares’ prices. In figure [24] the reader can see an
  illustration of the company object.
**Fundamental Ratio:** Abstract class extended by all ratios classes. This class was created in order to make it possible to deal with all ratios in the same way, as they were the same.

**Return On Assets, etc:** Example of a ratio class. All ratios are implemented as similar classes, having only one functionality, the `calculate()` method, which consists of the mathematical formula for the given ratio.

In figure 25 the reader can see the flow diagram of the process described above. Next, an exposition of the ratios used in this work is provided. A simple description of the ratio is provided, along with the motive why it is important and why it was selected for this work among the hundreds available. Notice that some of them already were indicated and explained on the section regarding fundamental ratios in the chapter 2, and therefore will not be mentioned again in this section. The two original rations are more exhaustively explained.
Figure 25: Fundamental Ratios Module data flow diagram
3.5.2 Net Income Growth Rate (NIGR)

The NIGR ratio represents the gain, positive or negative, from the last year net income. Through the calculation of the NIGR for periods of 5-10 years, the investor can have a good perception of the future net profits. To find companies with high growth potential, this ratio has a strong role to play. A NIGR ratio continuously growing at a stable rate for the last 5 years is a strong indicator of the company’s consistency and ability to produce profit, and so it is a very positive buy signal. Notice that NIGR ratio is independent from the real revenue, this is, companies may be generating large amounts of money every year, however if the profit is similar, this ratio will label the company as a “no buy”. This happens in more mature companies, that are not in their growing phase and have a tendency to slow down the profits growth, although providing a greater stability.

\[
Net\ Income\ Growth\ Rate\ (NIGR) = \frac{Net\ Income_{Current} - Net\ Income_{Last\ Year}}{Net\ Income_{Last\ Year}}
\]

(26)

3.5.3 Revenue Growth Rate (RGR)

The Revenue Growth Rate ratio is similar to the NIGR ratio. The difference is that in this case, the revenue is used instead of the net income. With the revenue information is possible to measure how fast a company is expanding and making money. The advantage of using this ratio is that with the NIGR ratio, investors only have a perception of the “clean money”, this is, the income after taxes and expenses. This may not be enough because some of the expenses may derive from investments made to expand the business. Calculating the RGR provides the investor with clear information about how much money is entering the company, due to sales, services or other company’s activities, and at which rate is it growing. For this work, the RGR is calculated from the last quarter revenue, instead of calculating it annually.

\[
Revenue\ Growth\ Rate\ (RGR) = \frac{Net\ Revenue_{Current} - Revenue_{Last\ Trimester}}{Revenue_{Last\ Trimester}}
\]

(27)

3.5.4 Revenue Rate (RR)

The Revenue Rate ratio is an original ratio created for this work. The main goal is to detect which companies are highly overvalued. The formula is the following:

\[
Revenue\ Rate\ (RR) = \frac{Stock\ Price \times Outstanding\ Shares}{Revenue}
\]

(28)

The fraction’s numerator represents the trading value of the company, this is, if all stocks were sold at the momentary price, how much money would be received. The denominator represents the revenue generated by the company. A RR value below 1 means that the company’s market value is lower than the money being generated by the company’s activity, which suggests that the company’s business market value may be undervalued and it will probably rise in the near future. On the other hand, if the RR value is too high, 10 for example, indicates that the company is selling 10 times over its worth in sales/revenue. In this situation, is mandatory to double check if the company has indeed revenue power to justify that value or if it is highly overvalued.

Assuming that the higher profit prospective a company can make a year is 20%, which is an hard reality even for the biggest corporations. A RR of 10, implies that the company’s market value is 10 times the revenue:

\[
Market\ Value\ (MV) = Stock\ Price \times Outstanding\ Shares
\]

\[
RR = \frac{MV}{Revenue} \land RR = 10
\]

\[
\equiv MV = 10 \times Revenue.
\]

Considering the PER formula and the restrictions:

47
\[\text{PER} = \frac{MV}{\text{Profit}} \land MV = 10 \times \text{Revenue}\land\]
\[\text{Profit} = 20\% \times \text{Revenue} \equiv \text{PER} = \frac{10 \times \text{Revenue}}{0.2 \times \text{Revenue}} = 50.\]

Conclusion, a RR of 10 implies a PER value of 50, which is a bad indicator and a strong signal that the company is overvalued. To keep up with a RR value of 10, companies’ profits should round the 50%, a very unlikely scenario.

That being said, the objective is to find companies with profit margins rounding the 20% and presenting RR values around 4-5, which indicates that the company is not overvalued and has a very good profit rate.

3.5.5 Price to Earnings Future (PER-F)

This ratio is the second original ratio created for this work. The objective of this ratio is to calculate the PER value 3 years from the day of its calculation.

\[\text{Price to Earnings Future (PER } - F) = \frac{\text{Stock Price}_{\text{Actual}}}{\text{EPS}_{3\text{ years}}}\]  

\[\text{EPS}_{3\text{ years}} = \frac{\text{Net Income}_{3\text{ years}}}{\text{Outstanding Shares}_{\text{Actual}}}\]

\[\text{PER} - F \equiv \frac{\text{Stock Price}_{\text{Actual}} \times \text{Outstanding Shares}_{\text{Actual}}}{\text{Net Income}_{3\text{ years}}}\]

Considering:
Market Value = Stock Price \times Outstanding Shares, the PER-F ratio is given by:

\[\text{PER} - F = \frac{\text{Market Value}_{\text{Actual}}}{\text{Net Income}_{3\text{ years}}}\]

The goal of this work is to find growing companies, and to achieve this goal, the PER ratio is very important and effective. Although, it may be misleading. By performing a simulation of the future, the investor can have an idea about the growing perspectives of the company. Also, is a good comparison metric between companies.

From the previous equations is possible to derive the basic idea of this ratio: assuming that the net income growth rate will be the same for the next 3 years, and assuming that the company’s market value is kept, what will be the PER value? For example, given two different companies, A and B. The market values are \(MV_A = 40\) \(\land\) \(MV_B = 40\). The net income is given by \(NI_A = 3\) \(\land\) \(NI_B = 4\). Calculating the PER with the previous values: \(PER_A = 13 \land PER_B = 10\). By analysing the current PER values, company A is more attractive.

But what are the perspectives for the next 3 years? Considering a current net income growth rate of \(NIGR_A = 20\%\) \(\land\) \(NIGR_B = 5\%\), and assuming that this values will be maintained for the next 3 years, the PER-F is: \(PER - F_A = 7.7 \land PER - F_B = 8.6\). Three years from now, under the same circumstances company B is slightly more attractive than company A.

3.6 Picking Module, The Genetic Algorithm

This module is the fundamental piece of this proposal. It is responsible for picking the most promising stocks among the available on S&P 500 and for the buy/sell decisions. As mentioned before, this module uses evolutionary computation to achieve its goals. The picking GA uses the fundamental ratios to choose the most suitable companies among the ones represented on S&P 500. Although, the picking has a particularity. In this work, an innovative way of picking companies will be tested. Instead of choosing the best companies from the general S&P 500 (for example), the companies are chosen by sector. As mentioned in sector 2, S&P 500 companies can be divided in ten sectors, each of them with different performance. In this scenario,
portfolio’s slots will be dynamically allocated, based on that performance. For example, supposing a portfolio with thirty slots, a normal approach would be choosing the top thirty companies to construct the portfolio. In this case, the best companies from each sector will be chosen, even if they are not in the top thirty. Being a dynamic approach, the number of slots destined to each sector is not static. If the performance of the sector is good, it may dispose of four or five slots, while the worst sectors only have right to one or two. Besides this previous brief explanation, further in this section, the several different strategies employed will be explained. Following this methodology, the risk of the portfolio can be reduced, minimizing the impact in case of financial collapse in a certain industry. In figure 26 an overview of the process is represented as a diagram. Further, all aspects of the algorithm are more carefully explained.
Figure 26: Picking Module Data Flow Overview
3.6.1 Chromosome Representation

The individuals, or chromosomes, are a simple array structure of real numbers. For the picking GA, the chromosome chosen has 19 genes. The genes indexed from 0 to 13 represent a weight given to a specific financial ratio, this is, the importance that the ratio will have on the decision whether to buy or not the stock. Besides the weights, assigned to the financial ratios, six more genes are present. These genes represent constraints necessary to the well function of the software, information regarding: the portfolio maximum size; bounds to define the minimum necessary rank an equity needs to obtain to enter or exit the portfolio; the number of sectors to invest; the time to sell the stocks, this is, the minimum time the stock needs to remain in the portfolio until it can be sold. Furthermore, the last gene represents the metric defining the score that the companies need to rank in order to be an acceptable company, this is, if the company’s score is higher than the value on the last gene, the company can be selected to buy. A representative image of the structure for the chromosome is displayed in figure 27. In the section 3.6.5 are explained all constraints genes, along with some additional constraints not present in the chromosome.

![Figure 27: Picking Genetic Algorithm Chromosome](image)

As the reader can observe from the previous image, each ratio has a random weight. The final score attributed to the individual is given by the following equation:

\[
\hat{\text{Score}}(X) = \sum_{n=0}^{N} W_i \text{Ratio}(X, i)
\]  

\[
0 \leq \text{Ratio}(X, i) \leq 5
\]  

\[
0.0 \leq W_i \leq 5.0
\]  

\[
0.0 \leq \sum_{n=0}^{N} W_i \text{Ratio}(X, i) \leq 325.0
\]

Where:
- \( \hat{\text{Score}}(X) \) is the score given to the company \( X \).
- \( W_i \) is the weight assigned to the fundamental ratio \( i \).
- \( \text{Ratio}(X, i) \) corresponds to the normalized result of the ratio \( i \) for the stock \( X \).

3.6.2 Selection - The New Generations

Having defined the individuals that will populate the GA, it is necessary to specify how the algorithm will select the individuals that will generate the next generation, this is, choose the breeders to generate new individuals or to simply remain in the population. The chosen process is known as the “Natural Selection” methodology, which mainly consists in sorting the individuals according to their fitness value and subsequently select some of them for reproduction. As the name indicates, it is a natural selection process, which means that the best individuals will “survive”, but it is not a guarantee, as it also happens in nature, sometimes the best ones are eliminated. Although, in this work, we want to prevent that from happening. With that being said, besides the already stated methodology, an additional feature is applied: In every reproduction process, the fittest individual is preserved and passes on to the next generation. Having the fittest individual stored, half of the population is selected to breed the new population, using genetic operators such as mutations and crossovers.
3.6.3 Individual’s Evaluation - The Fitness Function

In some phases of the algorithm it is necessary to sort the population. For example, to elect the half of the population that will breed the next generation, or to simply select the fittest individual in the last phase. In order to sort the population, from the best to the worst individual, this is, from the fittest to the most incapable, and this way converge on a optimal solution, the fitness function is applied on every individual. The fitness function in this work is the Return On Income (ROI), calculated with the output of the Portfolio Simulator. The ROI function is used to evaluate the gains obtain when investing during a specific period of time. Of course, the higher the returns are, the fitter the individual is, so the goal of the algorithm is to maximize the ROI. Further in this section, the functioning of the Portfolio Simulator is described. Also, in the Section 4, where the reader can read about the system validation, this measure has a more detailed description.

3.6.4 Genetic Operators

For the process of evolution, a key factor is the genetic variation. After the evaluation of the population, these operators are applied to the individuals. As explained before, the natural selection process is the first operator used. Once the selection of the fittest is concluded, the new population is completed through the reproduction (also known as crossover) and mutation. This is an essential operation to guarantee the genetic diversity needed for the evolution from generation to generation, combining existing individuals (with proven value) into new solutions, getting closer to the optimal solution at each generation. The two genetic operators used in this work are the crossover and the mutation. Both were briefly explained before, nonetheless, their practical implementation is described next.

3.6.4.1 Crossover

The crossover operation is based on the natural process of reproduction. Two different individuals are selected, and one or more new individuals are produced through the combination of the genes.

The parents, this is, the two individuals selected to combine and give birth to the new one, are randomly chosen from the population priorly selected by the selection process. Using the population after the selection process guarantees that the parents are good solutions (are the best available at least). By matching two good solutions, the genetic algorithm is more likely to find a better solution and ultimately converge in an optimal solution.

There are many possibilities for the combination of the genes when executing the crossover. The most typical methodology is to select two chromosomes, parent A and B, split them in half and create two new chromosomes, child X and Y: child X with the first half of parent A and the second half of the parent B; child Y with the first half of parent B and the second half of the parent A. This method is represented in figure 28.
Many other methods can be applied, for example, split the chromosomes in three parts instead of two or using three parents, etc. For this work, the previous method wasn’t good enough. The reason is that the explained method considers all genes to be the same, this is, all of them are weights for example, being freely inter-mutable. Although, as previously explain, the chromosomes implemented in this work are made of 14 weights and 5 constraints, being not inter-mutable. This problem was overpassed implementing a specific crossover method for this work, represented in figure 29.
The methodology is to combine only genes which represent fundamental ratios weights. One of the parents give the constraints’ genes to the child, and the weights’ genes are a combination of both parents.

3.6.4.2 Mutation

The mutation operation assures genetic diversity amongst solutions. As it is applied after the crossover operator, even if some new chromosomes are similar to their parents, with mutation they will have more incisive differences.

Contrary to the crossover procedure, only one chromosome is required to create a new one using mutation. A random individual is chosen from the population priorly selected by the selection process and updated with the crossover procedure.

Again, many possibilities exist to apply mutation to the chromosomes. The general case is plain simple: given a random chromosome, the algorithm creates a new individual by copying the genes of the parent, and modifying the genes accordingly with a pre stated mutation rate, this means that some genes are slightly modified and others stay equal to the parent.

In this work, the mutation rate is 10% and the methodology is represented in figure 30. Given one chromosome from the available population, each gene as an one in ten chances of being modified, accordingly to its range of values. To avoid having repeated individuals in the population and discourage the over fitting, if the new chromosome is exactly the same as the parent, this is, neither of the genes suffered modifications, the process is repeated until the child has some differences.
3.6.5 Parameters/Constraints of the Trading System

The last seven genes of each chromosome, or individual, represent parameters of the system. This parameters are used by the Portfolio Simulator module, defining the behaviour of the system. Besides the parameters provided in the individual’s genes, some constraints are present to all individuals. Notice that not all parameters are used in every simulation. Depending on the strategy employed, some of the parameters can be ignored. This variations are well described on the Portfolio Simulator module. In the next subsections these parameters and constraints are exposed.

3.6.5.1 Portfolio’s Size

Information present on each chromosome. As the reader can observe in the figure, the number can vary between 20 and 50. This value represents the maximum number of stocks that can be present in the portfolio at every moment. Notice that this is the maximum number, it does not mean that the portfolio cannot contain less stocks than the specified value. For example, in the first few days of simulation, usually the portfolio is not completely full.
3.6.5.2 Gain Stop

Like the portfolio’s size, this is a gene’s information. This gene represents the percentage of profit considered “enough” for the algorithm to finish. Given a certain moment, if the portfolio’s profit is superior to the value on this gene, the algorithm proceed to close all positions and take the profit home. The acceptable profit percentage in this work is 50%.

3.6.5.3 Minimum Jump To Entry/Exit the Portfolio

These two genes, one representing the value to entry and the other to exit, are very important genes for the portfolio management. First, clarify their meaning. Every simulation cycle, companies are selected and kept as possible companies to insert in the portfolio. This selection is based on their rank, calculated using the weights present in the other genes and the corresponding fundamental ratios. After the selection, the process of buying and selling begins.

Although, besides the actual rank of the companies, other values are considered. Before the algorithm makes the decision whether to buy or not the stock, it also considers the last rank of the company. For the decision to be positive, the following equation must be verified:

\[
\text{Jump} \geq M J_{\text{Entry}}
\]

\[
\text{Jump} = LR - AR
\]

Where:

- \( M J_{\text{Entry}} \) is the Minimum Jump To Entry value, present in the chromosome, varying between 20 and 30, as illustrated in the figure [31].
- \( LR \) is the last rank position of the company.
- \( AR \) is the actual rank position of the company.

This verification ensures that the algorithm only invests in a company that presents real improvements and high chances of growth, leaving behind companies that are basically as they were in the last simulation cycle (even if they are doing slightly better).

Analogous to the previous process, the exiting process also considers the last rank of the company. In this case, the equations to be verified are:

\[
\text{Jump} \geq M J_{\text{Exit}}
\]

\[
\text{Jump} = AR - LR
\]

Where:

- \( M J_{\text{Exit}} \) is the Minimum Jump To Exit value, present in the chromosome, varying between 20 and 30, as illustrated in the figure [31].

Contrary to the previous case, this verification ensures that the algorithm only sells companies that are presenting fairly poorer results when compared to the previous values.

An illustration is provided to clarify the Minimum Jump to Entry action on figure [31]. Let’s assume a minimum jump value of 20. As the reader can observe, in the first case, the \( SLB \) company was in the 29\textsuperscript{th} position at time \( N \). At time \( N + 1 \), the same company was positioned in the 9\textsuperscript{th} place. Following the previous equations, the jump was 20, so the company is good to buy. On the second case, the jump verified was only 8 positions, which is not enough and as a result the company is discarded.

56
Every simulation cycle the companies’ ranks are different. Without these verifications, buys and sells would happen everyday, which is not recommended for two main reasons: (1) If the company was bought a few days back, the price hasn’t got the time to raise so there is no profit in selling it; (2) every time the algorithm changes position, buy or sell shares, money is lost in transaction’s costs. This two verifications are very important as they prevent constant changes in the portfolio, avoiding premature sells and money waste in transaction’s costs.

### 3.6.5.4 Number of Sectors to Invest

Again, this is information passed as a gene of the chromosomes. As mentioned, the companies are distributed by 10 different sectors. Of course, every sector has good and bad companies. Regardless, in this solution, each simulation is limited to a certain number of sectors. This means that the simulator can only select companies amongst the X best sectors. For each trimester, a new rank of sectors is calculated, and the companies from the worst sector are discarded from the possibilities list. As illustrated in the figure, the value is an integer between 3 and 7, which means that the algorithm is forced to buy companies from more than one sector, but is also unable to buy from all sectors. This decision is based on two reasons: (1) prevent the algorithm from investing in only one sector. If this industry would collapse, the investor would lose all his money. (2) only buys from the best momentary sectors, avoiding at least the worst 3 in every simulation.
3.6.5.5 Time to Sell

Like the previous parameters, this is a value present in one of the genes of each individual, assuming an integer value between 15 and 60. The value’s unity is days. Similarly to the minimum jump to exit parameter, the objective of this gene is to avoid selling shares too soon, giving it time to raise its price. If the algorithm chooses to buy a given stock, it cannot sell it right away. To sell the stock, the following equations must be verified:

\[
DP \geq TTSell \tag{41}
\]
\[
DP = BD - PD \tag{42}
\]
\[
0 \leq BD \leq 365 \tag{43}
\]
\[
0 \leq PD \leq 365 \tag{44}
\]

Where:

- \(TTS\) is the Time To Sell value, present in the chromosome, varying between 15 and 60, as illustrated in the figure. \[27\]
- \(DP\) is the number of days passed between the bought and the present date
- \(BD\) is the Buy Day
- \(PD\) is the Present Day

3.6.5.6 Buy Signal

The Buy Signal value is the last gene of the chromosome. As already mentioned in the begin of this section, this value represents the minimum value that a company must score to be considered for buying. The score for each company is calculated daily, using the fundamental ratios’ values for the given day and the respective weights. The diagram in the figure \[32\] illustrates the procedure.
3.6.5.7 Other Constraints

- **Long only constraint** indicates that in this work only long positions are allowed. This means that the money invested in any share is always positive, no short selling is contemplated.

- **Transaction costs** are the costs associated to every buy and sell operation. These costs affect nega-
tively the profit of the portfolio. To avoid great costs in transactions, the frequency of trading must be reduced. In this work, the commissions paid for each transaction is 0.3% of the transaction’s value.

- **No dividends** are considered in this solution. Dividends are commissions paid by the companies to the shareholders. One more time, this constraint has a negative impact on the portfolio’s profit, as we are not considering income money.

### 3.7 Portfolio Simulator Module

The portfolio simulator uses each individual to create a portfolio, simulating a real portfolio management with real data, provided by the *Data Processing Module*. As mentioned, the simulation uses stocks from the S&P 500 index. This module is used by the genetic algorithm as the fitness function. For each chromosome, the simulator performs a real-life simulation, returning to the genetic algorithm the final cash balance produced. The genetic algorithm can then rank the chromosomes by their ROI.

Below, the reader can understand how this module is implemented. Also, the different investing strategies employed are described.

#### 3.7.1 Implementation and Functionality

As stated above, the *Portfolio Simulator Module*’s objective is to simulate the management of a portfolio. This simulation as an one year range, using the data accessible through the *Data Processing Module*. In the figure 33, the reader can observe a diagram representing the classes used to implement this module. The basic data flow of the algorithm is represented on the figure 34. Below, a description for each class is provided. Notice that this data flow and the definitions for each class presented below are the general concept. For some strategies, the data flow and the verifications performed during the process suffer minor changes. Those changes are explained later on the versions and strategies sections.

![Figure 33: Portfolio Simulator class diagram](image-url)
Figure 34: Portfolio Simulator flow diagram
**Simulator:** This is the class which controls the operations. It starts with the creation of a new portfolio, using the *Portfolio* class and the necessary parameters, available in the chromosome (maximum number of stocks, gain stop percentage, minimum jumps, time to sell and buy signal). Next, the main cycle begins. Considering the years in evaluation, either training or testing, the algorithm begins its execution on the January 1st and iterates through everyday until the 31st of December. For each day, two actions are considered: (1) Calculate the ranking for the day; (2) performing the adequate investment operations or move on to the next day. Both actions are carry out through the *Wall Street* class, explained below.

**Wall Street:** This is the main class of this module (the name is merely symbolic). The reason why this class is the most important class is because it represents the leader of the module, this is, the class using all other classes to simulate a real life portfolio management. In a figurative way, this class is the environment in which the simulation is running, keeping critical information like the companies’ data, the year of the simulation, what strategies are to deploy and the information extracted from the chromosomes. Two large operations can be considered: (1) calculate the ranking for a given day and (2) simulate the wall street environment for that day. Below the reader can read a more detailed explanation of these two operations.

1. The first operation has more purpose than a simple daily update of the companies’ ranking. The process is divided in three steps.
   - Calculate the rates of the companies based on their fundamental ratios for the specific day (update the ranking).
   - Use the *Ranking* class to sort the companies for each sector, this is, create different rankings for each sector.
   - Extract which companies are suitable to buy on that day, using one more time the *Ranking* class functionalities.

2. The second operation is the actual management of the portfolio. The first operation makes available the information concerning which companies to buy in that day. Although, it is not mandatory that the simulator will in fact buy those stocks. In this operation, based on the strategy, some verifications must me performed before the final decision. The process starts with an iteration through all companies on the “possible companies list”. For each company, the following rules are applied:
   - if the company’s rate is inferior to the buy signal, it is discarded and the algorithm continues to the next iteration.
   - if the portfolio has free slots, the company is automatically added to the “buy list”.
   - if the portfolio is full, it is necessary to verify if the proposed company has given the minimum jump to entry the portfolio and if it is better than any company that is already present in the portfolio.
   - if the proposed company verifies the immediately above condition, the worst company in the portfolio is added to the “sell list” and the proposed company is added to the “buy list”.

Finished the previous process, to conclude the second operation, the trading is executed. The trading operation is performed using the *Portfolio* class, to which is given the “sell list” and the “buy list”, along with the actual date.

**Portfolio:** The *Portfolio* class simulates a normal real-life portfolio. This object keeps information regarding the money invested/ money available, the total current value of the portfolio, which companies are in the portfolio at the moment and how many shares of each company were bough. Besides these informations, the *Portfolio* class is responsible for all trading functionalities. As mentioned above, to start the transaction process, the *Wall Street* class invokes the functionality for a specific day, sending to the portfolio a “buy list” and a “sell list”. After receiving the transaction order, to complete the operation, the following steps must be concluded:

- All stocks in the “sell list” must be sold. The sell operation for each stock comprehends two actions: (1) remove the company from the portfolio; (2) given the shares’ price for the day, increase the
available money of the portfolio accordingly with the equation:

\[
AvailableMoney = AvailableMoney + (SH \times price)
\]  
(45)

Where:

* \( SH \) is the number of shares the portfolio is holding for the given company.

– Having free slots in the portfolio, the companies from the “buy list” should now be acquired. Similarly to the sell operation, for each company, the buy operation follows two actions: (1) add the company to the portfolio constituents; (2) given the share’s price for the day, decrease the available money of the portfolio accordingly with the equation:

\[
AvailableMoney = AvailableMoney - (STB \times price)
\]  
(46)

Where:

* \( STB \) is the Shares to Buy. The number of shares to buy in each operation depends on how much money is available to invest in the company. This investment value is decided depending on which strategy is being used in the simulation. The different strategies to invest are explained on the next section.

**Ranking:** The *Ranking* class has several functionalities. Given the maximum number of stocks allowed in the portfolio and the number of sectors available to invest, this class provides the simulator with the following information:

– The global ranking of the companies, this is, the ranking of all S&P 500 companies, regardless of the sector.

– The ranking for each sector.

– The ranking of the sectors for the trimester in the moment.

– How many companies can the simulator buy from each sector. These numbers are calculated using the performance of each sector for the current trimester. The performance can be measure in many different ways, explained further in this document. Every sector has an associated percentage for the trimester, which translates to a number of companies. For example, considering that the performance is measured using the revenue. If the sector \( S \) represents 25% of the total index revenue for the trimester, then 25% of the companies on the portfolio will be from the sector \( S \).

– Which companies should the simulator choose that day. This functionality combines the ranking for each sector with the slots given to the sector, choosing the \( x \) top from each ranking.

In the figure [35] the reader can see examples of the information provided by the *Ranking* class.
3.7.2 Simulation Versions

There are an infinite number of possibilities to approach the stock market world when it comes to trading options. A considerable number of decisions must be made: which market, which kind of assets, which companies, etc.. In this work, some of these questions were already answered: the software should invest in companies’ stocks, from the S&P 500 index. That said, a few more things were to define. When should a company’s stock enter and leave the portfolio? The decision is made supported with some verifications to the system’s current state. If the conditions are the desired, then buy. If not, sell or do nothing. In this work, three different versions were implemented to deal with this problem. The versions are explained below. Following the previous ratiocination and terms, the difference between versions lies in the way that companies are packed into the “sell list” and “buy list”.

<table>
<thead>
<tr>
<th>Global Ranking</th>
<th>Sector Ranking</th>
<th>Nº Companies/Sector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pos</td>
<td>Company</td>
<td>Pos</td>
</tr>
<tr>
<td>1</td>
<td>AAPL</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>GOOG</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>AMZN</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>SLB</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>ZION</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>KMG</td>
<td>7</td>
</tr>
<tr>
<td>...</td>
<td>MSFT</td>
<td>...</td>
</tr>
<tr>
<td>500</td>
<td>L</td>
<td>10</td>
</tr>
</tbody>
</table>

**Ranking per Sector**

<table>
<thead>
<tr>
<th>Pos</th>
<th>Energy</th>
<th>Utilities</th>
<th>Health Care</th>
<th>Financials</th>
<th>Materials</th>
<th>Industrials</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>COG</td>
<td>NI</td>
<td>AMGN</td>
<td>TRV</td>
<td>DOW</td>
<td>DE</td>
</tr>
<tr>
<td>2</td>
<td>XEC</td>
<td>AWK</td>
<td>CELG</td>
<td>IVZ</td>
<td>CF</td>
<td>FAST</td>
</tr>
<tr>
<td>3</td>
<td>XOM</td>
<td>NE</td>
<td>HSIC</td>
<td>FITB</td>
<td>IP</td>
<td>HON</td>
</tr>
<tr>
<td>4</td>
<td>EOG</td>
<td>WEK</td>
<td>HCA</td>
<td>ICE</td>
<td>LYB</td>
<td>IR</td>
</tr>
<tr>
<td>5</td>
<td>CVX</td>
<td>XEL</td>
<td>MYL</td>
<td>TROW</td>
<td>T</td>
<td>MAS</td>
</tr>
<tr>
<td>6</td>
<td>MUR</td>
<td>NRG</td>
<td>PFE</td>
<td>USB</td>
<td>CTL</td>
<td>RHI</td>
</tr>
<tr>
<td>7</td>
<td>SLB</td>
<td>SO</td>
<td>STJ</td>
<td>AIZ</td>
<td>MLM</td>
<td>XYL</td>
</tr>
<tr>
<td>8</td>
<td>NOV</td>
<td>ED</td>
<td>REGN</td>
<td>C</td>
<td>AVY</td>
<td>CAT</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Figure 35: Different Rankings produced by the *Ranking* class
3.7.2.1 Benchmark

The benchmark version is the most simple version. This is the main version, as it is common to the other two versions, this is, the verifications performed in this version are also performed in Minimum Exit and Mandatory Timing versions.

The first verification made is if the company rate is higher than the buy signal. Next, if the company is not in the portfolio already and the portfolio is not full, the company is added to the “buy list”.

In the case where the portfolio is already complete, more actions should be taken to decide if the portfolio should remain as it is or if its time to substitute one of the companies for the new one. After calculate the leap that the company gave from the last ranking to the current one, the verifications continue. If the leap is smaller than the minimum jump needed to entry the portfolio, the actions terminate and the company is discarded. If the leap is bigger than the minimum jump, one last verification is required. If the rate of the worst company in the portfolio is smaller than the rate of the proposed company, then, the worst company is added to the “sell list” and the new company is added to the “buy list”.

3.7.2.2 Minimum Exit

This version appear due to the elevated number of transactions occurring with the other versions. As mentioned, for every transaction, 0.03% of the transaction’s value is paid in commissions. To reduce the number of transactions made, a new version was created. In this version, one more verification was added to the previous ones. A company in the portfolio may only be sold if it has a worst rank than the proposed company and if it has fallen considerably from the previous ranking to the current ranking. The leap that the company made is calculated and compared to the minimum jump needed to exit the portfolio. If the leap is higher, the company is added to the “sell list” and the proposed company added to the “buy list”.

This verification was added because in some cases, even considering that the worst company is not in the higher places of the current ranking, if it is only one or two slots below the allowed in the portfolio, is more than likely that in a near future the company will be on the top again. In those cases, the costs of sell and re-buy a few days later are not justified.

3.7.2.3 Mandatory Timing

The Mandatory Timing version introduces a new investment flow. Instead of free timing to buy and sell, a new time constraint is applied. Stocks can only be sold if they were kept in the portfolio for X days. Even if the company performed the worst of all companies, it cannot be sold until the minimum date is not due. The number of days is given by the gene time to sell.

Despite the fact that this version also reduces the number of transactions, it was implemented to solve a different problem. Some stocks were being bough and sold few days apart. In some cases, such as bankruptcy, this is a good decision. Although, in other cases, it is the worst decision the investor can do. If a stock is bought on day 1 and sold on day 2, the price is probably the same (or very close). The simulator was not giving enough time for the stock to raise its price and be profitable. Again, since the price was nearly the same, no revenue was made from the transactions and money was being lost in commissions. With this version the simulator is forced to keep the stocks in the portfolio long enough for the price to grow.

3.7.3 Investing Strategies

In this stage, the companies are chosen, both sell and buy list are ready. Only one last question needs answering: how much money should the simulator invest in each company. Again, this question has an infinite amount of possible answers. Three strategies were implemented and tested in this work:

3.7.3.1 Strategy A - Equal Wealth Distribution

Using this strategy, the methodology is simple. The portfolio sells all companies from the “sell list”, increasing the available capital. Then, the capital available is divided in equal portions by the companies in the “buy list”. The money invested in each company follows the equation:

\[
Investment_c = \frac{AvailableMoney}{BuyListSize}
\]  

\[\text{(47)}\]
Where:

- \( \text{Investment}_c \) is the investment for the company \( c \) in the “buy list”.

To illustrate the strategy, an example is provided:

The money available at a certain moment is 100 thousand dollars. The number of companies on the “buy list” is 5. Following the previous equation, each company will receive an investment of \( \frac{100000}{5} = 20000 \) dollars. As the reader can see in the figure 36, the wealth is equally distributed for the companies, regardless of their sector or performance.

\begin{align*}
\text{Investment}_c &= \frac{\text{AvailableMoney} \times \text{Performance}_s}{\text{BuyListSize}_s} \\
0.0 \leq \text{Performance}_s \leq 1.0
\end{align*}

Where:

- \( \text{Performance}_s \) is the weight represented by the sector \( s \) considering the entire index. For instance, if the total revenue of the index was 100 million dollars, and the sector \( \text{Healthcare} \) had a revenue of 15

---

Figure 36: Strategy A investment example

### 3.7.3.2 Strategy B - Money Makes Money

The second strategy is named after the known proverb “money makes money” because it was inspired in a basic principle - if one sector is making more money than the others, more money should be invested in that sector. Like the previous strategy, the first step is to sell all companies from the “sell list”. Next, the distribution of money is performed accordingly with the following equations:

\[ \text{Investment}_c = \frac{\text{AvailableMoney} \times \text{Performance}_s}{\text{BuyListSize}_s} \]  
\[ 0.0 \leq \text{Performance}_s \leq 1.0 \]

Where:

- \( \text{Performance}_s \) is the weight represented by the sector \( s \) considering the entire index. For instance, if the total revenue of the index was 100 million dollars, and the sector \( \text{Healthcare} \) had a revenue of 15
million dollars, the $Performance_{healthcare}$ is 0.15, meaning that 15% of the available money will be invested in companies from the Healthcare sector.

- $BuyListSize_s$ is the number of companies of the sector $s$ present in the “buy list”.

In figure 37, an example is provided to help the reader to better understand this strategy methodology. Considering that the simulator has 100 thousand dollars available at the moment and can only buy stocks from two different sectors, Health Care and Information Technologies, with a performance of a 0.7 and 0.3 respectively. If in the “buy list” are 5 and 3 companies from Health Care and Information Technologies sector, following the above formulas, each company from Health Care will receive an investment of $\frac{100000 \times 0.7}{5} = 14000$ dollars, while companies from Information Technologies will receive $\frac{100000 \times 0.3}{3} = 10000$ dollars each.

\[
\text{Performance}_X = 0.7, \quad \text{Performance}_Y = 0.3, \quad \text{BuyListSize}_X = 5, \quad \text{BuyListSize}_Y = 3
\]

\[
\text{Investment}_X = \frac{100000 \times 0.7}{5} = 14000, \quad \text{Investment}_Y = \frac{100000 \times 0.3}{3} = 10000
\]

\[
\text{Equation:} \quad \text{Investment}_s = \frac{\text{Available Money} \times \text{Performance}_s}{\text{BuyListSize}_s}
\]

This strategy has a problem though. The objective is to invest more money on the best sectors. The problem is that, if at a certain moment, the “buy list” contains many companies from the best sector and just one or two from the worst, as it divides the money for the number of companies in the “buy list”, the five or six companies from the best sector will end up with less investment than the one or two from the worst. The following example illustrates the strategy and this problem.

The money available at a certain moment is 100 thousand dollars. The number of companies on the “buy list” is 6, being 4 companies from sector $X$ and 2 from the sector $Y$. The $Performance_X$ is 0.3 and $Performance_Y$ is 0.15 respectively. Following the previous equations, each company from the sector $X$ will receive an investment of $\frac{100000 \times 0.3}{6} = 5000$ dollars. On the other hand, each company from the sector $Y$ will receive an investment of $\frac{100000 \times 0.15}{2} = 7500$ dollars.

Although the performance of sector $X$ is two times better than the performance of sector $Y$, the companies from sector $Y$ are receiving an extra investment of 2500 dollars. Despite the fact that this is not a regular
situation, a new strategy was implemented to counter this problem, explained next.

### 3.7.3.3 Strategy C - Take All You Can

The *Take All You Can* strategy tries to solve the problem mentioned above. In this strategy, best companies have more money invested. The “buy list” is mainly sorted by sector’s quality, and by company’s rank on a second phase. With that said, the investing starts in the best company and finishes at the worst. In the end, to avoid having available money when it could be invested, if there is still money available, it is equally distributed for all companies in the portfolio. To each company, the investment made follows the equation:

\[
Investement_c = AvailableMoney \times Performance_s
\]  

In the end of the process, if money is still available, a reinforcement is performed following the equation:

\[
ReinforceInvestement_c = \frac{AvailableMoney}{PortfolioSize}
\]

Where:

- \( ReinforceInvestement_c \) is the money to be invested in company \( c \).
- \( PortfolioSize \) is the number of companies in the portfolio.

To illustrate this strategy methodology, an example is presented:

The money available at a certain moment is 100 thousand dollars. The number of companies on the “buy list” is 5, being 2 companies from sector *Health Care*, 2 from the sector *Information Technologies* and 1 from the sector *Energy*. The *Performance\(_{HC}\)*, *Performance\(_{IT}\)* and *Performance\(_{E}\)* is 0.3, 0.2 and 0.1 respectively. The investments for each company are:

- \( Company_{HC1} = 100000 \times 0.3 = 30000 \) dollars
- \( Company_{HC2} = 70000 \times 0.3 = 21000 \) dollars
- \( Company_{IT1} = 49000 \times 0.2 = 9800 \) dollars
- \( Company_{IT2} = 39200 \times 0.2 = 7840 \) dollars
- \( Company_{E1} = 31360 \times 0.1 = 3136 \) dollars

At this moment, around 28000 dollars are available, so in every company from the portfolio is invested an equal share of the money. Supposing a portfolio with 10 stocks, each company will receive a reinforcement of 2800 dollars. The illustration for this example can be seen in the figure 38.
3.7.4 Sector Picking Choices

Being this work a solution which innovates by using the sector division of the companies, this is a major choice. Every trimester, sectors must be ordered by their “value” or potential. To dictate which sectors are the best, many options can be considered. In this solution, four different indicators were experimented.

- **Revenue:** The revenue of each sector on the given trimester. This indicator is calculated using the revenues of the companies from the sector. An higher revenue implicates high profit in the companies constituting the sector. Good trimesters for the companies in the sector suggest good opportunities to
invest. The profit each trimester made is presented in millions.

![Revenue Graph](image)

**Figure 39: Graphic for revenue of all sectors**

- **Revenue Growth**: The growth of the revenue in relation to the previous trimester. As this is a solution seeking for high growing companies, the growth indicator is very important. Some sectors may present high revenues every trimester, although not growing much. With this option, priority is given for sectors presenting high growth rates, even if their revenue is smaller when compared to the other sectors. The growth comes in percentage.
Earnings Per Share: The earnings per share indicator is also a very important indicator to consider, as it is calculated using the net income, which is the revenue subtracted by all taxes and expenses. Sectors reporting higher earnings per share are the sectors making more real money, so this can be a more conclusive indicator than the revenue.
Earnings Per Share Growth: The last indicator is the earnings per share growth in relation to the previous trimester. An high growing earnings per share not only indicates that the company is producing more cash every trimester, but also that the internal management is highly productive.
As the reader can see from the previous four figures, figures 39, 40, 41 and 42, all indicators result in very different graphics, which means that changing between these options will produce different results.

3.8 Chapter Conclusion

Within this chapter is a full description of the developed solution. The reader can have an insight on how the software is structured and implemented. For every module the reader has available several diagrams and full descriptions of its functionalities. The chapter has five main sections: the system overview and general architecture; the data module, where is available the methodology to download and process all data used in this work; the ratios module, where all ratios used in this work and the methodology to calculate them are described; the genetic algorithm module is described next, including all aspects relevant for the reader to understand how this algorithm works and was implemented. The genetic algorithm section includes detailed descriptions about the individuals implementation, selection and reproduction, along with the fitness function and several restrictions. For last, the portfolio simulator module provides information concerning the several software versions and investing strategies implemented and tested in this work.

Notice that almost every aspect in this solution is easily modified, this is, the modules are implemented in an abstract way, allowing future re-use of the software for different cases. For example, download and process data from other indexes, add new ratios, training/testing in different time spans, use different simulation strategies, change the stop orders and other restrictions concerning the genetic algorithm. On chapter 5, the reader can understand the current system’s problems/limitations and how it can be improved.
4 System Validation

In this section are explained the several validation metrics used to evaluate the system. It is important to validate a new system after its implementation. System’s results can not be considered positive if not compared to other systems and without some metrics calculation to verify that it has indeed achieved the objectives and surpassed previous solutions. In this work, the valuation comprehends three main calculus:

- **Performance:** The portfolio’s performance was evaluated using the fundamental indicator Return on Investment (ROI).

- **Volatility/Risk:** The portfolio’s volatility was evaluated calculating the variance of the portfolio.

- **Risk Adjusted:** The risk associated to the investments made throughout the portfolio simulation is calculated using the Sharpe Ratio.

The system validation is completed through the analysis of some quantitative measures, which will help the reader to understand the quality of the solution.

The section finishes with the presentation of the results obtained using this work’s final product, explaining in more detail two case studies.

4.1 Validation Measures

4.1.1 Performance - Return on Investment (ROI)

One of the most common ratios used to measure the performance and efficiency of an investment is the Return on Investment (ROI). The formula is a simple profit formula showing how much percent the user got from his investment:

\[
ROI = \frac{Final\ Return - Investment}{Investment} \times 100
\] (52)

In this work, once the system allows the portfolios to have several different stocks, the final return is given by:

\[
Final\ Return = \sum_{i=1}^{n} NS_i \times S_i \times P
\] (53)

Where:

- \( NS_i \) is the number of shares the portfolio has of company \( i \).
- \( S_i \times P \) is the current stock’s price of company \( i \).

Using the \( ROI \), an evaluation through comparative techniques is possible and very simple to perform. Besides the importance of the annual mean return from the portfolio itself, the performance of others takes also an important part. That being said, a way to investigate the performance of the chosen portfolio is calculating the ROI and comparing it with the mean return of the S&P 500 index, or eventually with the mean return produced by other existing solutions.

Notice that although it may be a strong measure for portfolios comparison, some cautions should be take. It is necessary to evaluate the period of the investment. A comparison between a two year investment and a one year investment can lead to incorrect conclusions about which one is better. Also, a risk and volatility comparison is important, as some investors would prefer less return if it means less risk and more stability. In the figure [13] the reader can see why the comparison of the ROIs for two portfolios may not be enough by itself.
Figure 43: ROI graphs for two investments

Although both investments ended up with a ROI of 18%, in the first case it took two years to achieve it. Also, looking at the curves, the second option is much more stable, with practically no loses at all, while in the first case the portfolio value adopted negative values in the end of the first year. These graphs prove that the ROI is a good measure, but not enough to conclude with certainty which investment is better.

4.1.2 Volatility/Risk - Variance

The variance is a probability distribution measure which measures how spread are the numbers in a data set. The calculation follows the formula:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$$

(54)

Where:

- $X_i$ is the portfolio value for day $i$. 

• $\mu$ is the mean of the portfolio value for all simulation period and is given by:

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N} \quad (55)$$

• $N$ is the number of the days in the simulation period.

The daily values of a portfolio for all simulation can be seen as a data set of numbers. Applying the variance formula to this data set provides the investor with an idea of how disperse are the portfolio values from the mean daily return of the portfolio. An high variance value indicates that the returns are not easily predicted, this is, one day with an high return does not indicate that the following day will also return high profit. An highly volatile portfolio is undesirable for most investors, as it means a great amount of uncertainty and risk about the portfolio’s value. On the other hand, with a low variance value, the returns tend to be more stable and predictable, this is, is not likely that the portfolio’s value will be cut in half overnight.

Besides the stability and peace provided to the investor, a less volatile portfolio allows the investors to explore techniques to improve profit, such as the leverage technique.

From the figure 44 the reader can see two graphics, representing an highly and low volatile portfolio. Although both have a final ROI of 20%, the second is much more stable and risk free, so considering the two measures, the second portfolio is the best.
In this work, in order to compare the volatility between the case studies’ results and the S&P 500, the data set used is not the portfolio’s value for each day, but the gain percentage from the previous day. This variation was implemented because the S&P 500 index is quoted in values around 2000 and the portfolio’s values in this work round the 1000000. Using the gain percentage, the values for both are normalized, which allows a fair comparison. In figure 45, the reader can observe the transformation and its motivation.

Figure 44: Variance graphs for two investments
4.1.3 Risk Adjusted - Sharpe Ratio

The Sharpe ratio is a referenced metric used in finance to evaluate the risk/return of an investment. To decide which investment is better, besides the two comparisons stated above, is also important to realise if the expected return of the investment is due to good investment decisions, or if the investor is choosing very profitable assets carrying big risks. It is well known in the financial world that there is no such thing as a free risk asset. Although, while some people are more willing to accept risk in order to achieve higher returns, other people prefer a more modest return, if it implies a safer and stable investment. This ratio gives a perception to the investor about how much extra volatility he will have to deal with for the extra returns, this is, how much more risk he is taking for the higher returns. The calculation of this ratio follows the formula:

$$\text{Sharpe Ratio}(P) = \frac{R_P - R_F}{\sigma_P} \quad (56)$$

Where:

- $R_P$ is the average rate of return, or the expected return, of the portfolio $P$.
- $R_F$ is the Risk-Free rate. The Risk-Free rate is a theoretical value which represents the rate of return an investor would achieve by performing an investment with no risk at all. It is called theoretical because no investment in the world is risk free. In other words, is the minimum return rate the investor is expecting to receive from the investment. An higher return, would imply an addition of risk to the investment. Although it is a theoretical value, the United States Treasury Bill is often used as reference, as it is considered the safest asset in the world.
- $\sigma_P$ is the portfolio $P$ standard deviation.

To illustrate the use of this ratio, a simple example is provided. Consider two investments managers, $A$ and $B$. Manager $A$ holds a portfolio which generates a return of 25%, while manager $B$ with his portfolio only generates a profit of 10%. At first sight, the best investment is obviously the one that manager $A$ holds. Although, on a second analysis, manager $A$'s portfolio has a standard deviation of 10%, while manager $B$'s
portfolio is much more stable, presenting a standard deviation of 2%. Assuming a risk-free rate of 5% and following the previous equations: \( \text{Sharpe Ratio}(A) = \frac{20 - 5}{10} = 1.5 \) and \( \text{Sharpe Ratio}(B) = \frac{10 - 5}{2} = 2.5 \).

Although portfolio \( A \) produces twice as much as the portfolio \( B \), the volatility (risk) associated to the portfolio \( A \) is higher, which translates into a lower Sharpe ratio, and consequently manager \( B \) generates higher returns when risk-adjusted.

In this work, because there is no absolute value for the Sharpe Ratio referent to the S&P 500 index, an estimate was calculated. Also, as risk-free rate, approximate values were considered: for 2011 the risk-free rate is 3%, for 2012 is 2%, for 2013 and 2014 is 2.5%. These are higher values than the real ones, however, because there is no certain value, we chose to be pessimists.

4.2 Quantitative Metrics

Additionally to the previous performance measure, the return produced, the variation and the risk-adjusted, some questions/metrics can also be used to evaluate the several strategies employed:

- **Total profit:** Annual return percentage? How much money did the strategy produced per year (mean).
- **Maximum draw down (MDD):** Which was the largest percentage ever lost during the simulation?
- **Number of positions:** How many positions, this is, changes in the portfolio, did the simulation took during the entire simulation. As explained on the chapter 3, many changes in the portfolio result in a large amount of money spent in commissions, which has a negative impact on the portfolio’s return.
- **Percentage of profitable decisions:** From the total number of decisions made, how many were lucrative? At the time of selling the assets, the price of the stock is higher than when it was bought? Or did the price fell and the simulator lost money?
- **Transactions costs:** The amount of money spent in commissions.
- **Sector distribution:** The stocks chosen to construct the portfolio verify the dynamic distribution proposed, this is, best sectors detains more slots than the weaker sectors? Is important that the simulator adapts to the new reality of the market, changing between the more lucrative sectors at the end of each trimester. This metric may adopt four different values in form of percentage intervals. Each interval represents the percentage of time that the sector distribution was correct, which means, the major number of companies in the portfolio were from the best sector. The intervals considered are: \([0\% − 25\%], [25\% − 50\%], [50\% − 75\%], [75\% − 100\%]\).
- **Beta:** Is the portfolio highly correlated with the market? Is it able to support a bearish market without collapsing? Many portfolios may present good results when the market is also presenting good results. The really good portfolio is the one that when the market falls, is able to sustain the returns on a good level (comparatively to the market). Like the sector distribution, abstract values are used to evaluate this metric:
  - \( \text{LOW} \): if there is no apparent relation between the graphs. For example, when the index is going down, the simulation graph is in an upward trend.
  - \( \text{MEDIUM} \): if the simulation graph sometimes follows the index trend, but it may also counter its movement.
  - \( \text{HIGH} \): if the graphs are very similar. When the index curves up/down, the simulation graph also presents those exact same curves.

4.3 Results

This chapter aims to present full documentation concerning the results obtained in this work. The chapter is divided in three sections. The first to provide a overview of the results obtained and the methods employed in the process. In the second section are described some specific cases and their results. The last section presents an analysis and the conclusions obtained from the several simulations and tests performed.
4.3.1 Overview of the Simulations & Environment

As described on the third chapter of this document, the developed solution has many available options to decide how the companies are chosen and how the money is invested. These choices are made by the user before each simulation. Five major choices must be made:

- The market desired. In this case, the market is represented by the S&P 500 companies.

- The initial budget. In this work, the initial available money for all simulations is 1000000 USD. In "real life", one can opt to invest only 1000 USD. Although, in this work, the initial budget chosen is on the millions level because some companies’ stocks cost around 1000 USD, or even more. To avoid having unreachable companies, or only buy 1 stock from each company, the initial budget should largely surpass those values.

- The period of simulation. The investor should select the training year. Consequently, the testing phase is performed in the following year, this is, if the train occurred in 2010, the test is executed in 2011. A different approach is also available, which consists in performing the training in a two year period, this is, for example, training in 2010-2011 and testing in 2012. The data gathered and stored is referent to five years, 2010-2014.

- The most important choice is concerned with the available investing strategies. Again, as mentioned in the chapter 3, the investor must define the simulation version, the investing strategy and the sector ordering. Currently there are three versions, three strategies and four ordering options implemented. Each of the choices are independent, so combining all of them, the number of possible different simulations is $3 \times 3 \times 4 = 36$. Considering a 5 year range, with trainings performed in one or two years, 7 possible test years are available. The number of simulations increase to $36 \times 7 = 252$. As the results from training the software in a two year period did not differ much from the results obtained while training in an one year period, in the following sections only the one year training cases are considered. This means that the final number of simulations is $36 \times 4 = 144$.

- For last, the investor must choose the number of executions of each simulation. In order to achieve a good reliability level, each simulation should be repeated (with the same parameters) and the final result calculated following the formula:

$$Final\ Result = \frac{\sum_{i=1}^{N} Result_i}{N}$$  \hspace{1cm} (57)

Where:

- $Result_i$ is the result obtained in the simulation $i$.
- $N$ is the number of simulations performed.

Executing each simulation several times eliminate the possibility of an outsider result, either too high or too low. In this work, the number of executions performed for each simulation is 50.

In the figure 46 are presented the general parameters of the environment set to the simulation process, both stated in the previous paragraph and others that are not entered/specified by the user. The parameters of the genetic algorithm are exposed in the figure 47.
The table 48 summarizes all results. For the 36 possible combinations and for the four possible testing years (despite having data stored for 2010, it can only be used for training and consequently the results are not contemplated in the table).

For each combination (ex. Benchmark-C-REVENUE-2011), five different columns are displayed, presenting:

- **Mean Result:** The mean of returns for the 50 executions.
- **Min:** The minimum return obtained in the 50 executions.
- **Max:** The maximum return obtained in the 50 executions.
Var: The mean of the variations obtained in the 50 executions.
Shar: The mean of the Sharpe ratios obtained in the 50 executions.

In the figures 49 and 50, are exposed the results in three different abstraction levels. Besides the average results, values for the maximum and minimum are also available. In figure 49, the reader can see which way to order the sectors performed best given the combination Version + Strategy (sector ordering abstraction). As an example, using the version Benchmark and the strategy B, in 2014 the better results were achieved when the sectors were sorted by their EPS Growth.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Var</td>
</tr>
</tbody>
</table>

Figure 48: Results Overview Table
Figure 49: Best sector ordering for each combination Version-Strategy.
From tables represented in 48, 49 and 50, it is already possible to state a few comprehensive appreciations:

- The Benchmark-A-Revenue Growth combination outperformed the other 35 combinations in 3 of the 4 periods. As it is the best combination, it will be more exhaustively explained in the next section as a specific case study.

- In 3 out of 4 periods, the Benchmark-A achieved the best results.

- In all 4 periods, the Benchmark version outperforms the other two versions.

- Each period has its own “favourite” way to order the sectors. Despite that ordering the sectors by their Revenue Growth surpassed the other possibilities when combined with the strategy A and the version
Benchmark, notice that for every other combinations, in 2011, ordering the sectors by their Revenue presented better results. The same happened in 2013.

- In 2014, the results suggest that using the Earnings Per Share or the Earnings Per Share Growth for sector ordering provide better returns.
- The best minimum results were obtained when employing the strategy A.
- The best minimum results are registered when employing the simulation version Benchmark for half the periods, while the best maximums were all registered with the Benchmark version.
- For bad years, like 2011, both other versions, Minimum Exit and Timing, present similar returns (slightly lower), and better minimum values.
- The variation values for the same period are similar, regardless of the combinations used for the simulation.

Next, two different cases are described, along with the detailed results and conclusions regarding the quality of the solution.

### 4.3.2 Case Study I - Benchmark A Revenue-Growth

The first case study represents the best results obtained. As mentioned in the previous section, the best results were obtained when the combination employed was: Version Benchmark + Strategy A + Sectors ordered by Revenue Growth. Notice that all results are related to an one year period, this is, the real simulation occurs from the January 1st to December 31st for example.

In the figure 51 the reader can observe a portion of the table represented in figure 48 concerning the case study I. For comparison purposes, the results for the S&P 500 index are also represented in this table, along with the improvement obtained when comparing with the results obtained in this simulation.

As the reader can see, the implemented solution surpassed the index every year. Also, the variance observed, was lower than the variance suffered by the index, which implies that the Case Study I provides more stability and less risk. Finally, the Sharpe ratio registered for the case study are higher, meaning that when adjusted to risk, the Case Study I investment is more attractive.

#### 4.3.2.1 Return on Investment (ROI)

The values already presented are mean values calculated with the results obtained from 50 executions of the same simulation. In the following images are exhibited the results obtained with this strategy and the distribution of results for the 50 executions.

In the figures 52 and 53 are represented the ROI evolutions for each one of the periods. The curves in the graph represent the mean execution, the best execution and the S&P 500 index performance.
Figure 52: Case Study I - ROI per year for the S&P 500, the mean and best execution - 2011 & 2012
In all four graphs, the curve representing the S&P 500 index and the curves representing the simulations are similar, although a few “percentage levels” above. This behaviour suggests that the solution is very correlated with the market. Nonetheless, the results are very positive, which can be explained by the fact that the algorithm is buying companies from the sectors presenting higher revenue growth, and by doing that it achieves higher returns than the index.
In figure 54 is represented the graph for the 4 year period. As the reader can see, the mean solution strongly surpasses the index, ending up with almost twice the index’s ROI. The best solution triples the index’s ROI. Of course, the best case is a one time situation, although it happened, so it means that is possible to achieve those results with the right combination of ratios, weights and strategies. Nonetheless, the mean solutions results are very satisfying.

The figure 55 is an histogram to demonstrate the ROI distribution for the 50 executions.
Notice that, every year, the ROIs registered with higher frequency in 50 simulations are higher than the ROI of the S&P 500 index. As a matter of fact, the number of executions producing worst ROIs than the index is very low, which means that this case study totally surpassed the index.

4.3.2.2 Variance

The mean variance of the executions for these case study is better than the variance displayed by the S&P 500 index. However, for a better comprehension of the solution quality, in the figure 56 are exhibited the ROI-Variance results for all executions, along with the index result.
Through the analysis of the graphs in the figure 56 the reader can notice that generally the case study performs better. For the periods 2012 and 2013, the variance values are very similar, but the ROIs produced by the case study are considerably higher. In the remain periods, 2011 and 2014, almost 100% of the cases present lower variance and better ROI. With that being said, the case study proves to present better returns for the same risk, and in some cases better return with lower risk.

4.3.2.3 Quantitative Metrics

The following images, 57 and 58, provide the performance according with the quantitative metrics established previously in this chapter. In figure 57 are the measures and metrics for the S&P 500 index and the average GA execution. To complete the study, the figure 58 shows the measures and metrics for the 50 executions and for the best one. Notice that in this second table the values are in form of interval, presenting the minimum and the maximum value registered in the 50 executions.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>S&amp;P 500 Index</th>
<th>Avg. GA Execution</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROI (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.015</td>
<td>0.008</td>
</tr>
<tr>
<td>Sharpe Ratio</td>
<td>-0.07</td>
<td>0.40</td>
</tr>
<tr>
<td>Annual Return (Avg %)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum Draw Down (%)</td>
<td>-18</td>
<td>-4</td>
</tr>
<tr>
<td>Number of Positions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Profitable Decisions (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transactions Costs (K $)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 57: Case Study I - Quantitative Metrics 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>GA (50 Executions)</th>
<th>Best GA Execution</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROI (%)</td>
<td>[-1.97, 15.59] [14.42, 30.96] [17.51, 39.41] [9.59, 28.18]</td>
<td>15.59</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>[0.013, 0.016] [0.006, 0.008] [0.004, 0.007] [0.005, 0.008]</td>
<td>0.014</td>
</tr>
<tr>
<td>Sharpe Ratio</td>
<td>[-0.27, 0.44] [0.79, 1.95] [1.49, 2.56] [0.58, 1.96]</td>
<td>0.44</td>
</tr>
<tr>
<td>Annual Return (Avg %)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Maximum Draw Down (%)</td>
<td>[-16, -12] [-5, -2] [-4, 0] [-6, -3]</td>
<td>-15</td>
</tr>
<tr>
<td>Number of Positions</td>
<td>[73, 1349] [25, 1049] [99, 1301] [142, 1222]</td>
<td>234</td>
</tr>
<tr>
<td>Profitable Decisions (%)</td>
<td>[46, 74] [67, 82] [71, 86] [59, 73]</td>
<td>72</td>
</tr>
<tr>
<td>Transactions Costs (K $)</td>
<td>[10, 90] [11, 86] [13, 89] [12, 90]</td>
<td>74</td>
</tr>
<tr>
<td>Sector Distribution</td>
<td>[[0%, 25%], [75%, 100%]]</td>
<td>[25%, 50%] [50%, 75%] [25%, 50%] [25%, 50%]</td>
</tr>
<tr>
<td>Beta</td>
<td>[LOW, HIGH]</td>
<td>HIGH</td>
</tr>
</tbody>
</table>

Figure 58: Case Study I - Quantitative Metrics 2

91
4.3.2.4 Chromosome example

The “winning” chromosome for 2013, along with the average chromosome, are illustrated in the figure 59. All genes are discriminated in the table.

![Figure 59: Case Study I - Chromosome example and best individual values.](image)

Looking into the genes of the chromosome which triggered the best execution, and comparing with genes from the average execution, its possible to deduct a few conclusions:

1. A key factor for the positive results is the limit imposed on the company selection. Only 4 out of 10 sectors are available.

2. The ideal portfolio’s size shall be around 10-15 assets. Focusing the investments in a small number of companies from the best sectors is more lucrative than spread the investment to much.

3. The “winning” fundamental ratios are: debt ratio, price to earnings growth, return on equity, revenue growth rate and revenue rate. All of them are growth potential indicators, which implies that the best results were achieved when money was invested in companies presenting high growth potential.

4.3.2.5 Results discussion

The Case Study I is proven to be a superior strategy. It is demonstrated that it offers higher profits and more stability than the S&P 500 index.
From the ROI evolutions charts (figures 52 and 53), the reader can observe that, despite the fact that both index and evolutionary strategy follow the same trends, the average execution of the evolutionary algorithm surpassed the index. This is justified with three reasons: (1) although the trends are similar, when the index curve is going up, the GA curve goes up further; (2) in several occasions, in a falling scenario, the GA curve can resist the down trend better than the index; (3) the GA performs intelligent investment decisions, this is, by choosing the best sectors, the ROI receives high boosts when in an upward trend, surpassing the S&P 500 index growth. Notice that the best GA execution is much more profitable than the index or the average execution. To future work, a study shall be performed about the decisions made in the best execution, in order to approximate the average curve to this one.

Looking into the variance graphs in figure 56, is notorious the supremacy of the GA solution. Almost every execution provides more stability for the same revenue, being that an high percentage of them even provides higher revenue for lower variance (risk).

4.3.3 Case Study II - Minimum Exit B Revenue

The second case study represents a different approach. The combination employed was Version Minimum Exit + Strategy B + Revenue. This case did not performed as good as the first case study, although it has surpassed the S&P 500 index in some periods. The reason to look more closely at this case is to show the reader a different combination of parameters, that despite the fact it was not the best, also performed good. As in the previous case study, all results are related to an one year period.

In the figure 60 the reader can observe the part the table represented in the figure 48 correspondent to the Case Study II. Again, for comparison purposes, the results for the S&P 500 index are also represented in this table, along with the improvement obtained when comparing to the results obtained in this simulation.

<table>
<thead>
<tr>
<th>Measures</th>
<th>S&amp;P 500 Index</th>
<th>Avg. GA Execution</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROI (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.015</td>
<td>0.008</td>
<td>0.007</td>
</tr>
<tr>
<td>Sharpe Ratio</td>
<td>-0.07</td>
<td>0.40</td>
<td>2.01</td>
</tr>
</tbody>
</table>

Figure 60: Case Study II - Validations results comparison

As the reader can see, the implemented solution surpassed the index in the first two years, 2011 and 2012, and presented worst results in 2013 and 2014. The variance observed was lower than the variance suffered by the index, which implies that the Case Study II, similarly to the Case Study I, provides more stability and less risk. Finally, the Sharpe ratio registered for the case study follows the ROI trend, this is, is better in the first two periods and worst for the last two. This results suggests that the case study works better when the market is more volatile and presenting accentuated value falls.

4.3.3.1 Return on Investment (ROI)

The values already presented are mean values calculated with the results obtained from 50 executions of the same simulation. In the following images are exhibited the results obtained with this strategy and the distribution of results for the 50 executions.

In the figures 61 and 62 are represented the ROI evolutions for each one of the periods. The curves in the graph represent the mean execution, the best execution and the S&P 500 index performance.
Figure 61: Case Study II - ROI per year for the S&P 500, the mean and best execution - 2011 & 2012
Similar to the first case study, the solution follows the index trends, however in this case the curves representing the S&P 500 index and the curves representing the simulations connect in several points. This behaviour suggests that the solution is correlated with the market, but in some periods the market actually performs better.
In figure 63 is represented the graph for the 4 year period. As the reader can see, the mean solution is very close to the index. The best solution doubles the index's ROI. The figure 64 is an histogram to demonstrate the ROI distribution for the 50 executions.
Notice that, in 2011 and 2012, the ROIs registered with higher frequency in 50 simulations are close or higher than the ROI of the S&P 500 index. Although, in the other two cases, is notorious that the index performed better. As the reader can see, the number of executions which outperformed the index is very diminished.

4.3.3.2 Variance

For a better comprehension of the results, in the figure 65 are exhibited the ROI-Variance results for all executions, along with the index result. One more time, the study case achieved better results for 2011 and 2012, but the S&P 500 index presented itself as a safer investment decision in 2013 and 2014.
Through the analysis of the graphs the reader can notice that, for the periods 2011 and 2012, the variance values are very close, but the ROIs produced by the case study are higher. In 2013, half of the executions exhibited less variance, but also lower returns. The other half produced more money, but shouldering higher risks. In 2014 nearly 90% of the executions produced lower returns, although the variance presented was also lower.

4.3.3.3 Quantitative Metrics

The quantitative metrics are the most relevant aspect in this case study. As the reader had the opportunity to verify, this solution did not achieved great results when comparing to the index when it comes to returns and volatility. Although, because this is a different version and strategy that the one employed in the first case study, the metrics’ values are very different. The following images, 66 and 67, provide the performance according with the quantitative metrics established previously in this chapter. In figure 66 are the measures and metrics for the S&P 500 index and the average GA execution. To complete the study, the figure 67 show the measures and metrics for the 50 executions and for the best one. Notice that in this second table the values are in form of interval, presenting the minimum and the maximum value registered in the 50 executions.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>S&amp;P 500 Index</th>
<th>Avg. GA Execution</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROI (%)</td>
<td>0</td>
<td>13.41</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.015</td>
<td>0.008</td>
</tr>
<tr>
<td>Sharpe Ratio</td>
<td>-0.07</td>
<td>0.40</td>
</tr>
<tr>
<td>Annual Return (Avg %)</td>
<td></td>
<td>13.6</td>
</tr>
<tr>
<td>Maximum Draw Down (%)</td>
<td>-18</td>
<td>-4</td>
</tr>
<tr>
<td>Number of Positions</td>
<td>-</td>
<td>22</td>
</tr>
<tr>
<td>Profitable Decisions (%)</td>
<td>-</td>
<td>57</td>
</tr>
<tr>
<td>Transactions Costs (K $)</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>Sector Distribution</td>
<td>-</td>
<td>[50%, 75%]</td>
</tr>
<tr>
<td>Beta</td>
<td>-</td>
<td>HIGH</td>
</tr>
</tbody>
</table>

Figure 66: Case Study II - Quantitative Metrics 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>GA (50 Executions)</th>
<th>Best GA Execution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Deviation</td>
<td>[0.012, 0.015]</td>
<td>[0.006 0.008]</td>
</tr>
<tr>
<td>Sharpe Ratio</td>
<td>[-0.25, 0.38]</td>
<td>[0.35, 1.27]</td>
</tr>
<tr>
<td>Annual Return (Avg %)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Number of Positions</td>
<td>[11, 52]</td>
<td>[16, 56]</td>
</tr>
<tr>
<td>Profitable Decisions (%)</td>
<td>[30, 73]</td>
<td>[60, 84]</td>
</tr>
<tr>
<td>Transactions Costs (K $)</td>
<td>[3, 5]</td>
<td>[3, 5]</td>
</tr>
<tr>
<td>Sector Distribution</td>
<td>[[0%, 25%], [75%, 100%]]</td>
<td>[50%, 75%]</td>
</tr>
<tr>
<td>Beta</td>
<td>[LOW, HIGH]</td>
<td>HIGH</td>
</tr>
</tbody>
</table>

Figure 67: Case Study II - Quantitative Metrics 2

99
Analysing the tables in the figures 66 and 67 there are two big differences: (1) the number of positions; (2) the percentage of profitable decisions. Running the software with the Minimum Exit version changes the number of positions drastically, reducing it in almost 10 times. With the reduction of buys/sells, the transaction costs also dropped significantly. Most importantly, the accuracy improved, achieving 84% in 2012 and 93% in 2013. With this level of accuracy, different approaches could be studied, using leverage techniques for example.

4.3.3.4 Chromosome example

The “winning” chromosome for 2013, along with the average chromosome, are illustrated in the figure 68. All genes are discriminated in the table.

<table>
<thead>
<tr>
<th>Genes</th>
<th>Chromosome Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average (50 Executions)</td>
</tr>
<tr>
<td>Current Ratio</td>
<td>2.31</td>
</tr>
<tr>
<td>Debt Ratio</td>
<td>4.33</td>
</tr>
<tr>
<td>Earnings Per Share</td>
<td>4.82</td>
</tr>
<tr>
<td>Net Income Growth</td>
<td>0.51</td>
</tr>
<tr>
<td>Price to Book Value</td>
<td>4.34</td>
</tr>
<tr>
<td>Price to Earnings</td>
<td>2.45</td>
</tr>
<tr>
<td>Price to Earnings Growth</td>
<td>1.41</td>
</tr>
<tr>
<td>Profit Margin</td>
<td>2.29</td>
</tr>
<tr>
<td>Return on Assets</td>
<td>0.12</td>
</tr>
<tr>
<td>Return on Equity</td>
<td>4.91</td>
</tr>
<tr>
<td>Revenue Growth Rate</td>
<td>4.11</td>
</tr>
<tr>
<td>Revenue Rate</td>
<td>4.83</td>
</tr>
<tr>
<td>Price to Earnings Future</td>
<td>1.75</td>
</tr>
<tr>
<td>Portfolio Size</td>
<td>26</td>
</tr>
<tr>
<td>Gain Stop</td>
<td>50</td>
</tr>
<tr>
<td>Number of Sectors</td>
<td>6</td>
</tr>
<tr>
<td>Time to Sell</td>
<td>44</td>
</tr>
<tr>
<td>Min Jump Entry</td>
<td>25</td>
</tr>
<tr>
<td>Min Jump Exit</td>
<td>24</td>
</tr>
<tr>
<td>Buy Signal</td>
<td>53.15</td>
</tr>
</tbody>
</table>

Figure 68: Case Study II - Chromosome example and best individual values.

Looking into the genes of the chromosome which triggered the best execution, and comparing with genes from the average execution, it's possible to deduct a few conclusions:

1. The requirement of a bigger jump to exit can have a negative impact on the performance.
2. Smaller portfolios produce better results. Focusing the investments in a small number of companies from the best sectors is more lucrative than spread the investment to much.
3. The “winning” fundamental ratios are: debt ratio, net income growth, return on assets, revenue growth rate and revenue rate. As it happen in Case Study I, all winning indicators are growth indicators, which implies that the best results were achieved when money was invested in companies presenting high growth potential.

4.3.3.5 Results discussion

The Case Study II’s results can be separated in two periods. In 2011-2012 and 2013-2014. In the first period, it has performed better than the S&P 500 index. On the second period, although not by much, the solution was surpassed by the index’s performance.

From the ROI evolution charts (figures 61 and 62) and variance charts (figure 65) is possible to verify the conclusion stated above. This behaviour can be justified by the change of strategy. In this case, the simulations were performed employing the version Minimum Exit, which dictates that to sell an asset, the company must have fallen several positions in the rank. By doing this, it may happen that a company is causing a negative impact in the portfolio’s value, but the simulator is not allowed to sell it. When the company is finally ready to be sold, the price is too low and large amounts of money are lost.

However, positive aspects can be observed: the maximum draw down is lower using this strategy, which indicates a better resistance to crashes and bearish markets; the number of positions (and money paid in transactions) is significantly lower; once the number of buy/sell operations is reduced, the sector distribution is more accurate and is kept for almost the entire simulation; the percentage of profitable decisions increased.

It is possible to conclude that, to future work, is important to add more clauses to the sell verifications, not having a general situation. Stopping the simulator from selling assets is a very dangerous limitation, and should not be based on a single value (the minimum jump to exit gene).

4.4 Conclusion

In this chapter was described how the system was validated and tested. A description and motivation for the three measures used can be found in the first sections of this chapter, followed by a statement of several qualitative metrics chosen to classify the system. In the second part of the chapter, an overview of the results is presented, along with a more detailed view on two case studies and the conclusions that can be derived from the results obtained in each case.
5 Conclusions and Future Work

From the several sections in this project is possible to understand the work’s motivation, the objectives aimed, the financial concepts related to the work and the several existing approaches to solve the same or similar problems. On chapter 2 several related works were presented. It is possible to observe that the use of evolutionary techniques is not a new thing. The optimization of financial portfolios using artificial intelligence gained many practitioners and although it is hard to evaluate and compare this kind of systems, it is possible to conclude that the results obtained using this techniques surpassed the competitor methodologies. In chapter 3 and 4 is the description of the implemented system and the validation method respectively. The system presents three innovations: using genetic algorithm combined with fundamental analysis, including fundamental ratios created by us for this work; stock picking giving special relevance to the sectors and industry classification; use of sectoral indicators, instead of indicators of the overall market.

In this chapter the reader have access to the general conclusions from this work. The achievements are clearly stated and described in the second section. The last section addresses the possible future work. In this section are presented the current limitations of the system, the possible improvements and the possible extensions.

5.1 Conclusion

The analysis of the results obtained in this work allows the conclusion that the Evolutionary Algorithm plus Stock Market combination is a very powerful tool when it comes to the management of financial portfolios. The ability to solve and optimize problems with thousands of variables in a few seconds demonstrated great potential when applied to the buy/sell of companies’ assets.

As demonstrated in the chapter 2, a lot has been already explored in this field. Nonetheless, that is just the tip of the iceberg, much more is still available and there are many ways to innovate. In this work the reader has at his disposal the results obtained from a software combining genetic algorithms with the American stock market, more precisely, companies from the S&P 500 index from 2010 to 2014. Several strategies were tested and documented. The validation performed on those strategies, in the two case studies provided more specifically, showed that the approach chosen for this work has a lot of potential. In fact, the results obtained are already superior to the results registered by the overall market. In conclusion, dividing the market by industries/sectors has proven to be a reliable and profitable game plan. Exploring sectors presenting higher revenue and revenue growth provide investors with good and more stable returns. Additionally, this strategy successfully selects the growing companies, based on fundamental ratios.

However, things can be improved and the average results may come close to the best results obtained. It is important to keep in mind that machines cannot fully replace the human investor, as some things are not digitalized and may escape the system’s perception, resulting in dreadful decisions. Although the results obtained using this application in full automatic mode are satisfying and motivational, results are expected to improve if some decisions could be made by the user in real time.

5.2 Achievements - Innovations Left

In this subsection the reader can see the innovations and objectives accomplished.

- Successful implementation of an investing strategy supported on an market analysis divided into sectors.
- Lucrative use of indicators concerning industries/sectors instead of using the usual global indicators.
- Creation and implementation of two original fundamental ratios prospecting for growth: Revenue Rate and Price to Earnings Future. These ratios were used in this work and had a relevant role to play on the stock picking operations.

5.3 Future Work

In this section are described several current limitations of the system, accompanied with the respective possible improvements. To conclude, a list of possible extensions is provided.
5.3.1 Current Limitations

- The data stored is a big “problem” in this work. The problem is that the data is not freely available on the web. Clean and organized companies’ fundamental statements are sold for high prices. In the best case, the last year values can be found for free. Although, for a work of this nature, information for a five year period is the minimum acceptable. For this motive, a new software was implemented to download the fundamental statements and further process to cleaning them. The result was very satisfying, but far from perfect. Some fields could not be found, meaning that for a small set of companies the statements were incomplete. This may have impaired the results of the genetic algorithm.

- This system do not contemplate a stop loss signal. The stop loss signal represents the maximum percentage the investor is willing to lose on a given period. For example, assuming a stop loss of 5%, if in a three day period the portfolio’s value drops 5%, all positions are closed immediately. This constraint could have improved the results in some cases. In the case studies presented in this document is possible to observe from the graphs that at the end of the certain periods exist an accentuated fall, which could be avoided using a stop loss signal.

- The current system only considers companies from the S&P 500 index. Due to the system’s philosophy of investing by industry/sectors, and because the industry division used is the classification provided by the S&P 500, the software is constructed to use only companies within the same index at a time, the S&P 500 in this case.

- The current system only permits to adopt one strategy during a certain period. The strategy (constituted by a version, investing strategy and sector sorting) is decided at the beginning of the software execution and is kept until the end.

5.3.2 Possible Improvements

- Optimize the Data Module to include all fields and create complete fundamental statements with 100% accuracy.
- Extend the Data Module to include market data from different indexes.
- Add new fundamental ratios to the chromosome.
- Incorporate macroeconomic indicators and technical indicators in the algorithm.
- Extend the algorithm to allow different strategies during the same test. It would be of great profit if the algorithm could adjust its strategy to the current market situation instead of using the same strategy during the entire period.
- Improve the “environment” of the simulation. New features should be included: allow short selling; implement dividends distribution.
- Explore new risk measures and ratios besides the variance and Sharpe ratio.
- Develop new operations concerning the genetic algorithm, including new selectors, crossovers and mutation operators.

5.3.3 Extensions

In this section are presented the possible extensions to the current software. Are considered as extensions situations where the current software can be used with small updates.

- Download data from the web for any American company of any index. This data includes financial statements and stock market prices.
- Employ different strategies using the portfolio simulator. The software is constructed in a blocks format, which allows to change the “strategy block” and keep the software running.
- Add new indicators to the algorithm. The unique requirement to add a new indicator is the mathematical formula.
References


