Investing in Credit Default Swaps using Technical Analysis Optimized by Genetic Algorithms

João Manuel Ribeiro Cardoso Barata

Thesis to obtain the Master of Science Degree in

Electrical and Computer Engineering

Supervisor: Prof. Rui Fuentecilla Maia Ferreira Neves

Examination Committee

Chairperson: Prof. António Manuel Raminhos Cordeiro Grilo

Supervisor: Prof. Rui Fuentecilla Maia Ferreira Neves

Members of Committee: Prof. Aleksandar Ilic

December 2017

To my Father

Acknowledgements

"Técnico is Técnico, if you found it hard to get in, you will see how hard it will be to get out". These were probably the words I heard the most during my first week in this university, and whoever came up with this was not completely wrong. From deliver countless work projects in an extremely reduced period of time, to have five exams in four days with questions like "If two plus two equals four, why is the colour of the room white?", Técnico has proven to be not an easy thing to do. And I would probably still be attending some 3rd grade class if it were not for some people that I can call my Friends, so I would like to use this Section to express my gratitude to some of them.

To my housemates, Maria João, Sara and Gabriel, for making me grow as a person and for cheering me up whenever I was in a bad mood, home is wherever we are together.

To my friend Miguel Maldonado for being the first friend I made in Técnico, the typical "Alentejano" that never failed on me. To Ricardo Gomes, for being always there to explain everything and for sending me the notes of all the classes that I missed during the semester that I had to abandon the university. To my main group of friends Cristiano Rebelo, Filipe Marques, José Sérgio Vilela, Natanael Castro, Miguel Rodrigues, Mikhail Sobolev, Tiago Machado e Tiago Marques who with me went to countless classes, and miss some whenever it was necessary, who with me made project groups, making the work much easier, who gave me their study notes and took care of mines, who never said no to a beer after an exam that did not went well, but most of all for being always there. To José Guita, for so much more than forcing me to study FTEL when I was convinced I was going to fail, well we did not! To Raul Ribeiro, for the countless rides for late night studies and for maintain always his characteristic sense of humour, even when there are people that clearly cannot use Windows! To Pedro Gouveia, for the company and all the pep talks he made during the time we were writing our thesis, they seem to have worked out.

To my supervisor, Professor Rui Neves, for all the support, patience and guidance he gave to me during the development of this work, and for all the transmitted knowledge, always done with a level of enthusiasm that only people who are really passionate about financial markets can have.

But most of all, to my entire family, for all the support not only during the course but also for being present during the good and bad moments we had since I can remember. However, I have to highlight two very important people, my Godfather for all the lunches, snacks and dinners we had, always in fantastic places, and mostly to the most important person in my life, my Mother. I would be lost without your support, especially after Dad left us.

Thank you.

Abstract

This thesis proposes a system based on Genetic Algorithms, aiming to create an investment strategy that can be applied on the Credit Default Swaps market, with the use of four commonly used technical analysis indicators and an indicator based on daily variation of the spread. Furthermore, the proposed system adds some additional features to the standard GA with the objective of improving the profitability of the strategy. The obtained results show that it is possible to create a profitable strategy using these techniques and that it is possible to take advantage of the volatility in the CDS spreads to obtain high values of Return on Investment, that in a common situation can easily reach up to 50%. Moreover, the results show that the use of an indicator that measures the daily spread variation can be really important in the CDS market.

Keywords

Credit Default Swaps, Genetic Algorithms, Technical Analysis, Investment Strategies.

Resumo

Com esta tese pretende-se criar um sistema baseado em Algoritmos Genéticos, com o objetivo de desenvolver uma estratégia de investimento para ser aplicada ao mercado dos *Credit Default Swaps*, e com o uso de quatro indicadores de análise técnica frequentemente utilizados e um indicador baseado na variação diária do *spread*. Para além disso, o sistema proposto acrescenta algumas funcionalidades ao Algoritmo Genético padrão com o objetivo de melhorar a rentabilidade. Os resultados obtidos mostram que é possível criar uma estratégia rentável usando estas técnicas e que é possível aproveitar a volatilidade dos *spreads* para obter grandes valores de Retorno de Investimento, que numa situação normal pode atingir os 50%. Para além disso, os resultados mostram que o uso de um indicador que meça a variação diária do spread pode ser de extrema importância no mercado dos CDS.

Palavras-chave

Credit Default Swaps, Algoritmos Genéticos, Análise Técnica, Estratégias de Investimento.

Table of Contents

List of Figures

List of Tables

List of Acronyms

AAL	American Airlines Group Inc
AI	Artificial Intelligence
AIG	American International Group
ALCA	Alcadon Group AB
ANN	Artificial Neural Networks
ANOVA	Analysis of Variance
APC	Anadarko Petroleum Corporation
ASFZ	Associates First Capital
ASVM	Adaptive Support Vector Machines
ATNN	Adaptive Time Delay Neural Networks
AUG	Auryn Resources Inc.
BKT	BlackRock Income Trust, Inc.
BMPS	Banca Monte dei Paschi di Siena S.p.A.
BP	Backpropagation Neural Networks
BRG	Brigadier Gold Ltd
CBR	Case-Based Reasoning
CDO	Collateralized Debt Obligations
CDS	Credit Default Swaps
CHK	Chesapeake Energy Corporation
CNG	Canadian Mining Corp
COP	ConocoPhillips
DJI	Dow Jones Industrial
DVN	Devon Energy Corp
EA	Evolutionary Algorithm
EBRD	European Bank of Reconstruction and Development
ECP	Eligible Contract Participant
EMA	Exponential Moving Averages
GA	Genetic Algorithm
HMA	Hull Moving Averages
HMM	Hidden Markov Model
ISDA	International Swaps and Derivatives Association
KOSPI	Korea Composite Stock Price Index
LogR	Logistic Regression Model
MACD	Moving Average Convergence Divergence

MCP	
ML	Machine Learning
NASDAQ	National Association of Securities Dealers Automated Quotations
NN	Neural Networks
NYSE	New York Stock Exchange
OBV	On Balance Volume
OTC	Over-The-Counter
ROC	Rate of Change
ROI	Return on Investment
RSI	Relative Strength Index
SVM	Support Vector Machines
TA	Technical Analysis
TDNN	Time Delaying Neural Networks
TSI	True Strength Index

List of Software

PyCharm	IDE used to develop the strategies
Microsoft Excel	Spreadsheet software
Microsoft PowerPoint	Slide show presentation and schematic design software
Draw.io	Schematic design software
Microsoft Word	Text editor software

Chapter 1

Introduction

This chapter gives a brief overview of this Master Thesis. Before presenting the system architecture and the related literature, the scope and motivations are provided. At the end of this chapter, the work structure is presented.

1.1 Overview

The use of automated trading relying on evolutionary techniques has been in the centre of attention of investors, exploring its predicting capabilities in order to improve the quality of their investments generating higher returns.

Those capabilities rely on the basic idea that “History repeats itself” [1], meaning that they use extensive historical price data to develop technical strategies, first by optimizing the strategy parameters using that data and, once the best strategy is found, it is used to execute automated orders in the market, aiming to achieve higher returns more often. This concept has been used in diverse financial markets, such as the Stock, Bond and Forex Markets [2], and in this work are going to be applied to a particular case of the Derivatives Market, the Credit Default Swaps Market. As its own name implies, the Derivatives Market is composed of securities with a price derived of one or more underlying assets, which for the case of the CDS Market are usually Bonds [3].

CDS have gain the interest of dealers, investors, and regulators since the last decade crisis. A CDS, as it will be explained during this work, can be seen as an insurance that an investor can get to a certain investment, assuring that he will get payed even if the company in which he is investing defaults [4]. The price to pay to insure a company strongly varies with time, that is what makes a CDS contract more or less valuable and consequently give investors the opportunity to trade them to generate profits.

CDS were in the genesis of the 2008 financial crisis, portrayed in the American film “The Big Short” [5]. The movie shows how a set of different investors saw what was happening in the housing market and were able to make huge profits buying insurance (CDS) on mortgage bonds. In the past, everybody payed their mortgages, so the risk on those bonds was considered low, although at some point people started to get seriously indebted, and could not afford to pay that mortgages anymore, making the bond worthless, that started to happen in 2007 and eventually caused the whole system to crash. At this point, their CDS were really valuable, their only concern was that the insurance companies from where they bought the CDS were not able to pay them. For this reason, it is critical to understand the potential for loss before start to trade in the CDS market.

However, the movie and the crisis have shown how profitable trading CDS can be, and combining that with the fact that evolutionary techniques are being widely used for investment improvements with very positive results, it seems possible to use them to generate even higher profits in this market. Hence, the objective of this thesis is to see if it is possible to create a trading strategy for the CDS Market using only its past data, with help of an Evolutionary Computation technique called Genetic Algorithms.

1.2 Motivation

Over the past years, intelligent systems have been widely applied to financial markets, but there is still a lot more that can be done in those fields, either by improving the efficiency of the algorithm or study its application into different markets. When developing an intelligent system for trading in a market this volatile as the CDS market, the ability to achieve a financial gain has to be the main objective, this way the main objective of this thesis is to develop a strategy that is profitable in the considered period of test.

From the engineering side, developing a machine learning based algorithm and apply it in a real application like the CDS market is a challenging work, combining two of the most attractive fields of research, namely computer science and economic research.

Furthermore, the fact that there are not many works relating CDS with ML and, more specifically, there are not works, to the authors knowledge, using ML to create trading strategies in this market appears as an opportunity to do something new in both fields.

1.3 Contributions

The main contributions of this thesis are:

1. Apply Evolutionary Computation to trade in a relatively new and little explored market as the CDS Market.
2. Use Genetic Algorithms to trade in the CDS Market and compare the results with strategies based on unoptimized parameters;
3. Make use of daily price variations as a technical indicator and confirm its relevance for trading in markets with high fluctuation in prices;
4. Add some features to the original GA and confirm if those features have consequences in the profitability of the algorithm.

1.4 Contents

The remainder of the thesis is structured as follows:

- Chapter 2 presents the theory behind the developed work, such as the introduction to Credit Default Swaps, the investment strategy used which relies on technical analysis and the fundamental concepts of genetic algorithms. Furthermore, the related work developed over the last years is presented here.

- Chapter 3 contains the explanation of the proposed system architecture, explaining the main modules of the system and its behavior. The options made for the GA parameters are presented as well as the used technical indicators.
- Chapter 4 shows the performed system validation through one case study on the CDS market plus three examples of different performing companies of this market. There is also presented the performance meters used to evaluate those case studies as well as some implemented strategies used to compare the performance of the system.
- Chapter 5 summarizes the work and presents some suggestions of future work.

Chapter 2

Background and State-of-the-Art

This chapter will be used to present the fundamental concepts applied in this thesis, namely financial knowledge and the fundamental concepts of genetic algorithms. This chapter will give also an insight on some related works developed in the recent years.

2.1 Overview

This chapter will present the financial market in which the algorithm to be developed will be applied, namely the Credit Default Swaps Market. Afterwards it will be done a presentation of the two most used methods to perform a market analysis, and the choice for the Technical Analysis will be justified. It will also be done a presentation of the evolutionary algorithm that will be used in this thesis, the Genetic Algorithm, and its main stages. To conclude this chapter, it will be presented some papers found in the literature on the themes addressed in this thesis, aiming to justify some of the choices made during the development of the algorithm.

2.2 Background

As stated before, this thesis will present an approach to trading in the Credit Default Swaps market using Technical Analysis and a machine learning technique called Genetic Algorithm. To a better and easier comprehension, this sub-chapter will present some of their key aspects.

2.2.1 Credit Default Swaps

The history of Credit Default Swaps [6], or CDS, goes back to 1994 when J. P. Morgan made a contract with the European Bank of Reconstruction and Development (EBRD) on a $4.8 billion credit line to ExxonMobil. In this contract was established that J. P. Morgan would pay a periodic fee to EBRD until the end of the contract or until the default of ExxonMobil, and in this case EBRD would cover J.P. Morgan's loss.

Succinctly, a Credit Default Swap is a credit derivative contract between two parties, a Protection buyer and a Protection seller. The latter, sells the CDS providing the buyer protection against specific risk of default of a third party, a Company, in exchange for a fee or a premium, called the CDS spread, over the maturity of the CDS contract [4], as illustrated in the Figure 2-1. At first, insurance companies were the main CDS protection sellers, while commercial banks were the main buyers, however hedge have increased their participation in the market [7]. CDS are the most frequently traded credit derivative.

CDS are designed to cover many different types of credit events, including, bankruptcies, credit rating downgrades and defaults, which happen when a company fails to make timely payments on debt. Commonly, CDS have low transaction costs and have a maturity of one to ten years with most of the liquidity concentrated on the five-year horizon [4]. In this situation, the contract will be open for five years, during which the protection buyer pays the referred premiums and, if there is no credit event during that period, the protection seller gets to keep them.

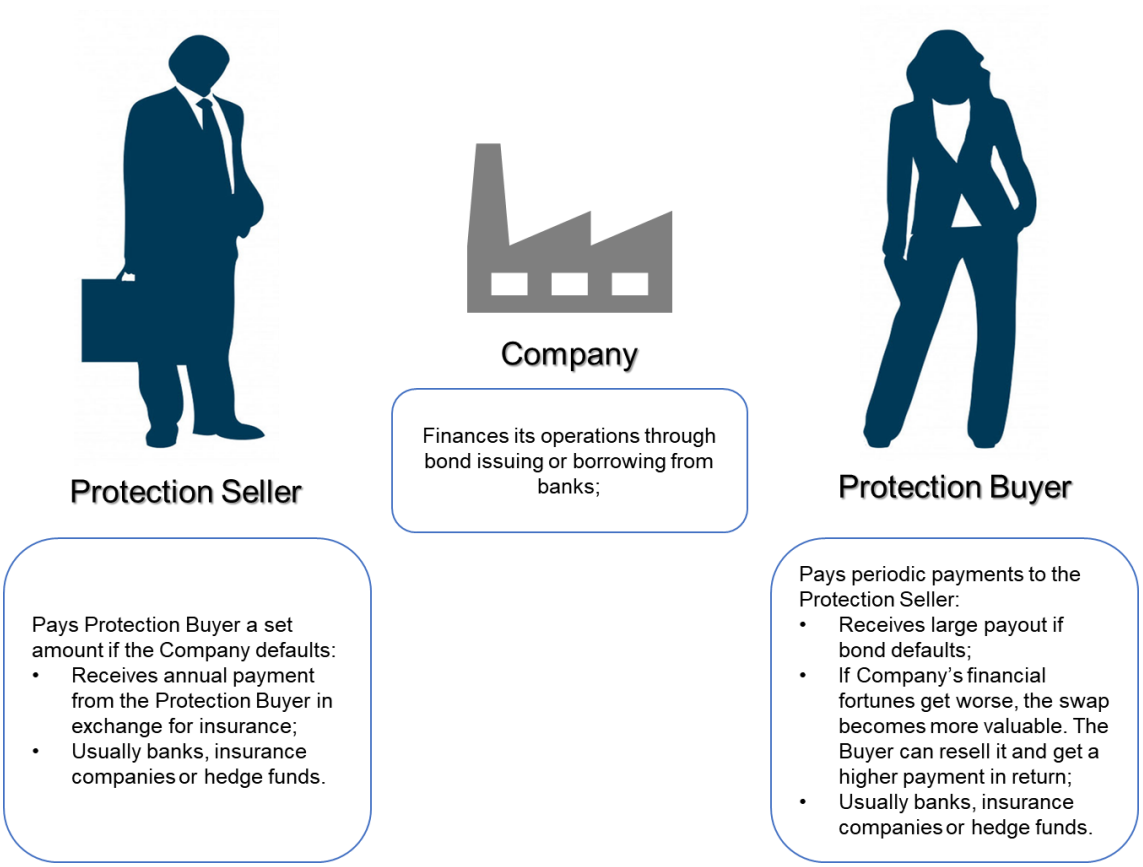

Figure 2-1: Intervenients in a CDS contract.

The standardization of CDS contracts was stipulated by the 1999 International Swaps and Derivatives Association (ISDA) Master Agreement, with constant reviews and publication of new definitions on the subsequent years. The CDS market has grown exponentially during the first decade of this millennium due either to these constant publications by ISDA and due to the emergence of new types of credit derivative products, including synthetic Collateralized Debt Obligations, or CDOs, for which CDS contracts are a crucial element [7]. Essentially CDOs repackage individual loans, like credit card debts, mortgages or corporate debt, and sell it to investors.

This growth allowed CDS to play a significant role in the credit crisis of 2007/2009 and the European sovereign debt crisis in 2010/2012, having a major contribution to the collapse of the world's largest insurance company, the American International Group, or AIG. At the time, AIG was selling large quantities of CDS, either to investors, aiming to insure a financial instrument, and speculators, to bet against other instruments that they did not own. Unlike traditional insurance, where someone can only insure something that he owns, in the derivatives universe everybody can insure something they do not own. Since CDS were unregulated and since they were sure that nothing relevant would happen to what they were insuring, namely the real estate market, AIG did not have to put aside any money to cover potential losses. This AIG credit default swap exposure resulted on an outsized bet against real estate, meaning that a decline in real estate prices would impose massive costs on AIG, which eventually happened and AIG was forced to pay redemption requests, what generated a large need for liquidity, forcing a government bailout [8] [9].

The default of a protection seller, like AIG, can affect many market participants and generate domino effects and default contagion, once in the presence of a CDS market the default of an entity incurs losses not only for its counterparties but also for protection sellers in CDS written on this entity. In the presented example, if a protection seller has insufficient reserves to cover its CDS responsibilities, the underlying credit event also results in the default of the protection seller, widening the scope for contagion [4].

After that period of crisis, there have been calls for tougher regulations and even an outright ban on certain types of CDS transactions, once until recently the CDS market was highly unregulated and the transactions were conducted mostly over-the-counter (OTC), meaning that they were traded directly between two parties, without the supervision of an exchange, such as NYSE or NASDAQ. The main market participants are almost entirely institutions once the market is limited only to Eligible Contract Participants (ECPs).

CDS trading is very complex, risky and, combined with the fact that credit default swaps are traded OTC, this market is susceptible to a high degree of speculation. In fact, as referred in the AIG case, speculators often buy Credit Default Swaps in bonds they do not own, meaning that they can collect the value of the CDS if the company defaults, without the risk of losing money on the bond. This particular type of CDS is usually called naked CDS.

Furthermore, it has been proven that CDS spreads start to change about two or three months prior to negative credit rating announcements [10], which gives it a predictive power for future stock returns, especially when it comes to negative changes to the credit quality of the firms. This is due mostly to the fact that the CDS market quickly and accurately incorporates public information, meaning that a firm with a high news intensity is likely to have a change in its spread, which eventually will have repercussions in the firm's value.

2.2.2 Investment Strategies

In order to find to the best trading opportunities, investors commonly perform a prior analysis to the market, whichever it is, seeking for the potentially best assets. The two most used methods are Fundamental Analysis and Technical Analysis, which give two different approaches to the prediction of the prices evolution problem.

- *Fundamental Analysis*

According to Fundamental Analysis, to understand what will happen to the value of an entity, one must perform a study on anything that can affect the entity value, known as the company fundamentals. Those include the overall economy and industry status and the company management and publicly available information, such as its past balance sheets [1]. Taking on these balance sheets, analysts can calculate several ratios from which one can see if the company's income is growing and if the company is able to pay its bills and still be profitable, and if this profit has been increasing for the last periods. Those ratios may be also compared between companies, usually from the same sector of activity.

Finally, based on these calculations, one can predict the behaviour of a company's value, and considering its future expectations of growth and profitability decide whether to invest or not, or choose the best of two companies of the same sector.

- *Technical Analysis*

On the other hand, Technical Analysis states that the price evolution of a certain asset is predictable through a statistical analysis of its past data, especially its past value and the volume of transactions, once they contain all the information that can affect the market's price, including even the fundamental factors [11]. Technical analysts strongly believe that history tends to repeat itself in terms of price movements.

Considering the two mentioned factors, the past value and volume, one can perform technical analysis in several ways, either by relying on chart patterns or by using technical indicators or oscillators. A technical indicator is a mathematical formula used to summarize all the relevant information of the past history of a financial time series into short-term statistics, which returns a value for each day and determines if an asset is under or overvalued. At the same time, through the analysis of charts, analysts can find certain patterns which may inform whether the trend is going to continue upwards/downwards or if it is about to turn. Trend lines are also commonly used to see the prevailing direction of the price, along with support and resistance lines, which connect the lower and higher values respectively, defining a region where prices are most likely to be, meaning that a price is most likely to fall after touch the resistance line and to rise after touch the support line [1].

Technical indicators fall mainly into two categories, momentum and trend indicators, which analyse different phases of the market behaviour and from different perspectives. A trend following indicator tries to identify a trend in the market, giving to the investor the big picture of what is happening by smoothing the price data and remove noise, following the current direction of the market with a lag. The simplest and most common trend indicator is the Moving Average. Alternatively, a momentum based indicator are used as leading indicators in order to identify if the current trend is losing strength or not, allowing analysts to anticipate the market by giving early entry/exit signals.

In his book, John J. Murphy [11] stated that the fundamentalist studies the cause of market movement, and the technician is more interested about the effect. Besides, Fundamental Analysis requires not only a better training from the user but also a wider data set, including accountable and financial information which are hard to obtain and usually unreliable, once a company might omit some information due to its self-interests [12]. Considering this, in this work it will only be used the Technical Analysis once we are not interested in the reason why the price is going higher or lower, we are only concerned in make a correct prediction of that movement. Although, these two are not excusive, meaning that can be used together [11], as seen in [13], [14].

The technical indicators used in this thesis are the Exponential Moving Average (EMA), the Double Crossover Method, the Relative Strength Index (RSI) and the Rate of Change (ROC). It will be also introduced an indicator based on the daily variations of the prices called Percent Changes. These indicators will be presented in depth in Chapter 3.4.

2.2.3 Genetic Algorithms

In the Financial Market, investors keep on looking for patterns that helps them decide if make a certain investment is a right thing to do. In the case of financial applications, Genetic Algorithms have been widely used to optimize trading strategies with very promising results, namely in the stock market. These results also encouraged its application in this thesis.

A Genetic Algorithm [15] corresponds to a search technique used to find optimal or sub-optimal solutions to search problems. Its behaviour is inspired on Darwin's Theory of natural selection, by defining an initial set of random solutions, which are iteratively refined, until an optimal or a sub-optimal solution to the problem is encountered.

In a GA, this set of random solutions is represented by a population of chromosomes (might also be called individuals), where each chromosome represents a potential solution. These chromosomes are defined by a string of bits, usually a vector of floats, the genes. To initialize the search process, it is necessary to create an initial population, which might be done by generating random vectors that belong to the solution space, as many as the desired population size. These chromosomes are then evaluated through a fitness function that attributes each one a score, or a fitness, according to how well a certain chromosome performs and ranks them all from the best to the worst, being the best ones the most likely to be used in the next stages. Here, using the resulting ranking, the GA performs a selection of the top chromosomes and uses them to create a new population through genetic operations, namely crossover and mutation. When crossover is performed, two or more parent individuals are used to reproduce and create an offspring, by means of combination of combination of their bit strings. On the other hand, mutation, analogously to the biological mutation, creates an alteration to one or more genes in a chromosome, which promotes evolution and avoids the algorithm to converge to some local maxima/minima by preventing chromosomes to became too similar to each other. This process is applied to each offspring individually after the crossover exercise and occurs with a small probability, with a typical value of less than 0.1 [16]. Once the GA requires a fixed size population, after the offspring is created it is necessary to fill the rest of the population, which is commonly done by create new random individuals. Each new population of offspring represents a new generation and the whole process must be repeated until a stable solution is achieved, meaning that a GA must have the necessary generations to allow it to happen.

One really simple and widely used example of an application of a GA is the so called One Max Problem [17]. The target to this problem is to achieve one individual which genes are only composed by 1's, starting with a population of random vectors filled with 0's and 1's. Once the initial population is generated the evaluation is performed using a fitness function who is just the sum of the chromosome genes, i.e. the number of 1's. The more 1's a chromosome has, the bigger the fitness value. Then the generational process begins and with a sufficient number of generations one can achieve an individual filled only with 1's, depending of course of the population and chromosome size.

In the Figure 2-2 are represented the main stages of a GA, followed by a more detailed explanation of those stages, initialization, selection, crossover and mutation.

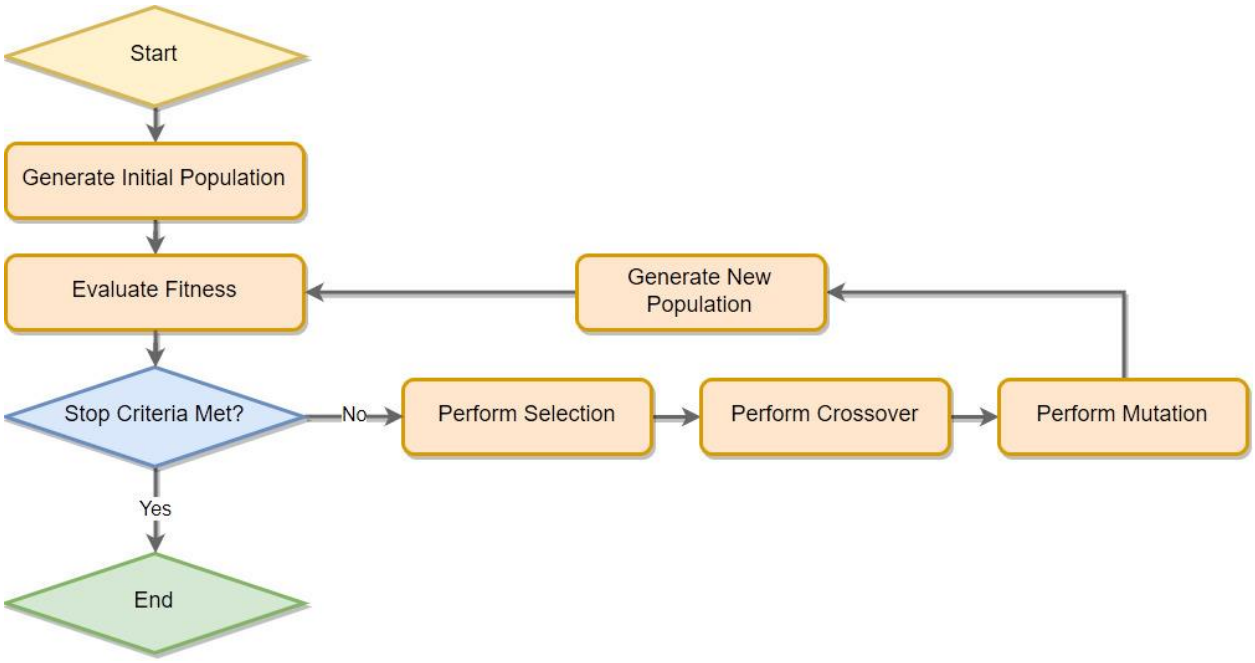

Figure 2-2: Genetic Algorithm flow chart.

- *Initialization*

In this stage, a population of genetic structures, the chromosomes, randomly distributed in the solution space, is selected as the starting point of the search. Each chromosome is composed by a certain number of genes, as represented in the Figure 2-3, where it is represented a population of M chromosomes of N genes. In this thesis, each gene is a float between 0 and 1 and is assigned to a specific parameter that will be optimized.

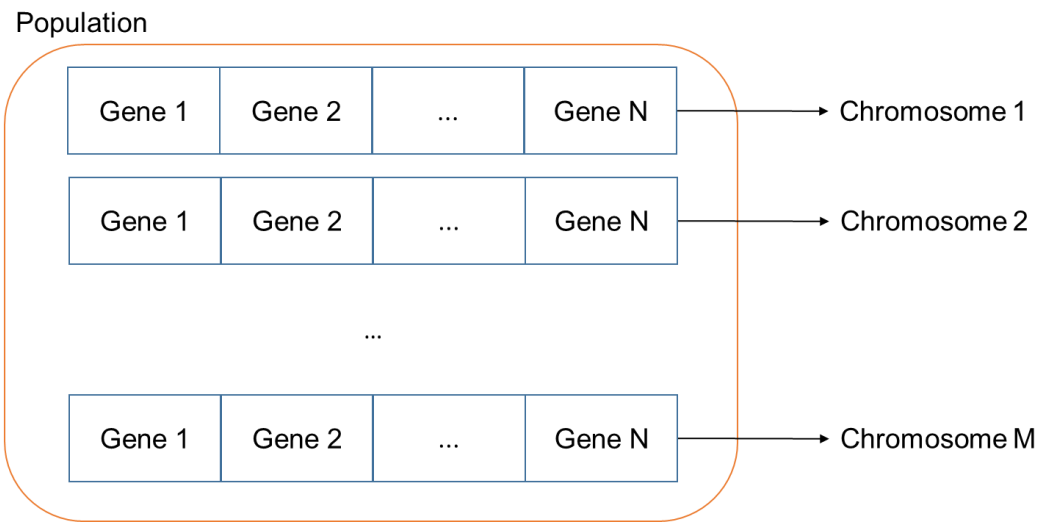

Figure 2-3: A population of chromosomes.

The size of the population has an impact in the performance and efficiency of the genetic algorithm, meaning that a large population although will likely achieve an optimal solution, will probably demand a high computational performance, due to a higher number of chromosomes that need to be evaluated, combined and mutated.

- *Evaluation*

Once the population is generated, each chromosome is evaluated using a user-defined fitness function, which assigns a score (fitness) to each chromosome in the current population, as illustrated in the Figure 2-4. The goal of this fitness function is to encode numerically the performance of each chromosome, which depends on how well each chromosome solves the problem in hand. For real-world applications of optimization methods such as GAs, the choice of the fitness function is the most critical step.

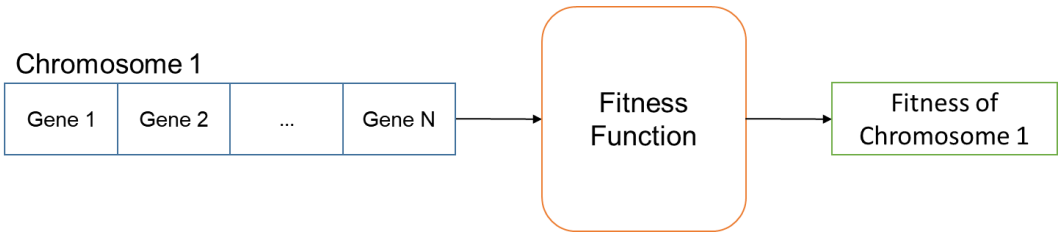

Figure 2-4: Fitness Function.

- *Selection*

In the selection step of the GA, chromosomes are chosen in a way that the high scoring members reproduce, preserving and propagating their characteristics to the subsequent generations. Well performing chromosomes may be chosen for replication several times whereas poorly performing structures may not be chosen at all. However, these worst performing individuals should not be discarded once they not only still have some chances to contribute to useful genetic material, but also ensure that the search process is global and does not converge to the nearest local optimum. Thus, the selection mechanism must consider a compromise between the selection of the best individuals and some of those who do not perform so good [18].

Through the Selection function, in each generation half of the individuals are discarded and the other half is selected to preform Crossover. During the Crossover, each two individuals produce another two individuals, maintaining the number of individuals constant during the whole process. For example, the GA starts with an initial population of 200 individuals, through selection 100 of them are rejected and the other 100, the best performing, are selected for crossover, giving a total of 50 “couples”. For each “couple”, two offspring are produced, meaning that a total of 100 new individuals result from crossover, and joining them to their parents it gives a total of 200 individuals, the number of the initial population.

This process causes an increasing number of best-performing chromosomes to find an optimal solution. There are many techniques to perform selection, for example the elitist selection, where the fittest members of each population are selected and always pass to the next generation, the tournament where the individuals of a population are distributed by groups and compete between them for a place in the next generation or the steady-state selection, in which some individuals (the ones with the highest

fitness) are selected to reproduce, some individuals with low fitness are ignored are replaced by the resulting offspring and the rest of the population survives to the next generation. The choice of the selection strategy used in the GA process must be done carefully, once it might affect significantly the performance of the algorithm.

- *Crossover*

After the individuals are selected, the crossover is performed, forming a new offspring between two randomly selected "good parents". In the crossover process, corresponding segments of a string representation of the parents are swapped, extending the search for a new solution. Crossover only occurs with some probability, the crossover rate, and can be of many different types, namely one-point, multi-point and uniform [19]. As shown in Figure 2-5, at one-point crossover a random crossover point is selected and one parent contributes with his code until that point, and the other parent completes the code.

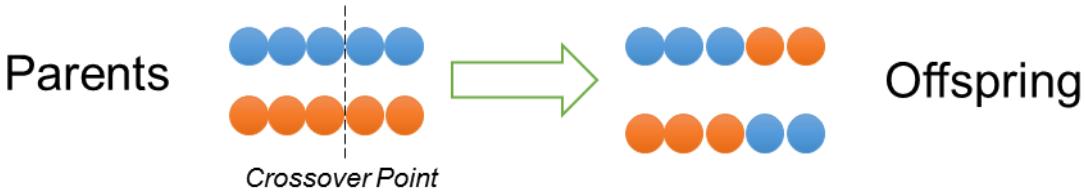

Figure 2-5: One-point crossover.

The same thing happens for the multi point case, shown in Figure 2-6 for two crossing points.

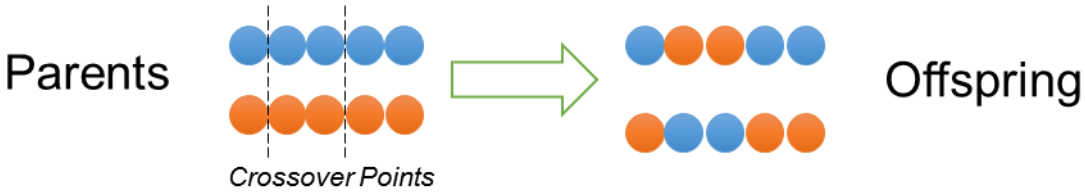

Figure 2-6: Multi-point crossover.

Another type of crossover is the uniform, where the value of the offspring chromosome at some point can be the value of any parent at that location, with equal probability, as illustrated in Figure 2-7.

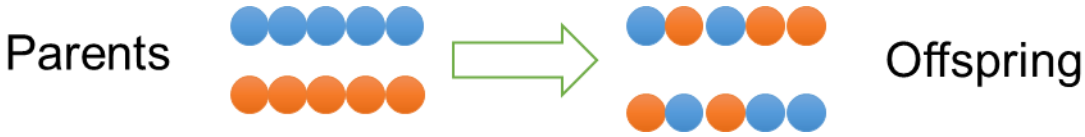

Figure 2-7: Uniform crossover.

- *Mutation*

Finally, mutation is the GA mechanism where an individual of the population is randomly chosen and a randomly chosen gene of this individual is altered to a random value that belongs to the interval of that gene, Figure 2-8. Like crossover, mutation is responsible for the manipulation of the selected individuals to form the subsequent generation.

Although its occurring probability is typically very low, it introduces diversity in the population and ensures that the search for the optimal solution will not stop. Mutation occurs with a certain mutation rate, that defines the probability of each attribute to be mutated. The general rule for mutation operators is that they only mutate, meaning that an independent copy of the individual must be made before the process, in order to be kept as a reference to another individual [20]. Thus, after mutation is performed, the convergence for an optimum is checked and if the mutant member is feasible, it replaces the member which was mutated in the population.

The two most common used mutation methods are Uniform mutation, who replaces the value of the chosen gene with a uniform random value, and Gaussian mutation, who is based on the gaussian distribution and where the gene to be mutated becomes the mean of the distribution and the mutated gene is more likely to be near that point.

Figure 2-8: Mutation.

2.3 State-of-the-Art

After presenting the essential background about the key themes discussed in this thesis, a study of what has been done in previous works is presented in this sub-chapter. Substantial literature can be found in applying technical analysis in trading, the clear majority in stocks market. There will be also presented works on Genetic Algorithms applied to financial markets.

2.3.1 Works on Credit Default Swaps

Although they are a relatively new financial instrument, the remarkable growth of the Credit Default Swaps market has attracted the attention of several researchers, studying its relationship with other markets and the information one can obtain from an analysis of a CDS spread.

In a time where the credit derivatives market was small and still starting to develop, Blanco et al. [21] conducted a study addressing the validity and implications of the theoretical relationship between CDS prices and credit spreads, as the CDS originally came up as an insurance to bonds. A credit spread

represents the difference in yield between a U.S. Treasury bond and bond with the same maturity but with less quality. They found evidence that the CDS market is more liquid than the corporate bond markets, in the sense that new information is reflected in the CDS premia more rapidly than into corporate bond prices. They also examined the determinants of changes in CDS prices, concluding that changes in firm-specific factors, such as implied volatility and stock prices, are significant for CDS prices, and not so relevant to variations in credit spreads.

Another application of the CDS market was presented by Hull et al. [22], where apart from corroborating the theoretical relationship between bond yield and credit default swaps spreads, explored the relationship between the CDS market and credit rating announcements, showing that CDS spread changes have a predictive ability essentially for guess downgrades by Moody's, however the results for positive rating events were much less significant.

One of the most cited studies was done by Longstaff et al. [23], in which they used CDS premia as a tool to measure the size of default and non-default components in corporate spreads. They document the differences between CDS spreads and corporate bond yield spreads and found that the differences between those spreads highlight the relative importance of default risk and liquidity for corporate bonds. Their results indicate that the default component represents the majority of corporate spreads.

Fung et al. [24] studied the relationship between the stock market and high yield and investment grades and the CDS markets in the United States and found that the relationship between them depends on the credit quality of the underlying reference entity. They also figured that CDS spreads contain valuable information that can benefit stock market investors, presenting the example of the 2007 crisis where the increasingly high CDS spreads were ignored by stock market, and at that time that rise on the spreads could be indicating a deterioration of the credit worthiness of several companies. They also referred that the CDS market was leading either the stock and bond markets at the time.

Tang et al. [6] conducted a research in which they found that, while fundamental factors such as stock volatility and leverage are important drivers of CDS spread changes, corroborating the work of Ericsson et al. [25], the CDS spread movements are also affected by supply and demand imbalance and market liquidity. They also concluded that CDS spreads move together much more during credit crisis than in normal times, and that those movements are mostly explained by the firm fundamentals during the credit crisis.

Avino et al. [26] took a step forward in the CDS market research and performed a study in whereas the CDS market is predictable using the referred economic factors that influence its spreads. They employ both linear and non-linear forecasting models. For the linear model they used a structural and an autoregressive model and for the non-linear a Markov switching model, both structural and autoregressive as well. Their models were evaluated by using statistical metrics (RMSE, MAE and MCP) and by their economic performance, tested through trading strategies based on iTraxx CDS spreads, from where they download their datasets. Their results show evidence on predictability of iTraxx index spreads.

An extensive research of what exists in the literature has been done by Augustin et al. [7]. In their work, a large variety of research domains is covered, from CDS pricing to its relationship with other markets. To a more complete review on the existing literature the lecture of this work is suggested.

2.3.2 Machine Learning and Financial Markets

A lot of research has been done combining the use of technical analysis and fundamental analysis with different machine learning algorithms. Due to the lack of existent solutions applying those techniques to the CDS market, this Section will give an insight on the existent literature applied to other financial markets.

One of the first applications of Neural Networks (NN) to the field of financial economics was done by Kaastra et al.[27] where they presented an eight-step procedure to design a NN forecasting model for financial time series. Some years later, Thawornwong et al. [28] evaluated the predictive power of technical indicators by employing a Neural Network as a decision maker, helping investors to perform stock trading when the information received by the different technical indicators contradict each other. They used five technical indicators and concluded that using them as input variables of a NN, a more accurate stock trend prediction is achieved.

In 2003, Kim [29] applied Support Vector Machines (SVM) to predict the direction of change in the daily Korea Composite Stock Price Index (KOSPI), using also technical indicators as input variables. The developed model was then compared with other computing techniques, namely BP and CBR. The author concluded that SVM achieved the best hit rate (57%), improving the prediction performance relatively to CBR by almost 6% and by 3% when compared to BP.

Similar conclusions were achieved by Cao et al. [30] for five futures contracts gathered from the Chicago Mercantile Market, showing that SVM forecasts significantly better than the BP network, except for one case. They also performed an investigation on the SVM parameters and concluded that they play an important role on its performance, proposing a model with adaptive parameters, ASVM, that outperformed the standard SVM.

Hassan et al. [31] proposed the forecast of stock price of four airline companies through the application to the Hidden Markov Model (HMM) technique, using the open, close, high and low prices and taking as target the next day closing price of the stock. The obtained results of this model are comparable to the Artificial Neural Networks model, with the advantage of this model being explainable, unlike to what happens with ANN.

Another application for Artificial Intelligence (AI) methods was proposed by Zan Huang et al. [32], where they applied SVM together with backpropagation neural networks to the problem of credit rating prediction. They prepared two bond-rating data sets from the United States and Taiwan markets, developing two models for each market, a simplified one with the five most commonly used fundamental financial variables, and a complex one, with all the financial ratios available. For the four tested models, results show that SVM achieved comparable accuracy to the ones of NN, and both outperformed the Logistic Regression Model (LogR), used as benchmark.

Cavalcante et al. [33] recently performed a complete study on what has been done in the literature, presenting multiple studies on diverse markets using several machine learning techniques.

Literature presents multiple algorithms that can be applied to financial markets in order to predict future asset prices or returns or predict price movements directions. They all have strengths and weaknesses, for example NN might present high error rates since the amount of noise in the financial markets is usually high and, on the other hand, the SVM high levels of prediction accuracy might not lead to high profits [16].

Genetic Algorithms are also widely used in the literature, and once they are the ones to be applied in this thesis, a more complete literature analysis will be done in the next sub-Section.

2.3.3 Works on Genetic Algorithms

After their presentation in 1989 [15], Genetic Algorithms have been widely used in several different applications for parameter optimization problems. A lot of works in financial markets were developed, proposing also some improvements to the original model. This Section will present some of the most relevant works on this field.

Potvin et al. [34] proposed a Genetic Programming model to automatically generate short-term trading rules on 14 Canadian companies listed on Toronto stock exchange, based on their historical prices and transaction volume data. The generated trading rules are adjusted individually to each stock. Their results show that the use of trading rules is only better than the buy and hold strategy when the market is falling or stable. However, 9 of the 14 companies show a positive return in the testing period.

Taking into account the difficulty of choosing the appropriate parameter values for technical indicators, Fernández-Blanco et al. [12] applied an EA to determine the best number of days for the MACD indicator. Their results have proven that a strategy based on a MACD parameters optimization using a GA can outperform both the traditional MACD (using 12, 26, 9 as parameters) and buy and hold strategy. The authors referred that these encouraging results should be explored in future works, thus this idea will be adopted in this thesis.

Oh et al. [35] proposed a portfolio optimization scheme for index fund management using GA, aiming to prove that the index fund can increase its performance using GA and matching the benchmark indexes. The GA scheme is applied to the 200 companies of the Korea stock price index (KOSPI) 200 from January 1999 to December 2001 and the results show that their GA approach has significant advantages over the conventional portfolio mechanism. Furthermore, they concluded that GA becomes quite effective when market volatility has increased, but it has a mediocre performance when the market is flat.

Gorgulho et al. [36] developed a system also based in GA, with the objective of manage a financial portfolio using 7 technical indicators (EMA, HMA, ROC, RSI, MACD, TSI, OBV), with data from DJI ranging from 2003 until 2009, meaning that it included the 2008 financial crash. Their strategy is based on scores attribution to the different technical indicators according to some defined rules. Their system clearly beats the buy and hold and random investment strategies, even avoiding losing money during

the financial crash. Despite they did not calculate the best parameters to the technical indicators, their idea of score attribution to each technical indicator is promising and will be adopted in this thesis.

It is also commonly found in the literature the application of GA together with other AI techniques. In 2000, Kim et al. [37] proposed the use of GA to perform feature discretization and to determine the connection weights of an ANN, aiming to predict the price direction of stock index KOSPI. In this study, GA is applied not only to improve the learning algorithm but also to reduce the complexity in the feature space, eliminating irrelevant factors by mapping continuous variables into discrete sets. They concluded that their algorithm produced valid results, outperforming two other models used for comparison.

Kim et al [38] investigated the effectiveness of a hybrid time series prediction relying on GA, Adaptive Time Delay Neural Networks (ATNN) and Time Delaying Neural Networks (TDNN) on the search for temporal patterns. The GA was used for optimization of the number of time delays and network architectural factors, to improve the effectiveness of the ATNN and TDNN models. They concluded that this optimization improved the performance of those models, resulting in a higher accuracy when compared to the same systems without the use of GA.

Choudhry et al. [39] tested GA together with SVM (GA-SVM) for stock market direction price prediction. They used a set of 35 technical indicators as input features, where the GA is in charge of selecting the most informative ones. Their system achieved a higher performance in terms of hit ratio when compared to the stand alone SVM system, meaning that the GA-SVM system correctly predicted the direction of price a higher percentage of times. Yongchen Li and Honge Xu [40] applied the same two techniques together expecting to improve credit rating prediction accuracy. The GA algorithm was used in order to optimize SVM parameters. They ran analysis of variance (ANOVA) on the 21 financial ratios available on the Chinese data set, to test whether the financial variable was considered informative or not, regarding the bond-rating decision. They constructed two different models, one with the 6 most relevant financial rations an another with all the 21. The results showed that GA-SVM results were comparable to the ones of SVM, respectively 81.64% and 79.08% for the model constructed with only the most relevant variables, which achieved a better performance than the most complex one. This latter conclusion is relevant, meaning that the use of too much information can lead the system into erroneous predictions.

Besides those hybrid systems, the original GA has been subject to several efforts to improve its performance, increasing diversity so that the GA can continue to look for the optimum solution. Particularly two people stood out, John J. Grefenstette in 1992 with the introduction of the concept of Random Immigrants [41], where some individuals, usually the worst evaluated on the current generation of the algorithm, are replaced by randomly generated ones and, based on the previous work of Hellen Cobb on triggered hypermutation [42], one year later the two together introduced a mechanism called Hypermutation [41]. The Hypermutation increases the mutation rate whenever it is triggered, what happens when there is a degradation in the performance of the time-averaged best performance, and on the other side Random Immigrants replace part of the population, usually the least fit in each generation, with randomly generated values. Those mechanisms were tested in [43] and proven to have advantages over the standard GA. In this thesis the implementation of these mechanisms will be considered during the implementation of the system.

2.4 Chapter Conclusions

This chapter presented all the fundamental concepts required for the understanding of this thesis. It was made an analysis on the different investment strategies, both technical and fundamental analysis and it was justified the choice of the first one.

It was also made an explanation on the main concepts of Genetic Algorithms, including its different stages. This algorithm will be implemented aiming to find the best parameters to perform technical analysis.

Finally, it was presented the state of the art on the main topics covered by this thesis. Although, to the author knowledge, there have been no studies done on trading CDS using technical analysis and AI techniques, it was presented some works done on CDS, showing its relationship with other markets. Besides, to show the important role that machine learning algorithms play to investors, some works have been presented using different algorithms and different markets. Since the algorithm to be applied in this thesis is a GA, a sub Section was dedicated to show previous works containing some ideas that will be adopted in the implementation of this thesis system.

In Table 2-1 are presented the most relevant aspects of some of the most important papers considered for this Thesis.

Table 2-1: Resume of the most relevant State of the Art papers.

Work	Year of Publication	Methodologies Used	Performance Measures	Financial Market	Dataset Period	Average Return
[26]	2014	Structural and Markov Switching Models	Number of Trades, Return, Sharp Ratio RMSE, MAE, MCP	European Corporate CDS	09/2005 – 09/2010	14.09% (for the best model & period)
[28]	2003	NN	Number of Trades, Return, Risk	Stocks	08/1996 – 12/1999	0.15% (best model)
[29]	2003	SVM	Hit Ratio (prediction performance)	KOSPI	01/1989 – 12/1998	-
[30]	2004	SVM	NMSE, MAE, DS	Futures	01/1988 – 06/1999	-
[32]	2003	LogR, SVM, NN	Prediction Accuracy	Bonds	1991 – 2000	-
[31]	2005	HMM	MAPE	Stocks	12/2002 – 09/2004	-
[34]	2004	GP	Return	Stocks	06/1992 – 06/2000	-4.85% for Short -3.59% for Long
[36]	2011	GA	Number of Positions, Avg. profit per position Return, Sharp Ratio, Sortino Ratio	Stocks	01/2003 – 06/2009	62.95%
[39]	2008	GA-SVM	Hit Ratio	Stocks	08/2002 – 01/2008	-

Chapter 3

Proposed Architecture

This chapter aims to describe the implemented system architecture developed to approach the problematic of investment in the CDS market. It will give a general overview of the main stages of the system, from the data pre-processing to the investment simulator module. It is also made a brief explanation on the used technical indicators.

3.1 Overview

This chapter aims to describe and analyse the main modules of the proposed architecture for the system. Based on the ideas of previous papers [12], [36], the system will implement an investment strategy through the use of technical analysis, where the indicator's parameters will be optimized through a Genetic Algorithm.

The system architecture is composed of four main stages, each one designed to deal with a specific part of the program. The GA is the main core of this thesis, and for that reason a more thorough explanation of the options made in each stage of it is required. Besides, an explanation about the technical indicators that will be used is also required, in addition to the previously explanation about Technical Analysis in Chapter 2. The behaviour of other two modules is also described.

Due to its countless libraries that allow to implement complex functions easily, the whole system was implemented in Python. Furthermore, the implementation of the Genetic Algorithm was based on the DEAP [44] evolutionary computation framework.

3.2 General Description

In order to develop a solution to the investment in the CDS market, several steps should be defined. Based on several existent solutions, the path taken was to divide the problem in smaller parts, which correspond to different units of implementation, each one with a specific role within the system. As a result, there were defined four main layers for the system, namely the Data Module, the Technical Rules Module, the Optimization Module and the Investment Simulator Module, as represented in the Figure 3-1.

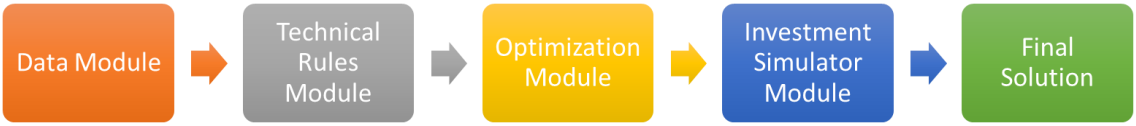

Figure 3-1: Stages of the system.

A brief introduction to the four main modules which compose the system is done below.

- Data Module: It is the module responsible to the pre-process of the downloaded data, making it ready to be used by the Technical Rules Module. This pre-processing consists on removing irrelevant information to the system, compacting the files that will be used, on selecting the tickers with enough data to be analysed, on organizing the data according to the maturity of the

CDS and finally on the splitting the datasets in two, in order to create a dataset for training and another of unseen data to test the viability of the system.

- Technical Rules Module: This module takes on the data from the previous years, provided by the Data Module, and calculates the technical indicators which will be used to define the investment strategy. It creates a new file for each company with all the technical indicators, calculated for all the days in analysis.

- Optimization Module: Taking on the data retrieved by the Technical Rules Module, this module is responsible to find the best strategy using a Genetic Algorithm, namely the best indicators to use and its respective parameters, according to the global profit made.

- Investment Simulator Module: It is first used to simulate all the strategies proposed by the GA, working as its evaluation function, since the best individual will be obtained according to the highest profit. Then, when the best strategy is found, it is tested in a specific period of unseen data, which will evaluate whether that strategy could be profitable or not.

This modular approach was also adopted in order to make potential changes to the application easier, making it possible to change a module without affecting the other modules, or in the worst-case scenario, make just small alterations. Figure 3-2 summarizes the system main phases, showing which module is responsible for each phase.

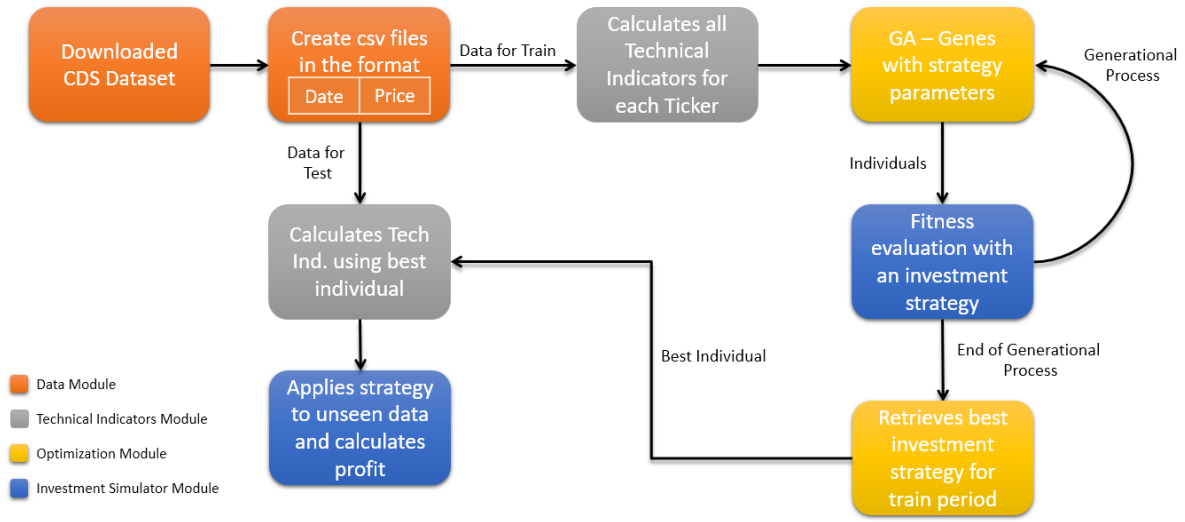

Figure 3-2: Main phases of the system.

3.3 Data Module

The CDS spreads data was obtained from two sources, CMA Datavision and Thomson Reuters CDS, with the first going further back than the latter but not being updated since 2008, and for that reason a combination of both sources was used [45]. The data set consists of historical CDS spreads of 236 companies, with maturities of 6 months, 1 year, 2, 5 and 10 years, during the period from January 1, 2005 to December 31, 2016, however many of the companies do not have values from that time. This thesis will focus on the CDS with a maturity of five years once they are by far the most liquid in the credit derivative market and the most used in the literature, according to Bai and Collin-Dufresne [46].

The downloaded *csv* file has three columns. The first one represents the Thomson Reuters CDS code which consists of a ticker, a number indicating the maturity, a symbol indicating the currency ("E" for euro and "S" for USD for example), and a variant code that indicates the type of restructuring constitutes a credit event, which have no relevance to this work. In the second column it is represented the date in the format "yyyy-MM-dd hh:mm:ss", yet since the spreads are updated daily one can simplify to the format "yyyy-MM-dd". Finally, in the third column is the value of the CDS spread.

The downloaded *csv* is then submitted to a pre-processing procedure, where the different tickers will be grouped in a different file according to its maturity. During this process, tickers who do not have values or which spread did not change for a long period of time are discarded. Then for each ticker in this file, the period of train and test will be defined, according to the available period data of each ticker, what will generate two more files. The most common periods found in the literature to make this split are 80% for train and 20% for test.

Once the data is divided, the train sample is passed to the Technical Rules Module to calculate the value of the different technical indicators, with all the used periods. This way, once these values are calculated only once, one will avoid that the GA constantly calculates them each time it runs, which speeds up the process. The test sample is also passed to the Technical Rules Module, but only when the Optimization Module finishes the search process, this way for that sample of data only the used technical indicators are calculated.

3.4 Technical Rules Module

As referred in Chapter 2.2.2, in order to make substantiated investments, investors usually rely on Fundamental Analysis and Technical Analysis, but only the latter will be used in this thesis.

Technical analysis has a great variety of technical indicators, meaning that a careful selection of those indicators is crucial to assure the correct behaviour of the system, as well as the weights assigned to each one. It is also convenient to consider the type of market in which one is investing, in the case of this thesis the market in study demands a special attention due to its high fluctuations and for that reason some indicators might not be as informative as wished. Besides, due to the inexistence of data about

the volume of transactions, indicators that require volume to make predictions, such On Balance Volume (OBV), will be discarded.

For the development of this thesis were chosen some of the most commonly used technical indicators in the literature, two of them are trend indicators namely an EMA and a Double Crossover method, and two momentum indicators, the Relative Strength Index (RSI) and the Rate of Change (ROC). Moreover, in order to study the variations in the price over the time it was used a simple mathematical concept called Percent Changes, that although it is not a technical indicator *per se*, can be used to measure if a price has suffered a big change lately, indicating a potential entry or exit signal.

The chosen technical indicators require the specification of some parameters, namely the period in which they are calculated. In the literature, there are easily found typical values proposed by specialists for the used indicators, however it might be an interesting approach to assign this task to the GA in an attempt to obtain better results, once this strategy has proven to be effective in previous works [12].

Have chosen the set of technical indicators, it is necessary to attribute a weight to each one since there is no proof that a certain technical indicator is better than the other. To work around this problem this thesis will follow the approach made by Gorgulho et al. [36], which created a score system to evaluate the importance of each indicator on each day, taking into consideration their recent values and the corresponding variation. The score system is identical to the one used by that author, meaning that based on the recent values of the indicator, a new value is obtained to show how strongly that indicator advises to buy or sell the CDS, with the following four grades:

- ✓ Very High Score – Indicates a strong buy signal and gives a score of 1 point to the indicator;
- ✓ High Score – Indicates a buy signal and gives a score of 0.5 points to the indicator;
- ✓ Low Score – Indicates a sell signal and gives a score of -0.5 points to the indicator;
- ✓ Very Low Score – Indicates a strong sell signal and gives a score of -1 point to the indicator;

Once the scores are attributed, they are used to define the final position of the system in the market, which takes the value of the sum of each indicator score multiplied by a weight attributed to that indicator by the genetic algorithm. This weight represents the relevance of each indicator, meaning that all the weights sum 1. Mathematically speaking one has:

$$Position = \sum_{i=0}^{N} W_i \cdot Score(X, i) \tag{3.1}$$

$$-1 \leq Position \leq 1 \tag{3.2}$$

$$0 \leq W_i \leq 1 \tag{3.3}$$

Where W_i represents the weight attributed by the GA to the indicator i and $Score(X, i)$ represents the grade attributed to the CDS by the indicator i at the day X.

The rest of this Section will be used to explain in detail the used technical indicators.

3.4.1 Exponential Moving Average

A Moving Average is a widely used technical indicator that helps smoothing the price by filtering noise from random price fluctuations, however, as a trend indicator, it can tell that a trend has begun only after that happening. By averaging the price data, it produces a smooth line that can be easily perceived, in opposition to the price line that contains more irregularities [11].

There are different types of moving averages, with Simple Moving Average (SMA) and Exponential Moving Average (EMA) being the two most widely used. The major difference between those two lies in the fact that the latter assigns a greater weight to the more recent data, giving it the ability to react faster to recent price variations.

The rules for trading are simply defined by the crossing points of the EMA with the price line. When the price is trading above the EMA one should adopt a buy position and, in the opposite side, when the price crosses bellow the EMA line is a good time to sell.

The EMA can be calculated through the following expression:

$$EMA_t(n) = price_t \times \left(\frac{2}{n+1}\right) + EMA_{t-1}(n-1) \times \left(1 - \frac{2}{n+1}\right) \tag{3.4}$$

Where n represents the selected time period for the EMA and t is the present day. Based on this indicator, one is able to define the rules for the score attribution present in the Table 3-1.

Table 3-1: Score attribution rules for the EMA indicator.

	$\boldsymbol{EMA(n)}$
Very High Score	Price surpasses the value of the EMA
High Score	EMA is rising
Low Score	EMA is decreasing
Very Low Score	Price crosses down the value of EMA

For the n parameter, the limit values defined are represented in the Table 3-2.

Table 3-2: Limit values for the EMA indicator.

Parameter	Lower bound	Upper bound
Days of EMA	2	50

For a better comprehension, it is represented in the Figure 3-3 an EMA of 12 days, with an indication of the score values at different moments.

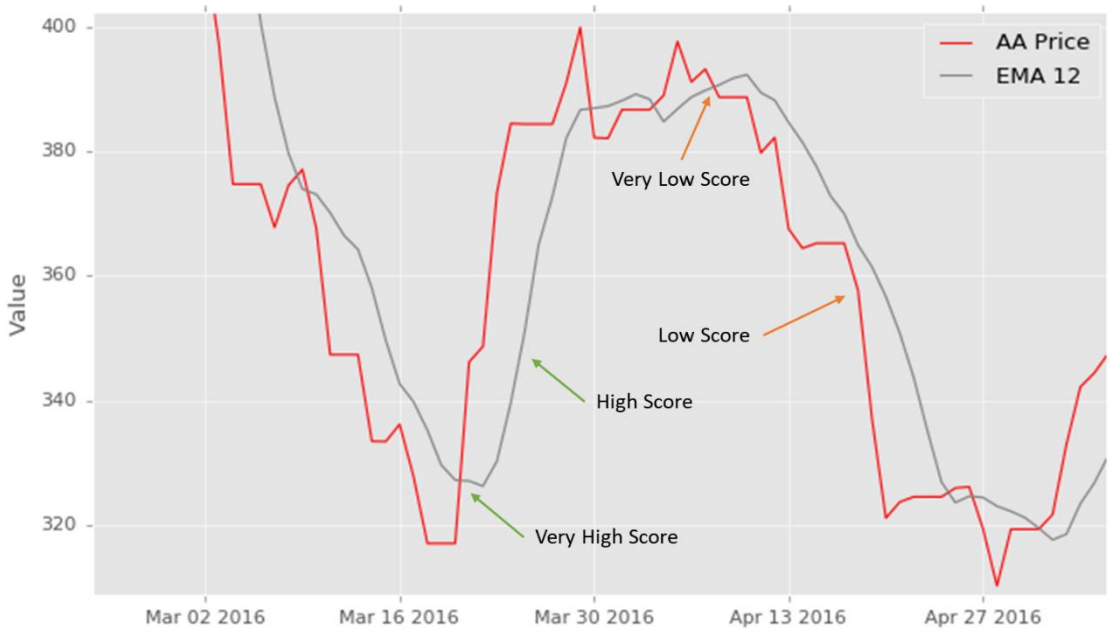

Figure 3-3: Example of score attribution by the 12 days EMA.

3.4.2 Double Crossover Method

This technical indicator uses two exponential moving averages to generate the market signals, one with a slower trend line using a longer time period identifies the principal trend, and another with a shorter period, which is more sensible to market variations and therefore reacts faster. Comparatively to the previous explained EMA, this technique lags the market a bit more, but produces fewer whipsaws, which occur when a price heads in one direction but it is quickly followed by a movement in the opposite direction.

The rules for trading using the Double Crossover Method are similar to the ones defined for the EMA, a buy signal is obtained when the price of the faster moving average crosses the slower moving average going up, meaning that a sell signal is obtained when the faster moving average is decreasing and crosses the slower moving average. Taking this into consideration, one can define the rules presented in the Table 3-3 for the Double Crossover Method.

Table 3-3: Score attribution rules for the Double Crossover indicator.

	$\boldsymbol{EMA(n) - EMA(N)}$
Very High Score	Faster $EMA(n)$ crosses above the slower $EMA(N)$ line
High Score	Both EMAs are rising
Low Score	Both EMAs are decreasing
Very Low Score	Faster $EMA(n)$ crosses below the slower $EMA(N)$ line

For the two parameters to optimize, n and N, the limit values defined were the ones represented in the Table 3-4.

Table 3-4: Limit values for the Double Crossover indicator.

Parameter	Lower bound	Upper bound
Days of Fast EMA	2	50
Days of Slow EMA	50	200

As stated before, the time spans of the two moving averages will be determined through the optimization process. An example of the Double Crossover Method is presented in the Figure 3-4, with a Fast EMA of 16 days and a Slow EMA of 55 days.

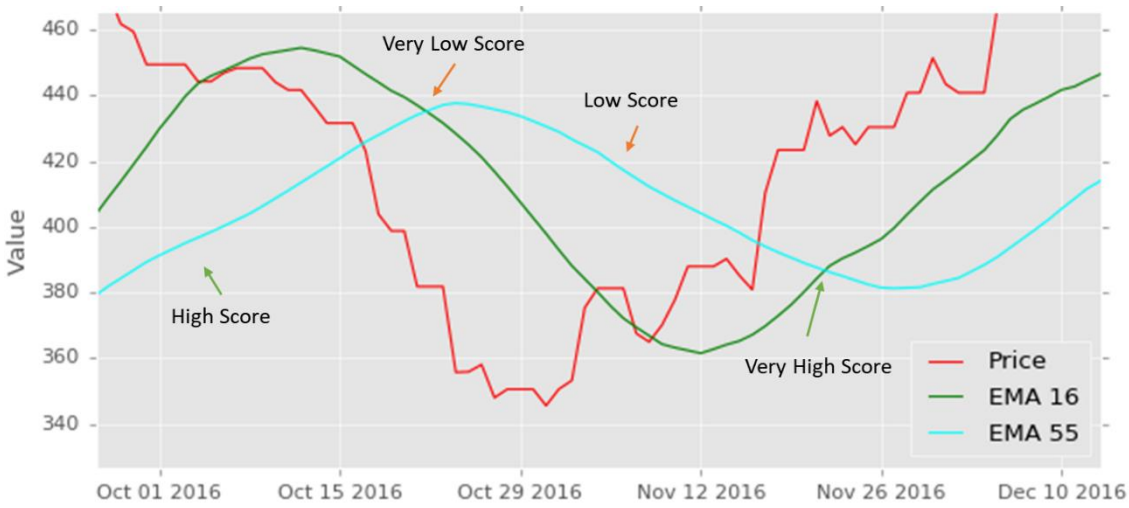

Figure 3-4: Example of score attribution by the Double Crossover method.

3.4.3 Relative Strength Index

The Relative Strength Index or RSI is an extremely popular momentum oscillator indicator that measures the speed and change of price movements. RSI is plotted on a vertical scale of 0 to 100 and through the analysis of its variation over time it allows to identify situations in which the market is overbought or oversold. Movements above 70 are considered overbought while movements below 30 are considered oversold. As a result, when the RSI line crosses 70 it could lead to a downward movement of the price once it has been rising for a long time without much pullback, which could mean that it is a good time to sell. On the other hand, when the RSI line crosses below 30 the prices have been falling for a long time, which can represent a good opportunity to buy at a lower price. It is important to note that one should only buy when the RSI crosses 30 from down to up and sell when it crosses 70 from up to down because one do not know how far the price might continue to follow the actual trend.

The RSI indicator for a period n can be obtained through the formula:

$$RSI(n) = 100 - \frac{100}{1 + RS(n)} \tag{3.5}$$

With RS given by the ratio between the average gains and losses in the considered period:

$$RS(n) = \frac{Average\ Gains(n)}{Average\ Losses(n)} \tag{3.6}$$

Taking into account the regions in which the RSI line stands and the direction of its movement, one can define the rules presented in the Table 3-5 in order to attribute the scores for this indicator.

Table 3-5: Score attribution rules for the RSI indicator.

	$\boldsymbol{RSI(n)}$
Very High Score	RSI line crosses above 30
High Score	RSI line is rising inside the threshold values
Low Score	RSI line is decreasing inside the threshold values
Very Low Score	RSI line crosses below 70

The period n of the RSI indicator was defined to be between the values presented in the Table 3-6.

Table 3-6: Limit values for the RSI indicator.

Parameter	Lower bound	Upper bound
Days of RSI	5	50

In the Figure 3-5 one can see what it is like to apply the rules of the Table 3-6 into a specific period of time.

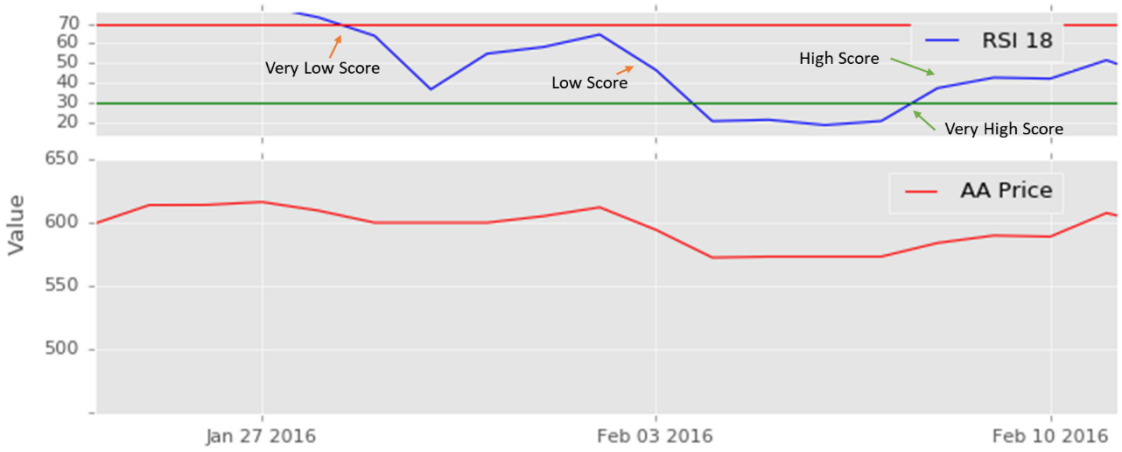

Figure 3-5: Example of score attribution by the RSI.

3.4.4 Rate of Change

The other momentum oscillator used in this thesis is the Rate of Change, or ROC. ROC measures the speed at which prices are changing, i.e., the difference between the current price and the price from n periods ago. It is calculated by the formula:

$$ROC(n) = \frac{Price(t) - Price(t-n)}{Price(t-n)} \times 100 \tag{3.7}$$

Where $Price(t)$ corresponds to the current day price and $Price(t-n)$ the price of n periods ago. In this case, the zero line becomes the midpoint line, if the current price is higher than the price n days ago, the resulting ROC would be above zero and, in the opposite side, if the current price is lower than the price of n days ago the ratio would be below zero [11]. In the first situation, the prices are rising which is an opportunity to buy whereas in the latter situation the prices are falling and the opposite attitude should be adopted. Based on this assumption, it is possible to define the rules of the Table 3-7 for the ROC indicator.

Table 3-7: Score attribution rules for the ROC indicator.

	$\boldsymbol{ROC(n)}$
Very High Score	ROC line crosses above 0
High Score	ROC line is decreasing, price is rising – bearish divergence
Low Score	ROC line is rising, price decreasing - bullish signal
Very Low Score	ROC line crosses below 0

Similarly to the period set for the RSI indicator, the same boundaries were determined for parameter n of the Rate of Change, as stated in Table 3-8.

Table 3-8: Limit values for the ROC indicator.

Parameter	Lower bound	Upper bound
Days of Rate of Change	5	50

Figure 3-6 illustrates the information provided by the Rate of Change.

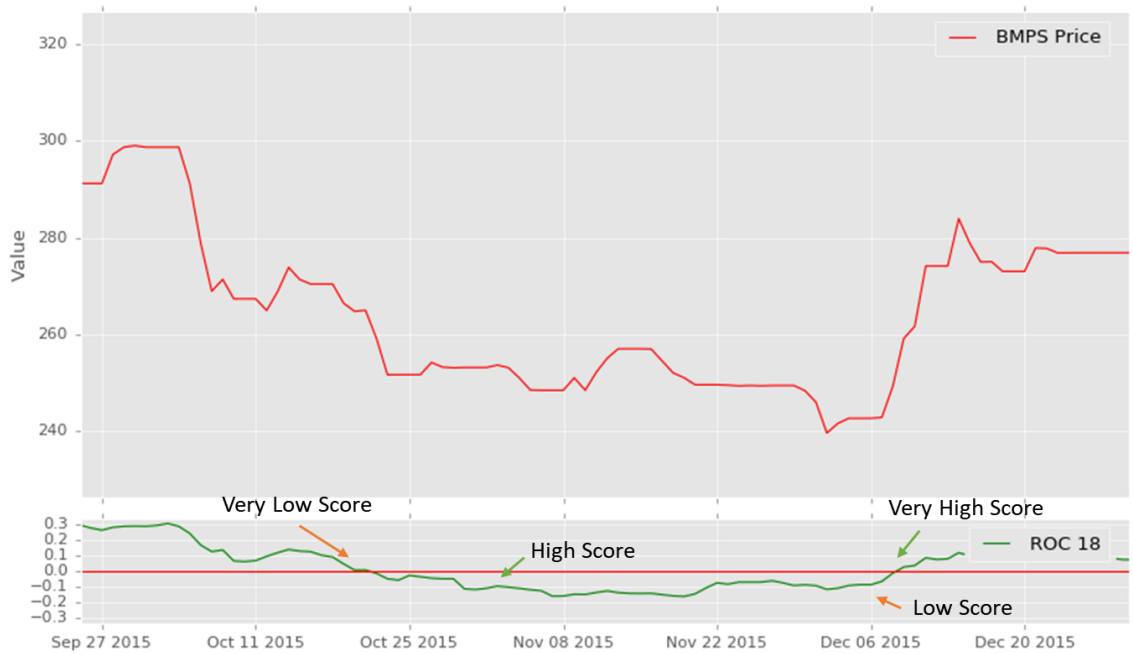

Figure 3-6: Score attribution by the ROC indicator.

3.4.5 Percent Changes

A Percent Change is a mathematical concept that represents the degree of change over time, calculated in the exact same way as the ROC indicator. However, with this indicator one will try to find some threshold values of variation from which one will decide rather to invest or not. If a company has raised its value more than a certain requirement in the last n days, it might be an interesting company to buy. On the other hand, if a company has seen his value fall more than another requirement in that period it might not be a good investment at the moment. Considering this, it is possible to define the score attribution rules and they are presented in the Table 3-9.

Table 3-9: Score attribution rules for the Percent Changes indicator.

	$\boldsymbol{PercentChange(n)}$
Very High Score	$PercentChange(n)$ is above a buy threshold
High Score	$PercentChange(n)$ is positive
Low Score	$PercentChange(n)$ is negative
Very Low Score	$PercentChange(n)$ is bellow a sell threshold

The two thresholds will be obtained through the GA module, as well as the periods to be considered for the calculation of the percent changes, meaning that 3 parameters need optimization for this indicator. Those periods belong to the interval represented in Table 3-10.

Table 3-10: Limit values for the Percent Changes indicator.

Parameter	Lower bound	Upper bound
Days of Percent Change	1	20
Buy Threshold	-1	1
Sell Threshold	-1	1

Figure 3-7 shows the application of a two-day Percent Change to the CDS spread of the Italian bank BMPS, with a buy requirement defined by the green line and the sell requirement by the red one.

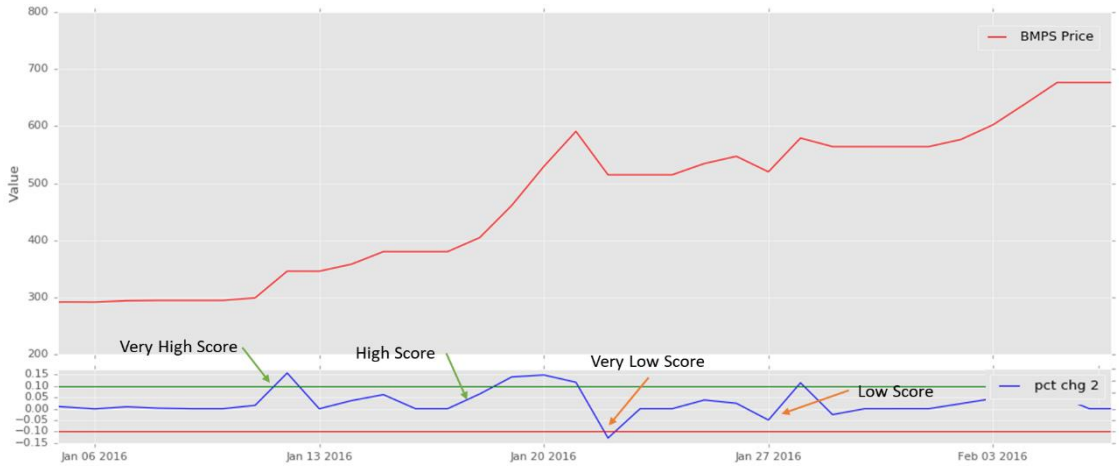

Figure 3-7: Score attribution by the Percent Changes indicator.

3.5 Optimization Module

As previously stated, the technical indicators used in this master thesis have multiple parameters that require optimization, for which is responsible a Genetic Algorithm. In Chapter 2.3 has been already done a previous approach to this subject, covering the most important aspects. This Section will be used to discuss the chosen parameters for the different stages of the GA, as well as the required structure of the individual to fulfil the required optimization.

During the search process, the GA generates individuals that could represent a possible solution to problem. Each produced individual determines the time spans of all the technical indicators and how they are going to be combined. This individual is then supplied to the Investment Simulator Module, which during the search process will work as the evaluation function of the individual, once it retrieves the profit that the GA intends to maximize. The technical indicators for each company have already been calculated in the Data Module, meaning that they can be easily accessed by the Investment Simulator, without having to be calculated each time the GA runs.

3.5.1 Chromosome Representation

As explained in the Chapter 2.2.3, the chromosome contains in its genes the information required by the Technical Rules Module to generate the set of technical indicators. This chromosome is composed by a vector of floats with values in the interval [0,1]. Notwithstanding, it was referred in the previous Section that some parameters of the technical indicators are integers, such as the periods of calculation of the exponential moving averages. Considering this, it will be necessary to map this values into their respective intervals. The reason why one needs to bound the values into the interval between [0,1] is to provide a universal mutation operation function, since if there were used genes with different boundaries it would be necessary to adapt the mutation function.

The chromosome used by the system contains 15 genes, as represented in the Table 3-11. It is composed by the eight parameters of the technical indicators referred in the previous sub-Sections, followed by the weights of each indicator, in order to let the system itself try the most appropriate value for it. Finally, two additional genes were added to decide, taking in consideration the weighted sum of the scores of each indicator, the threshold values from which the system adopts a buy or a sell position.

Table 3-11: Chromosome Representation.

Chromosome Representation

EMA	Double Crossover		RSI	ROC	Percent Changes		
Nº of days EMA	Nº of days Short EMA	Nº of days Long EMA	Nº of days RSI	Nº of days ROC	Nº of days	Buy req.	Sell Req.

Weights					Requirements	
Weight EMA	Weight Double Crossover	Weight RSI	Weight ROC	Weight Pct Changes	Requirement to Buy	Requirement to Sell

3.5.2 Selection

Once the GA's population is defined, the generational process begins with the selection of the individuals, which will choose the individuals to mate or mutate, hoping the individuals of the subsequent generation have a higher fitness. The fitness function of the GA relies on the Investment Simulator Module where, as it will be explained further, the generated individuals will be evaluated according the profit generated, in an attempt to achieve increasingly higher profits until the generational process is finished.

However, there is not a general guideline concerning the way of choosing the best selection method for each problem [47], some works have proven that the tournament selection outperformed the other strategies, achieving the best solution quality with low computing times, which is mainly due to the fact that tournament does not require sorting the entire population [18], [48].

In the tournament selection, a set of k individuals are randomly chosen from the population of size n and compete against each other. The individual with the highest fitness wins and will be included as one of the next generation population. The whole process is repeated n times for the entire population. Considering this, the choice of the tournament size is also important, once a larger value of the tournament size leads to a higher expected loss of diversity since less individuals are chosen to participate in the next steps. Therefore, the tournament size chosen was 3.

3.5.3 Crossover

The third step of the GA is the crossover of the selected individuals to form the next generation. As discussed in 2.2.3, there are three main ways to perform crossover, one-point, multi-point and uniform, each one with advantages and disadvantages. For example, uniform crossover might create offspring that will be very different from its parents if these are not similar but, if they are similar, the offspring will be similar as well. This last situation also happens when using one-point crossover, although it is very simple, it only has one swapping point, meaning that the offspring will be less diverse and quite similar

to its parents. Thus, the crossover technique used was two-point crossover, as it was designed to overcome some of the disadvantages of both uniform crossover and one-point crossover [49], as allows that the problem space to be searched more thoroughly.

Regarding the crossover rate, the chosen a value was the most found on the literature, 50%.

3.5.4 Mutation

Along with the crossover process, the Mutation is also responsible for the manipulation of the selected individuals to form the next generation. From the two presented methods in Section 2, the mutation method chosen was the Gaussian Mutation, as it stands as the most popular operator for real value floating point vectors [50]. It was used with a standard deviation σ and a mean μ taking the value of the gene to be mutated. This distribution is presented in the Figure 3-8.

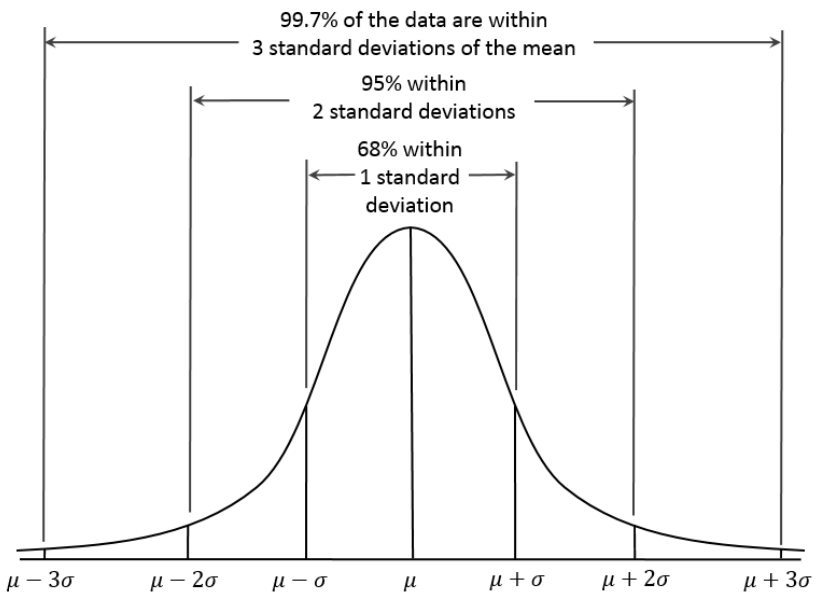

Figure 3-8: Gaussian Distribution. Source: [51]

According to this distribution, the height of the curve represents the likelihood of getting that particular value of the x axis, meaning that the mutated gene is more likely to be near the mean than far from it. To the mutation rate was defined the value 20%, in order to respect the optimal values found in the literature [52].

3.5.5 Improvements

Sometimes, crossover and mutation might generate weaker offspring when compared to the parents, and good candidates might end being discarded. Despite there is a chance that the GA would end up re-discovering those better individuals in a subsequent generation, there is no guarantee that it is going to happen, and it might entail an impact in the GA performance.

To solve this issue, in addition to the stages described and performed by the GA, another feature is added, called Elitism. The role of Elitism is to create an “Hall of Fame” of a chosen number N of best individuals that ever lived in the population during the evolution process. As any hall of fame, it is sorted from the fittest individual ever to the Nth fittest. At each generation, the hall of fame is updated with the population by replacing the worst individuals in it by the best individuals present in the population, but only if they are better obviously, in order to keep the hall of fame size constant. In this work, there were kept the top-10 performing individuals, since those are still eligible for selection as parents.

Another improvement attempt to the GA was made by using the approach proposed in [43], namely the implementation of Random Immigrants. This technique simply replaces a part of the population of each generation with randomly generated values. Those replaced individuals are the worst performing, and due to that fact, they would never be selected for crossover, the convergence of the algorithm will not be affected, if the percentage of the population that is being replaced, called the replacement rate, is properly chosen. In this thesis the replacement rate chosen was 5% of the entire population.

Finally, the last attempt of improvement made was the implementation of the already referred Hyper-Mutation. When the algorithm does not show a better result in the last three generations the hyper-mutation is triggered, meaning that the mutation rate is doubled and continues that way until the end of the algorithm.

3.6 Investment Simulator Module

In order to test the performance of each hypothetical model obtained with the GA, it was implemented an Investment Simulator Module to create the investment strategies, meaning place the buy and sell orders, according to the values retrieved by the algorithm to the technical indicators. As input, the module receives a ticker and an individual, from which the technical indicators will be calculated for that ticker.

The Investment Simulator Module is necessary both for the train and test periods. During the train period, this module calculates the profits for the different individuals, working as the fitness function of the Genetic Algorithm. During the period of test, it takes on the obtained best individual and for each ticker calculates the Return on Investment, which will be explained further. Furthermore, it is printed the detailed graphic of the ticker, with the CDS spread and the markers indicating the buy and sell orders.

The Investment Simulator has four states, Buy, Hold, Sell and Out. A Buy order is placed when the score of each indicator multiplied by its weight is greater or equal to the requirement to buy. The value of the purchase is stored for future calculation of the profit and this position will remain on Hold until the referred sum became lower than the value of the requirement to sell. In this case, the investment simulator adopts a Sell position and calculates the trade profit, making the difference between the buy and sell values. After the sell, the simulator stays off the market until the requirement to buy is reached again.

The simulator only calculates trade profits or losses, returning the final profit for each ticker when the period in analysis is complete. This means that for each ticker, the investment amount is equal to the CDS spread value were the buy requirement is triggered and the final profit is just the summation of all the trades profits and losses. Those calculations are performed for the selected tickers.

In an attempt to simulate what happens in the real life, to each trade made by the simulator is associated a fee that depends on the trade profit, meaning that when Sell is triggered one has to pay 1% of the trade profit to the broker.

3.7 Chapter Conclusions

The implemented system is composed of four main layers, the Data Module, the Technical Rules Module, the Optimization Module and the Investment Simulator Module, who have dependencies between each other.

The Data Module is responsible to make the necessary alterations to the downloaded data, making it ready to be used by the Technical Rules Module. This module retrieves a file for each company, with all the technical indicators calculated for the periods defined and that will be required by the Optimization Module. In its turn this module, making use of a Genetic Algorithm, will pick the best indicators to use and its respective time spans. The GA will find the best individual according to the profit calculated by the Investment Simulator Module, which during the search process functions as the GA's fitness function. When the best individual is found, the investment simulator returns graphical information about the trades as well.

The system will be evaluated in the Chapter 4.

Chapter 4

System Evaluation

In this chapter are presented the tests performed to the system described in Chapter 3. It is also done an evaluation of the system, comparing it with other investment strategies.

4.1 Overview

In this chapter, the proposed algorithm will be submitted to validation tests and its performance will be benchmarked against other strategies. Those tests will be performed using data from the CDS market, being necessary to split the entire dataset into train and test, the first is necessary to find the best strategy and the latter to evaluate that strategy. This evaluation will rely on some performance measures that will be presented in this Chapter.

4.2 Train and Test split

In order to evaluate the performance of the algorithm described in Chapter 3, it was applied a testing strategy widely used by investors and brokers called Backtesting [53]. This method consists in apply the developed strategy to relevant historical data and check what would have happen if the strategy was faithfully followed, which makes it possible to confirm the viability of the strategy through the analysis of the obtained levels of profitability in that period.

This way, to build the model and posteriorly confirm its validation, the Data Module is responsible to divide the entire dataset in two sets, one for train and another for testing. The train dataset is the one used by the GA to find the best strategy (individual), through the calculation of the profit made in that period. When this step ends, the strategy will be applied in the test data, unseen until that moment. Nevertheless, it is important to notice that obtaining a positive profit during the train period does not imply that the profit obtained during the test period will be positive as well.

In the case of this thesis, as referred in Chapter 3.3, the dataset available ranges from January 1, 2005 to December 31, 2016. It was defined that the first 80% of the whole dataset would be used for train and the last 20% to test.

4.3 Performance Measures

Generally, with machine learning, testing accuracy defines how well the algorithm works, but in investing is not only about accuracy but also about our performance, one might have an 80% accuracy but still have a non-profitable algorithm. This way, besides measuring accuracy in terms of profitable trades, several other performance measures were defined.

4.3.1 Return on Investment

One of the most important performance measures of an investment is the Return on Investment (ROI). This measure essentially represents the amount of return on an investment relative to the investment cost, being easily calculated using the equation:

$$ROI = \frac{Gain - Cost}{Cost} \tag{4.1}$$

Where $Gain$ corresponds to the return obtained after applying the investment strategy to the initial funds, $Cost$. Due to the fact that ROI is measured as a percentage, it can be used to compare the profitability of different investments and strategies, taking into consideration that it is calculated in the same period of time.

When considering multiple investments, one can also obtain the Average Return on Investment by simply calculate the mathematical average of all the ROI.

4.3.2 Classification Parameters

In addition to the rentability measure, it is a good idea to have a few metrics on hand to compare the developed strategy either with other strategies or some hypothetical model. However, it does not exist a classification parameter that works for every strategy, once they might have different risk tolerances and requisites to perform a trade.

Below is presented a list of the used classification parameters:

- **Number of Trades -** The number of trades done in a specific period. Measures how active in the market a system is, being necessary to obtain some of the other parameters. Besides, an excessive number of trades might be responsible for the reduction of trading returns, once transaction costs are taken into consideration.
- **Percentage of Profitable Trades -** Measures the number of trades that generated a positive return, after deduction all the trade commissions, divided by the total number of trades. Some traders do not pay much attention to this factor, in fact if they can handle streaks of losing trades, they prefer a more profitable system even if has a win rate of 40% or less.
- **Percentage of Non-Profitable Trades -** On the opposite side, this measures the number of trades that, after discounting the commissions, generated a loss.
- **Average Profit per Trade -** Indicates the average profit/loss of all the trades within a specific period, that is the average amount of money that was won or lost per trade. It is calculated by dividing the total net profit by the total number of trades. The higher this value is, the better.
- **Greatest profit -** Represents the most profitable trade done in the period of analysis, being calculated as a percentage. It can serve as an interesting term of comparison between

strategies once it can tell how likely a strategy is able to take advantage of a market opportunity to make higher profit.

- **Greatest loss -** On the other hand, and for identical reasons, this measure can also be an interesting term of comparison between strategies.

4.3.3 Used Strategies

Due to the lack of existent solutions for trading in the CDS market, it was developed a set of strategies to serve as a term of comparison to the developed one.

- **Random -** According to some studies [54], the best strategy for investment is to invest randomly. The basic idea behind this hypothesis is that no matter how much expert a trader is, its knowledge about the markets is always limited, which might lead to misleading strategies and poor investment decisions. The same authors concluded that a random investment strategy is much less volatile than others, which means that although this strategy might not be as profitable as others, it will not lose as much as either. Taking this into account, a random strategy was applied by creating a set of 150 purely random individuals and testing them in the same period of the other strategies.

- **Literature -** This strategy was implemented with the intent of compare a strategy based on some of the most typical values found for the technical indicators, [55]–[57], with the ones retrieved by the evolutionary strategy and confirm if the usage of computational techniques can improve the obtained results, mostly in terms of profitability, or not. This way, to apply the strategy, were used the parameters presented in the Table 4-1:

Table 4-1: Parameters of the Literature Strategy.

Chromosome Representation for the Literature Strategy

EMA	Double Crossover		RSI	ROC	Percent Changes		
12	21	55	14	12	5	0,2	0,2

Weights					Requirements	
20%	20%	20%	20%	20%	0,5	0,2

- **Standard GA -** The first evolutionary strategy defined is based on the standard Hollands [58] implementation of the Genetic Algorithm, using the same chromosome structure, crossover probability, selection tournament and probability of mutation of the proposed architecture, however without the improvements to the algorithm introduced in the Section 3.5.5.

- **Improved GA -** This strategy aims to improve the results of the previous strategy using the improvement features defined in the Section 3.5.5, namely Elitism, Random Immigrants and Hypermutation. Moreover, for a more legitimate comparison of these two strategies, the number of individuals and generations used during the search process are equal.

4.4 Case Studies

In this Section, the case studies and the main results obtained through the application of the previously described strategies are presented.

4.4.1 CDS Market

This case study consists on the application of the four defined strategies to the CDS Market, more specifically to 177 companies selected after the pre-processing done by the Data Module. Taking into account what was previously stated in the beginning of this chapter, the strategies will be applied from September 2015 until December 2016, as this was the resulting test dataset after the split process. The remaining dataset was used to train both the Genetic Algorithms. Those periods are explicit in Table 4-2.

Table 4-2: Train and test periods.

Parameter	Value
Train Period	12/2007 – 08/2015
Test Period	09/2015 – 12/2016

The parameters of the GAs were the represented in the Table 4-3.

Table 4-3: Parameters of Genetic Algorithms.

Standard GA	Improved GA
100 Individuals 50 generations 50% Probability of Crossover Tournament Selection of 3 individuals Mutation Probability of 20%	100 Individuals 50 generations 50% Probability of Crossover Tournament Selection of 3 individuals Mutation Probability of 20% Hypermutation is triggered after 3 generations without fitness improvements and doubles Mutation Probability (40%) Elitism of the best individual 5% of Random Immigrants

The performed tests led to the graph of the Figure 4-1, where each curve represents the ROI of a certain strategy. In this Figure, it is possible to observe that both the GA outperform significantly the remaining strategies and that the Random Strategy can achieve a better ROI than the one achieved by following the most used technical indicators. The results show that both Genetic Algorithms can make a better use of the opportunities to make a higher profit, as is it is possible to observe between January and March of 2016 where the ROI increased exponentially. They both can generate an increase of returns over 50% higher than the other strategies, with the Improved GA achieving almost 90% in only 16 months. It is important to take in account that once the creation of a portfolio was not a concern, those returns are calculated for all the 177 companies, which could mean that the results could improve even more if a careful selection of the companies was made.

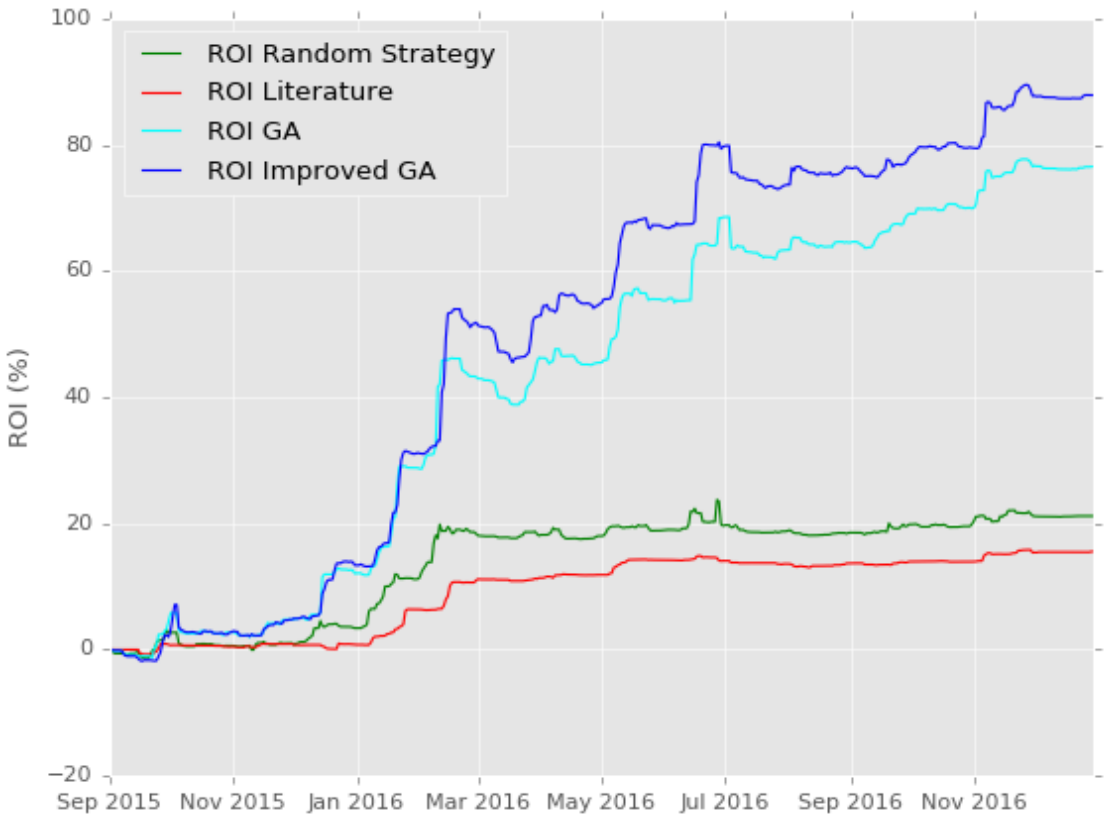

Figure 4-1: Evolution of the Average ROI for the defined strategies.

To support this hypothesis, Table 4-4 displays the 5 companies with the highest ROI during the test period, for the Improved GA strategy. It is a curious fact that of these 5 companies, 4 are oil exploration companies.

Table 4-4: Companies with the biggest Return on Investment.

Ticker	ROI (%)
CHK	947.61
COP	868.51
APC	744.73
DVN	471.63
AAL	223.19

On the opposite side, Table 4-5 shows the companies with the lowest ROI.

Table 4-5: Companies with the lowest Return on Investment.

Ticker	ROI (%)
ALCA	-67.39
ASFZ	-56.76
CNG	-31.58
BRG	-31.28
AUG	-23.42

Furthermore, a consequence of the non-creation of a portfolio is that there are times where the developed strategy has a vast number of open positions, as it is possible to see in Figure 4-2. In fact, there is a mean of around 77 open positions per day, meaning that this case study represents a hypothetical situation where one would not take into account the risk of having such amount of money invested.

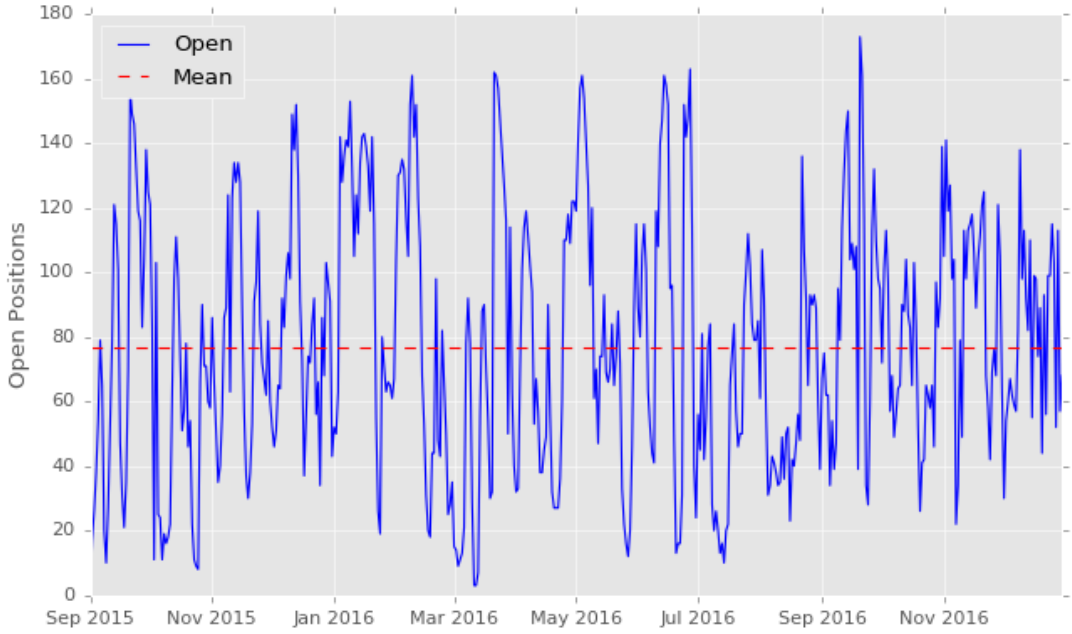

Figure 4-2: Number of open positions during the train period.

Table 4-6 summarizes the performance of the applied strategies, according to the previously defined parameters, for this Case Study.

Table 4-6: Classification Parameters for Case Study 1.

Parameter	Random	Literature	Standard GA	Improved GA
Number of Trades	17462	2350	10387	10366
Profitable Trades (%)	27.83	33.36	31.41	28.81
Non-Profitable Trades (%)	72.17	66.64	68.59	71.19
Average Profit per Trade (%)	0.19	0.95	0.87	0.98
Most Profitable Trade (%)	91.17	156.15	95.98	150.88
Biggest Loss (%)	-27.05	-27.05	-24.72	-24.72
Average ROI (%)	21.18	15.70	65.89	87.84

From the analysis of Table 4-6 it is possible to conclude that, despite not being the most accurate in terms of profitable trades, the Improved GA can achieve a slightly bigger average profit per trade when compared with the Literature and Standard GA strategy. The Literature strategy achieves the highest number of profitable trades and even the most profitable one, however as a consequence of being the one who performs the smallest amount of trades, it is the one who achieves a smaller Average ROI. On the other hand, the two evolutionary strategies achieve a very similar performance, although it is interesting to point that notwithstanding the fact that the Improved GA has a lower percentage of profitable trades, it still achieves a higher Average ROI than the Standard GA, what can be partly explained considering the fact that the first has a superior most profitable trade. The Standard GA also performs more trades, which means that the trade commissions paid are higher, causing the average profit per trade to drop.

For a better comparison between the two evolutionary strategies, it is also interesting to observe the obtained best individuals during the training period for this case study. To facilitate the reading, the best individual was mapped from the values used during the execution of the algorithms, meaning the interval [0,1], to the ones defined for the technical indicators intervals, presented in Chapter 3.4. Table 4-7 shows the parameters obtained for the used technical indicators.

Table 4-7: Case Study 1 - Indicators parameters for both evolutionary strategies.

Indicator	Parameters	Standard GA	Improved GA
EMA	Number of Days	2	4
Double Crossover	Days Short EMA	20	46
	Days Long EMA	116	116
RSI	Number of Days	48	48
ROC	Number of Days	4	28
Percent Changes	Number of days	2	4
	Requirement to Buy	0.322	0.503
	Requirement to Sell	- 0.180	- 0.260

It is perceivable that the Standard GA used shorter exponential moving averages, which might be one of the reasons for its higher number of trades. The number of days used for the RSI calculation is equal in both cases though the Improved GA uses a higher number of days to obtain the ROC. Table 4-8 shows, in percentage, the attributed weights to each indicator by the two algorithms.

Table 4-8: Case Study 1 - Weights of indicators for both evolutionary strategies, in percentage.

Indicators	Standard GA	Improved GA
EMA	28.25	28.73
Double Crossover	17.86	10.96
RSI	2.99	0.98
ROC	15.13	23.41
Percent Changes	35.78	35.92

In both cases the most relevant indicator is the Percent Changes, which confirms its suitability to very volatile markets, as it is the case of CDS. On the contrary, the RSI indicator is the least used. To conclude this case study, Table 4-9 presents the requirements to buy and sell a CDS defined by the Genetic Algorithms, based on calculated Position as explained in Chapter 3.4.

Table 4-9: Case Study 1 - Requirements to buy and sell for both evolutionary strategies.

	Standard GA	Improved GA
Requirement to Buy	0.089	0.182
Requirement to Sell	-0.023	- 0.062

The next example will take on the results of the Table 4-4 to show how profitable can be to invest in a sector that has been particularly affected by the financial crisis in the last years, using CDS.

4.4.2 CHK – Chesapeake Oil

In this example, the system will be evaluated in a common situation of the CDS market, which happen recently to a company of the Energy sector called Chesapeake Energy. In the early 2016, concerns over its worsening liquidity and its ability to manage nearly $10 billion of obligations [59], along with the fact that oil prices plunge to their lowest levels in a decade [60] had contributed to a crash in their stock values and on the opposite side an explosive rise on their CDS spreads. This “extreme fear” period eventually ended and value of their CDS spreads went back to normal. This period is represented on Figure 4-3.

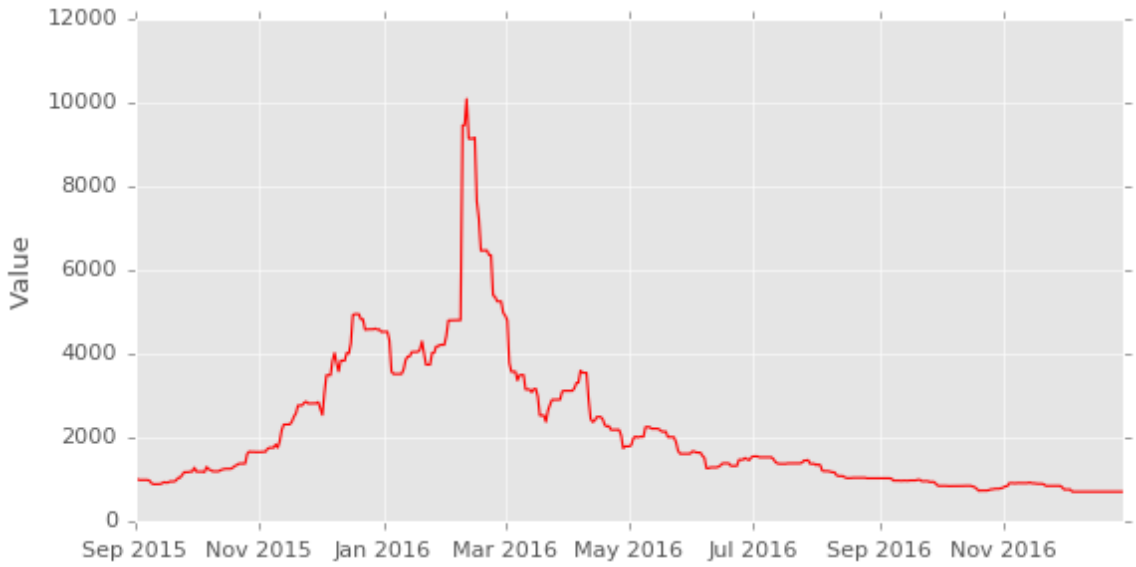

Figure 4-3: CDS spread of Chesapeake Energy during the period 09/2015 - 12/2016.

The periods for this test are the same as the ones used in the previous case study (Table 4-2), as well as the parameters for the GAs (Table 4-3). This way, the period of test coincided with the referred period.

The strategies presented in Chapter 4.3.3. were applied to this data set, and the obtained results are summarized in the Table 4-10. The evolution of the ROI is also presented in Figure 4-4.

Table 4-10: Classification Parameters for CHK example.

Parameter	Random	Literature	Standard GA	Improved GA
Number of Trades	78	6	52	41
Profitable Trades (%)	53.85	50	44.23	46.34
Non- Profitable Trades (%)	46.15	50	55.77	53.66
Average Profit per Trade (%)	2.21	13.89	5.55	7.64
Most Profitable Trade (%)	91.17	90.91	90.62	127.87
Biggest Loss (%)	-17.86	-5.43	-17.88	-5.94
ROI (%)	301.89	76.61	785.80	1194.23

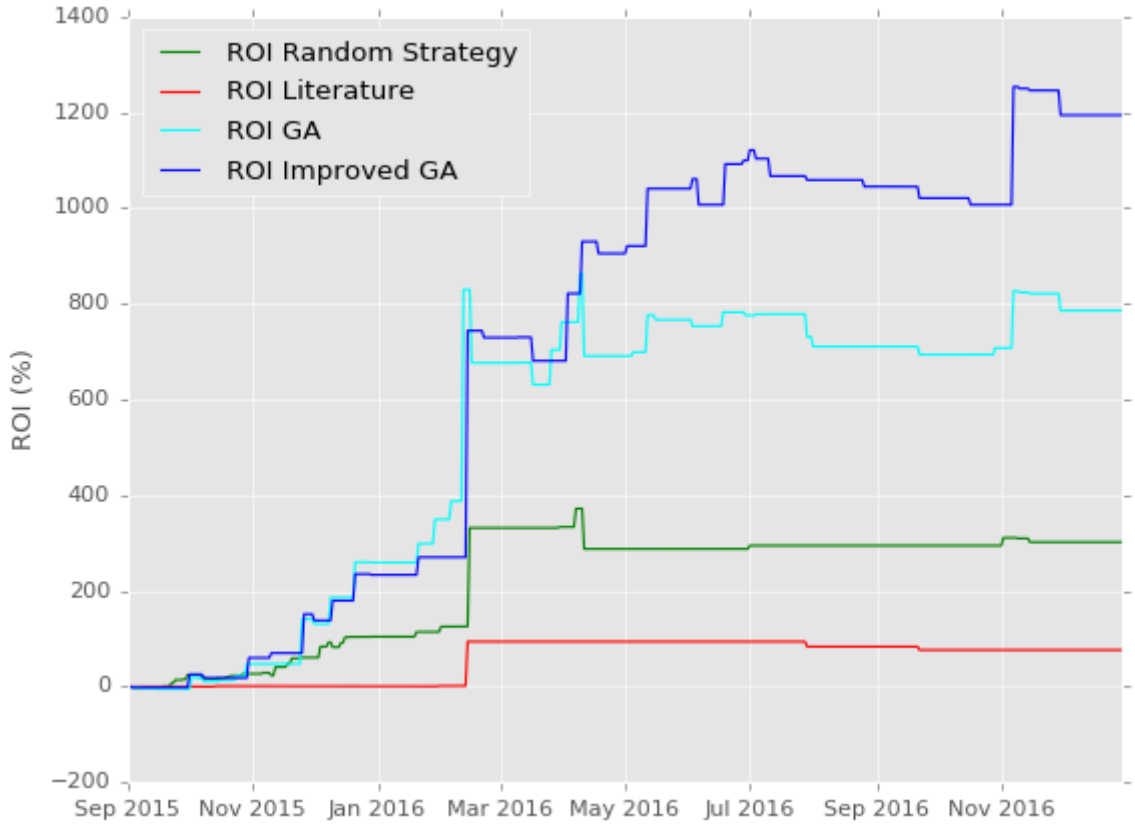

Figure 4-4: Evolution of the Return on the Investment for the CHK example.

From the analysis of the Table 4-10 and the Figure 4-4, it is clear that the Improved GA strategy clearly outperforms the other strategies, achieving the highest ROI of them all. Besides, it is the one that can take a greater advantage on the explosive growth verified in February 2016, seizing that opportunity to make the highest most profitable trade. The number of profitable trades is around 50% in all the strategies. It should be noted that the Literature Strategy performs a considerable less number of trades

than the others, which makes it the one with smallest ROI. Both the Random and the Standard GA strategy achieve a similar ROI, however despite the latter has a lower accuracy considering the number of profitable trades, it can achieve a bigger average profit per trade, once it also performs fewer trades.

In order to have a clear overview of when the trades are made, in the Figure 4-5, Figure 4-6 and Figure 4-7 are presented the buy and sell points defined by the Improved GA strategy. With the purpose of providing a better visualization of the trades, the period in analysis was divided in three graphs. The markers in green represent buy points whilst the red one represent sell points.

Figure 4-5 shows the first months of the test period. In this period is easily observed that the spread of CHK is rising and that the strategy is able to take advantage of this, with the majority of the preformed trades being profitable. During this period, two trades stood out, namely one concluded in October 30 profiting 35%, and the other concluded at November 26 with a total profit of near 48%.

Figure 4-5: Buy and sell points resulting of the application of the Improved GA strategy to CHK (Sep 2015 - Dec 2015).

Figure 4-6 contains the highest rise of the spread, where the strategy is able to perform its most profitable trade, with a return of almost 128%. However, in the three subsequent trades it does not have a chance to make a good profit due to the fall in the spread, it is able to continue to perform profitable trades when the spread is rising.

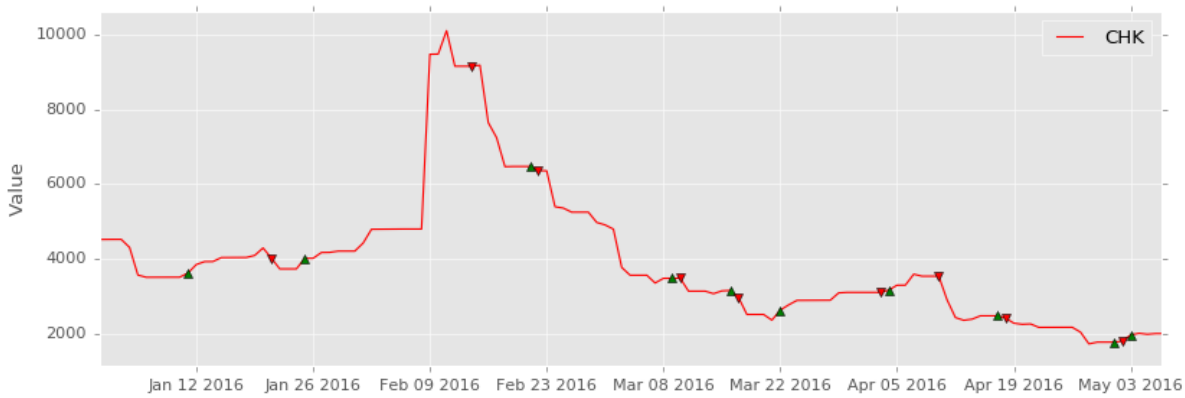

Figure 4-6: Buy and sell points resulting of the application of the Improved GA strategy to CHK (Jan 2016 - Apr 2016).

Figure 4-7 contains the time period where the strategy has more difficulties to generate profits, which is mainly due to the fact that the spread is constantly falling. During this period, despite 21 trades are made, only 7 represent a positive return, however the mean profit during this period remains positive, which is mostly due to the fact that there are still highly profitable trades, such as the first trade concluded in November which has a profit of more than 22%. Also responsible for the positive profit during this period is the fact that the non-profitable positions do not last long and are lower in value.

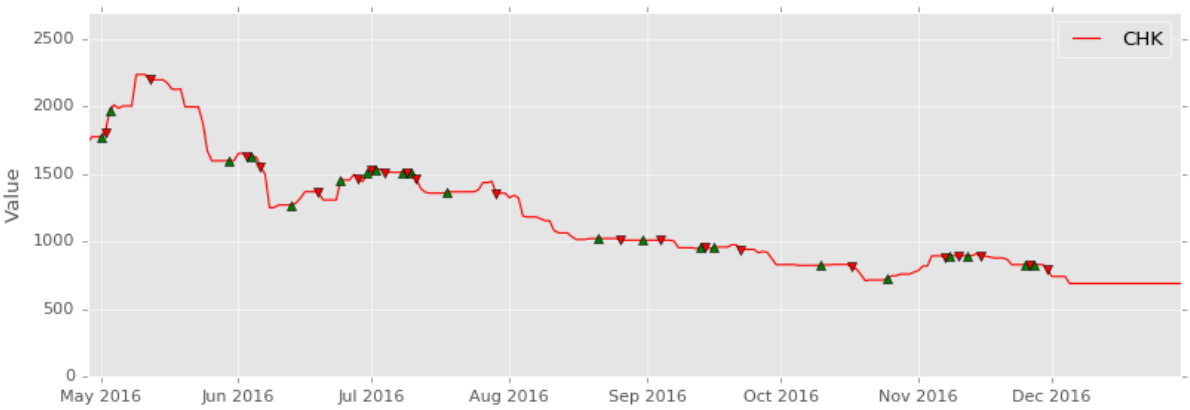

Figure 4-7: Buy and sell points resulting of the application of the Improved GA strategy to CHK (May 2016 - Dec 2016).

In the Table 4-11 are presented the parameters of the technical indicators obtained by both Genetic Algorithms, Table 4-12 presents the weights attributed to each indicator and the

Table 4-13 the required values to buy and sell the CDS.

Table 4-11: CHK - Indicators parameters for both evolutionary strategies.

Indicator	Parameters	Standard GA	Improved GA
EMA	Number of Days	8	4
Double Crossover	Days Short EMA	32	46
	Days Long EMA	100	116
RSI	Number of Days	18	48
ROC	Number of Days	18	28
Percent Changes	Number of days	2	6
	Requirement to Buy	0.023	0.503
	Requirement to Sell	- 0.839	- 0.639

Table 4-12: CHK - Weights of indicators for both evolutionary strategies, in percentage.

Indicators	Standard GA	Improved GA
EMA	16.51	28.73
Double Crossover	8.84	10.96
RSI	8.43	0.98
ROC	12.32	23.41
Percent Changes	53.90	35.92

Table 4-13: CHK - Requirements to buy and sell for both evolutionary strategies.

	Standard GA	Improved GA
Requirement to Buy	0.269	0.178
Requirement to Sell	- 0.241	- 0.062

Analysing the tables, it is possible to observe that the Improved GA attributed a high number of days and a small weigh to the RSI indicator, suggesting that it is almost irrelevant to the trading strategy. On the opposite side, both strategies attribute the highest weight to the Percent Changes indicator, which states its validity when it comes to trade in markets with this kind of volatility. Finally, the requirement to buy the CDS when considering all the indicators is smaller in the Improved GA, but it also sells earlier than the Standard GA.

On a final note, it is also important to note that, despite the very attractive ROI, this kind of investment is extremely risky. If on one hand it is necessary to assure that the protection seller has the funds to pay the determined amount in case of the default of the company, on the other hand if one invests in the wrong moment, for example when the spreads for CHK were worthen around 10000 USD, it could entail to an extreme loss since they never worth as much again. Besides one must not forget that, to hold the contract, it would also be necessary to pay that CDS spread to the protection seller.

4.4.3 ALCA – Alcadon Group

This example takes on the results of the Table 4-5, aiming to understand why the system cannot obtain a good result in those cases. The Alcadon Group is a Swedish company involved in the electric components industry, offering a wide range of products such as cooper cables, optic fibre cables and optical fibre accessories [61]. The CDS spread of this company during the period where the test was performed is represented in the Figure 4-8. It is possible to observe that its spreads have been falling during this entire period. In fact, in the entire available dataset it is possible to observe that its spread never has been so low and that the fall has begun in 2012, coinciding with the time at which the company

was acquired by DistIT [58].

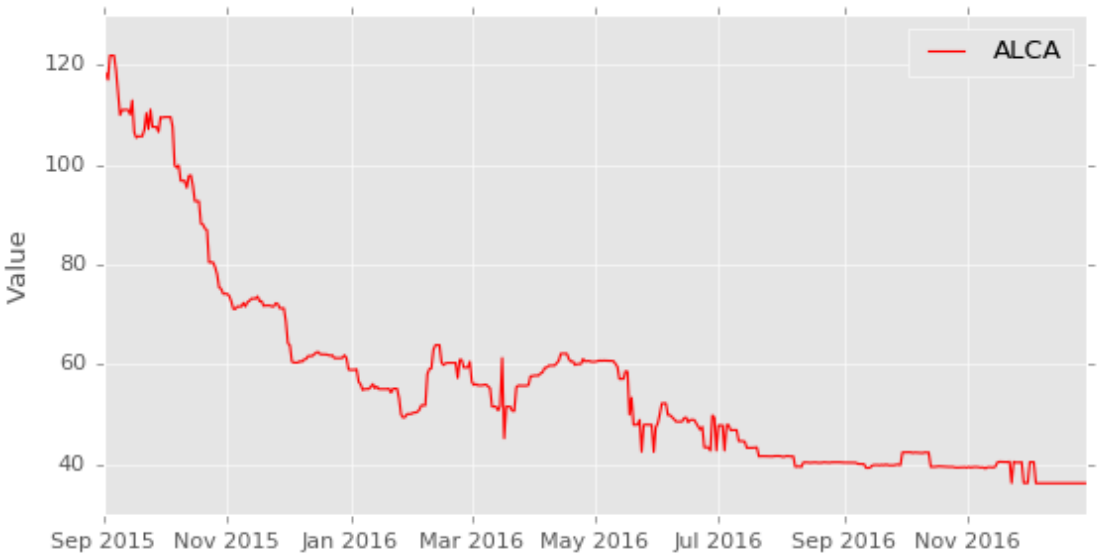

Figure 4-8: CDS spread of ALCA during the period 09/2015 - 12/2016.

Similarly to what has been done in Chapter 4.4.2, the strategies presented in Chapter 4.3.3 were applied to this data set, and the obtained results are summarized in the Table 4-14.

Table 4-14: Classification Parameters for ALCA example.

Parameter	Random	Literature	Standard GA	Improved GA
Number of Trades	62	15	74	44
Profitable Trades (%)	24.19	26.67	21.62	31.82
Non- Profitable Trades (%)	75.81	73.33	78.39	68.18
Average Profit per Trade (%)	-1.03	-1.96	-1.29	-0.89
Most Profitable Trade (%)	6.30	2.25	27.62	14.08
Biggest Loss (%)	-26.30	-26.30	-13.07	-12.58
ROI (%)	-50.29	-28.63	-65.18	-34.86

Analysing the Table 4-14 it is clear that none of the strategies can perform well during the referred period, since they all have negative ROIs and a negative average profit per trade. The fact that the Literature strategy performs fewer trades can be the reason why it achieves the higher ROI, nonetheless the Improved GA achieves a higher number of profitable trades and a higher average profit for each trade. The evolution of the ROI is also presented in Figure 4-9, where it is possible to see that the Improved GA has the best ROI until November, however two non-profitable trades make its value decrease and be surpassed by the Literature strategy.

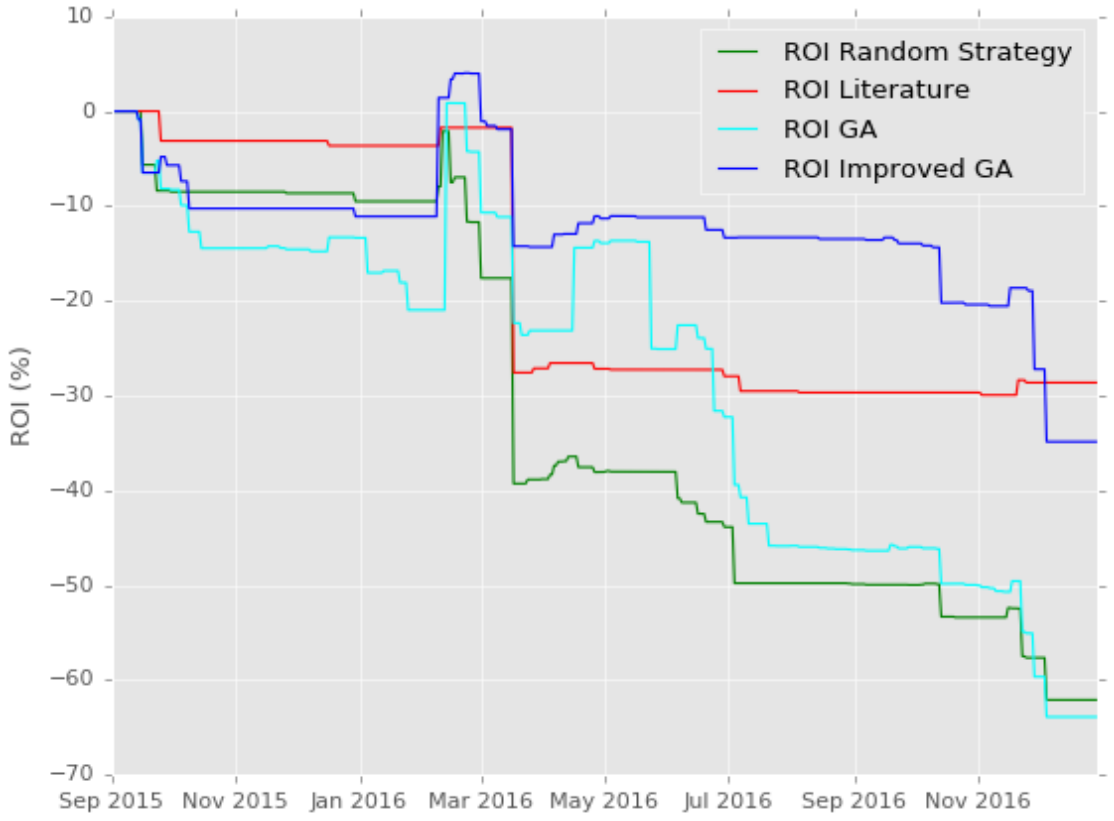

Figure 4-9:Evolution of the Return on the Investment for the ALCA example.

Just as in the previous example, Figure 4-10, Figure 4-11 and Figure 4-12 divide the period in analysis in three and show the buy and sell points defined by the Improved GA strategy.

Figure 4-10 represents the begging of the test period, were it is possible to see a decrease in the value of the spread of more than 50%. During this period, only one trade is profitable.

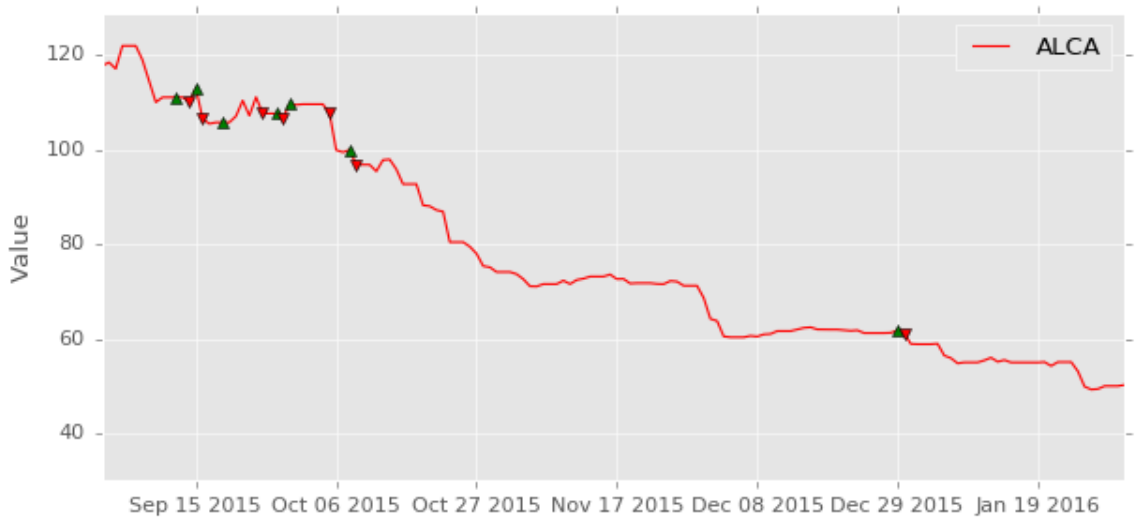

Figure 4-10: Buy and sell points resulting of the application of the Improved GA strategy to ALCA (Sep 2015 - Jan 2016).

Figure 4-11 represents the period where the strategy can perform the higher number of profitable trades, and even though the most part is residual it can still do a trade with a 14% profit during February. However, during periods of more stability the algorithm does not have a chance to make higher profits and when the price has high fluctuations, for example in March, it might end up with considerable losses.

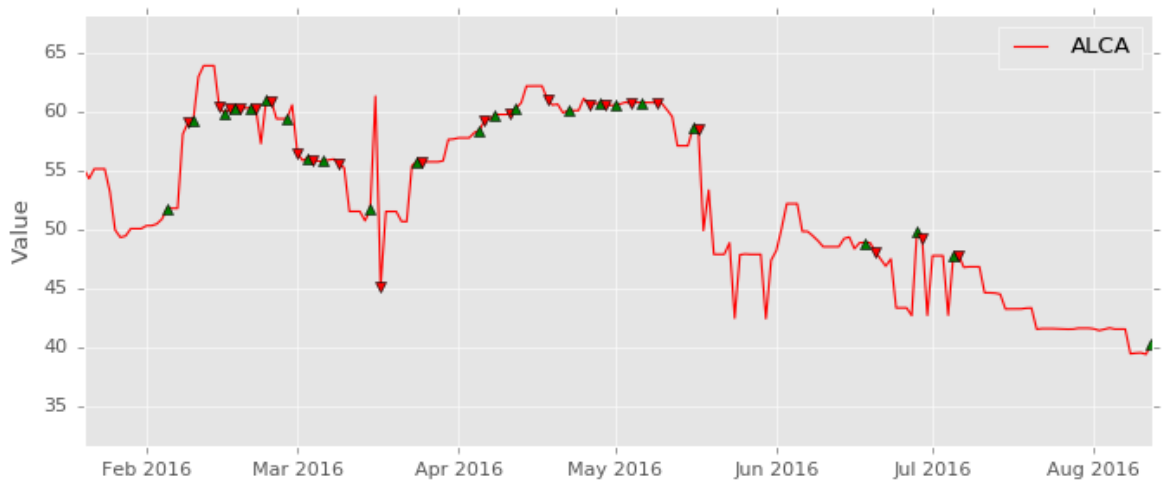

Figure 4-11: Buy and sell points resulting of the application of the Improved GA strategy to ALCA (Jan 2016 - Aug 2016).

The end of the test period represented in Figure 4-12 contains a period of more stability in the spread, and as in the previous time span, the algorithm is not able to be profitable, accumulating a number of residual losses that contribute for the diminishing of the ROI. At the end of the test period it is possible to see two trades that represent a loss of 10% each.

Figure 4-12: Buy and sell points resulting of the application of the Improved GA strategy to ALCA (Aug 2016 - Dec 2016).

In conclusion, the main reason why the system is uncapable of achieving a positive ROI during this period is mostly due to the fact that it does not perform so well during times when the spreads are falling

or periods of more stability. One of the possible reasons behind this fact is that the system trains during a period of high instability of the firm, with considerably higher spreads and high volatility, as presented in Figure 4-13.

Figure 4-13: CDS spread for ALCA during the train period (2008 -2015).

An interesting approach to do when the spreads are falling could be to sell protection instead of buying it. Looking at Figure 4-10, if Investor 1 makes a contract where he is selling protection on ALCA to an Investor 2 at the October 6, for example, where the Investor 2 is responsible to pay the spread value at that time (about 100 USD), and if in October 27 the Investor 1 buys protection to ALCA at the current trading spread (about 80 USD), it is possible to conclude that Investor 1 has made a profitable business. Investor 2 has to do periodic payments of 100 USD to Investor 1 and he only has to pay 80 USD to secure the same company. This could be compared to adopt a short position in the stock market, that simply is to sell an asset that one does not own at a higher price, expecting to buy it at a lower price in the future.

4.4.4 BKT – BlackRock Income Trust Inc.

After presenting examples of the best and worst situations that can happen in the CDS market, this example will take on the company that achieved the median ROI during the test performed in Chapter 4.4.1, more specifically BlackRock Income Trust, aiming to describe the most frequent situation in the CDS market. BKT is a diversified, closed-end management investment company, with a portfolio composed by corporate bonds, United States treasury obligations, municipal bonds, and short-term securities, among others. Its CDS spread for the period in analysis is presented in Figure 4-14.

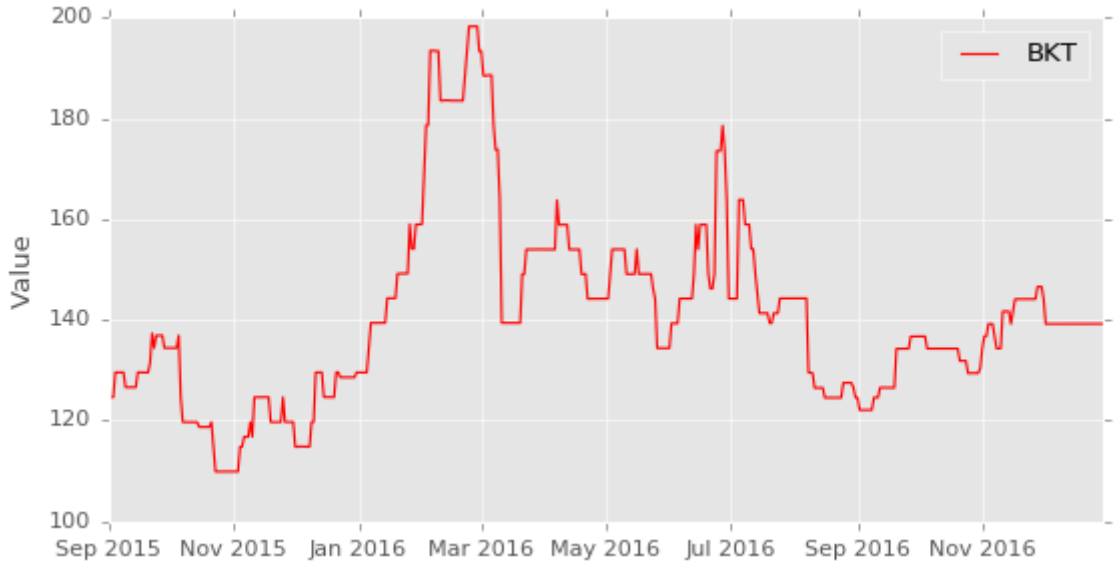

Figure 4-14:CDS spread of BKT during the period 09/2015 - 12/2016.

As it has been done on the previous examples, Table 4-15 contains the defined classification parameters for all the strategies applied to BKT, during the test period.

Table 4-15: Classification Parameters for BKT example.

Parameter	Random	Literature	Standard GA	Improved GA
Number of Trades	85	16	68	63
Profitable Trades (%)	47.06	37.50	30.88	33.33
Non- Profitable Trades (%)	52.94	62.50	69.12	66.67
Average Profit per Trade (%)	-0.12	-0.57	0.52	0.85
Most Profitable Trade (%)	8.26	2.58	21.60	21.66
Biggest Loss (%)	-9.37	-8.96	-12.53	-10.23
ROI (%)	-12.76	-9.27	32.41	58.31

Table 4-15 clearly states that both strategies based on evolutionary techniques perform much better than the two other strategies, achieving a considerable higher ROI. Despite their higher number of profitable trades, neither the Random and Literature strategies can achieve a positive average profit per trade. Figure 4-15 shows the evolution of the ROI for BKT during the period in analysis.

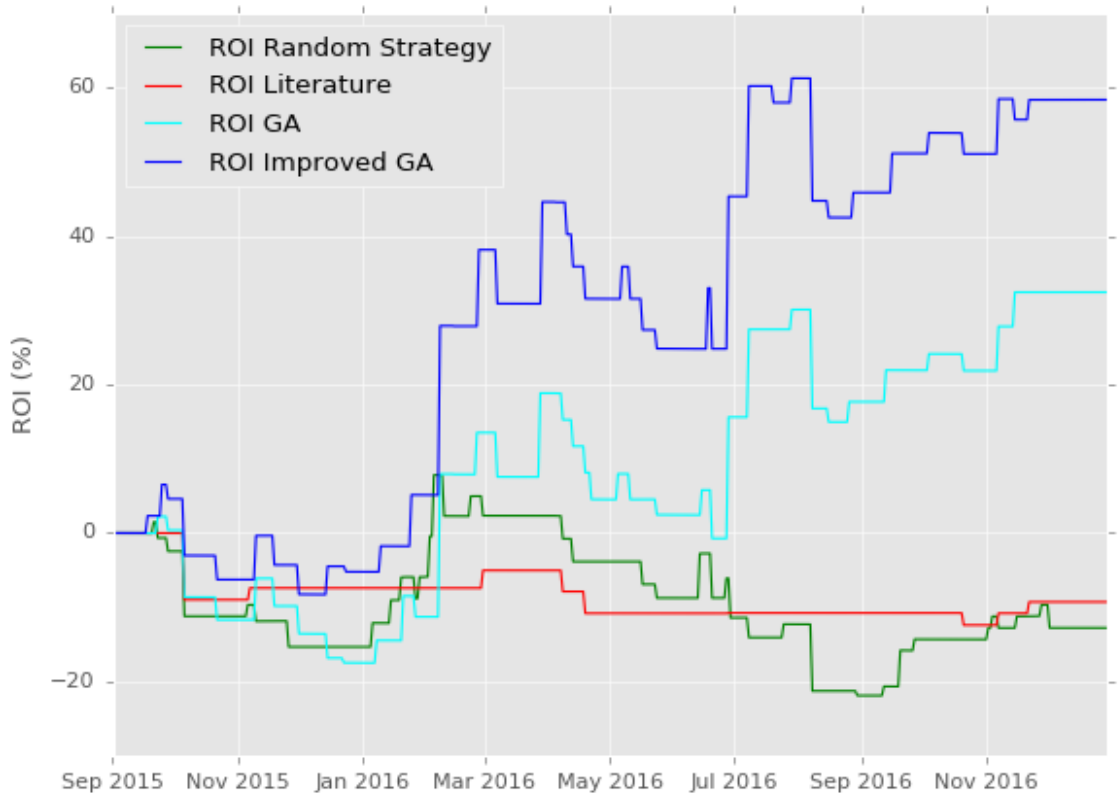

Figure 4-15: Evolution of the Return on the Investment for the BKT example.

As observed in Figure 4-15, the reason why both evolutionary techniques have a higher ROI is mostly due to two investments, one in February 2016 and the other at the end of June/starting of July. Observing the spread variation of BKT in the Figure 4-14, it is possible to observe that these periods represent a time where the spread has experienced an explosive growth, that both the GA based strategies can take advantage of to increase their ROI. However, when the spread falls abruptly as in the case of August 2016, they both experience higher losses.

Figure 4-16 represents the buy and sell points of the first months of the period in analysis. As in the previous examples, it is possible to observe that this strategy continues to make profitable trades when the spreads rise but when they fall quickly it might have considerable losses, like the one in the begging of October. Although, these losses are a consequence of trading in a market like this, where the spreads can rapidly rise or fall, and state the importance of using an indicator that measures volatility. Taking a look at the referred trade, it is possible to note that the CDS was sold only one day after the fall started, meaning that the strategy reacts quickly to these variations avoiding even higher losses.

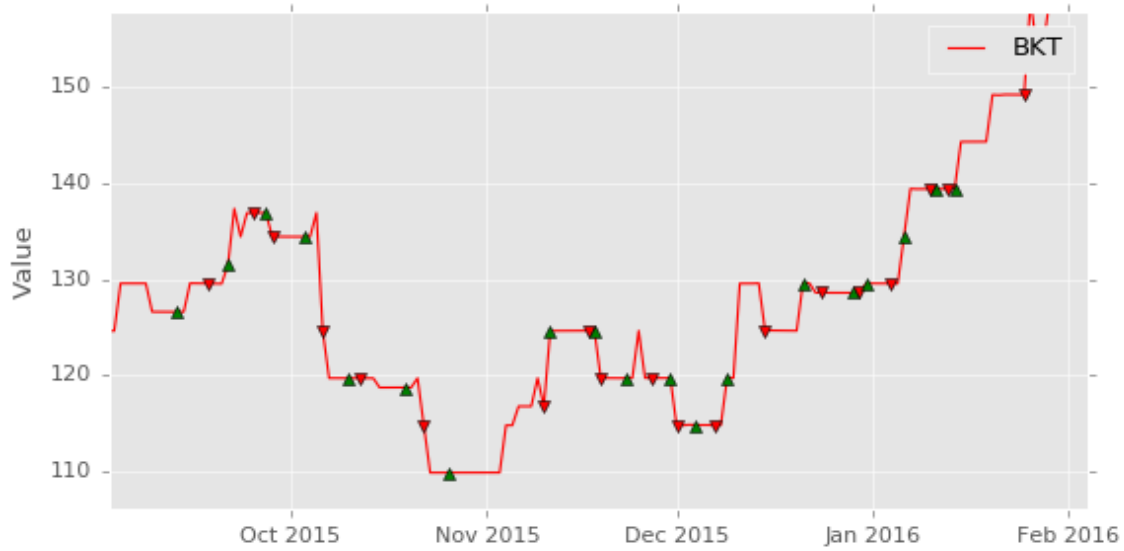

Figure 4-16: Buy and sell points resulting of the application of the Improved GA strategy to BKT (Sep 2015 - Jan 2016).

The subsequent months in analysis are displayed in Figure 4-17.

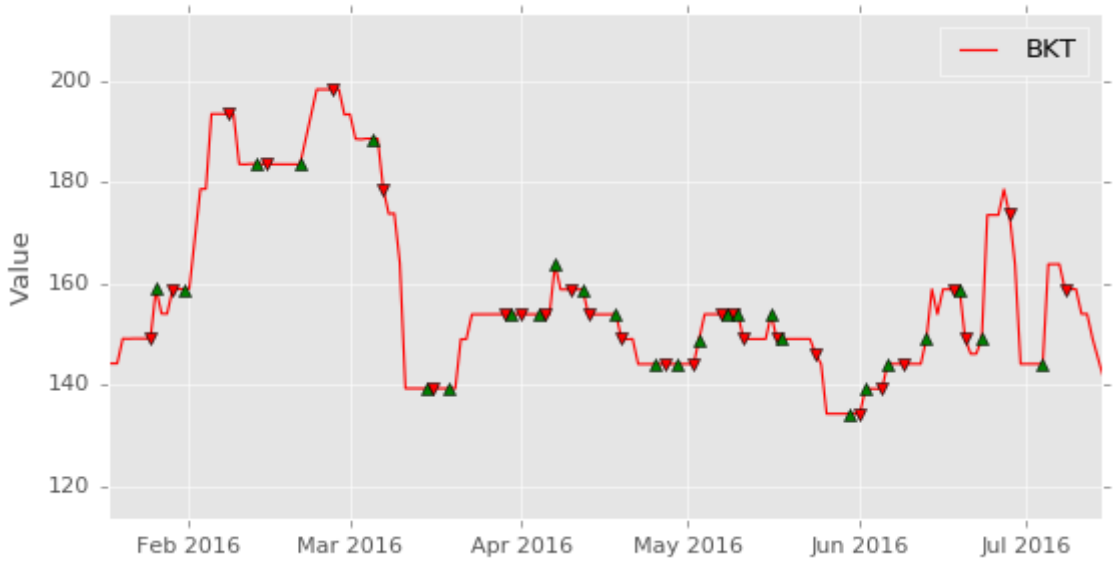

Figure 4-17: Buy and sell points resulting of the application of the Improved GA strategy to BKT (Feb 2016 - Jul 2016).

During this period, the system is capable of increasing its ROI mostly due to the trades performed in the begging of February and in the end of June. The final period is represented in Figure 4-18.

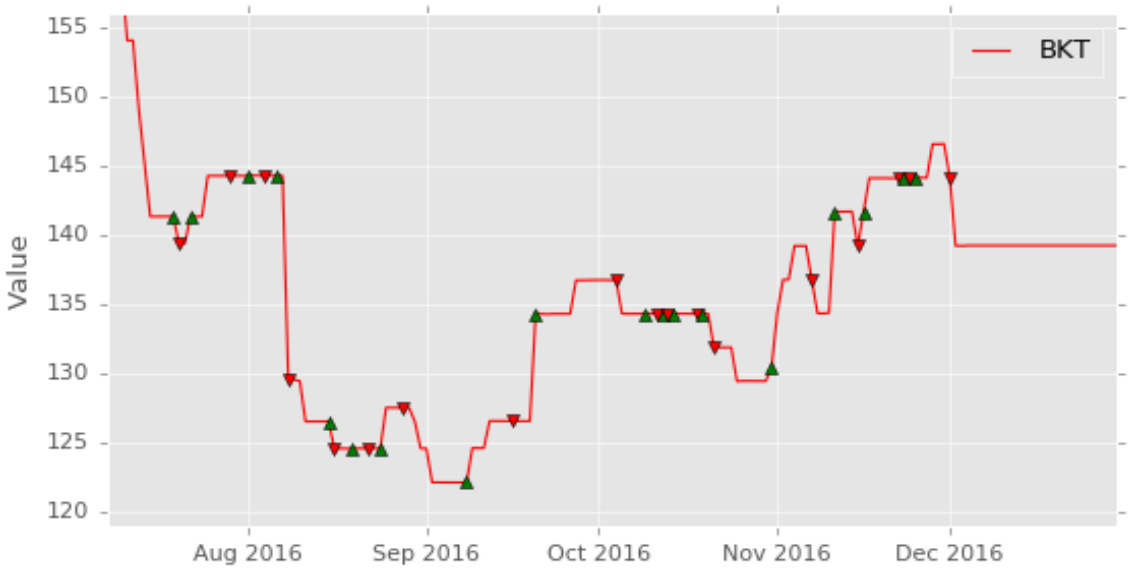

Figure 4-18: Buy and sell points resulting of the application of the Improved GA strategy to BKT (Aug 2016 - Dec 2016).

This period contains a fall in the spread during August, causing a drop in the ROI from whom the strategy cannot recover. Even though this period contains profitable trades, the fact that the higher, registered in November, has only a profit of around 5% makes it difficult to reach the ROI verified at the begging of this period. However, once the strategy was able to take advantage of some spread rises between February and August, it still ends with a high ROI. It is also important to notice that the spreads of this example are far from the ones of the Chapter 4.4.2, which could be seen as a sign that invest in BKT is less risky than investing in CHK, and might come as an alternative to investors that are not willing to take the risk of trading with such high spreads.

4.5 Chapter Conclusions

This chapter presents the tests performed to the developed strategy and compares it with other three strategies, one random, another based on the most commonly used technical indicators and another using a standard GA. Those strategies are compared based on the Return on Investment generated by each one, and on some defined Classification Parameters.

From the first case study it is possible to confirm that trading on the CDS market can be very profitable, nevertheless it is crucial to perform a correct choice of the CDS to invest in, meaning that one could end up with a huge profit but also with significant losses. Taking this results in consideration, another three tests were performed to three different companies, the one that performed the best during the case study, the median performing company representing the average situation, and the worst performing company, in order to see why the algorithm could not be profitable in this case.

Analysing these examples, it was clear that the algorithm could achieve a good performance whenever

the market is rising, but it has difficulties in being profitable on the opposite situation. A possible solution to this last scenario could be to sell protection when the spread starts to fall, which would make the contract more valuable when the spread eventually reached a lower value.

Nevertheless, the studies show that the Evolutionary Strategies can achieve better returns than the others, especially the developed strategy using the alterations to the Standard GA, namely Hypermutation, Elitism and Random Immigrants. Those alterations allow to the Improved GA strategy to react quickly when the market is raising rapidly. Another conclusion is that the evolutionary strategies tend attribute more weight to indicators that measure volatility, as is the case of Percent Changes.

Chapter 5

Conclusions

This chapter finalizes this work, summarizing the most relevant aspects, pointing out the most relevant conclusions and suggesting some aspects to be developed in Future Works. .

5.1 Final Conclusions

In this work it was developed a trading strategy using Genetic Algorithms and Technical Analysis, to be applied in the Credit Default Swaps market.

Credit Default Swaps are a relatively new developed financial instrument, which has received a lot of attention by the investors mainly during the last decade. A CDS can easily be seen as an insurance against a default of a certain company, where the person that buys the CDS pays a regular fee to the CDS seller, called the CDS Spread. The value of this spread varies with time, becoming higher if the company conditions deteriorate and when that happens the buyer can resell the CDS, generating a profit from that trade. Considering this, there is an opportunity to develop trading strategies to this market and, taking into account what has been done in other financial markets, those strategies can be improved through the application of Technical Analysis and Machine Learning.

Using Technical Analysis and through a set of indicators, it is possible to forecast future price changes using past data. The technical indicators used in this work were some of the most common, such as the EMA, RSI and ROC, but it was also introduced a new one, the Percent Changes, which rely on the calculation of price variation in the last days and how encouraging that variation was to decide whether to invest or not.

Afterwards, a Genetic Algorithm was used to optimize these trading rules, that is the values to use for each parameter which achieve a higher profit. This work conducted tests on two implementations of GA, one more similar to its original formulation and another including some enhancements found in literature. The performed tests shown that, using those enhancements, the results can improve in terms of profitability.

Furthermore, the obtained results show that investing in the CDS market can be really profitable, but also really risky due to its unpredictable and abnormal changes in the spreads. It is important to invest carefully and be sure of the ability of the protection seller to pay their obligations in case of default, and at the same time one must think twice before buy a CDS with a high spread value, because as spreads raise rapidly, they might fall quickly as well, meaning that the investment done might never be recovered.

5.2 Improvements and Future Work

As previously stated, there is still a lot of work that can be developed in the CDS Market. Besides, the developed algorithm can be improved by applying some alterations and creating new features. The following are suggested:

- Create an optimization portfolio strategy composed only by CDS where variables such as the risk are considered, in addition to the profitability, considered during this work. That way it would be possible to perform a multi-objective optimization, maximizing the ROI and minimizing the risk.
- Use parallel processing to speed up the algorithm, since the evaluation of the individuals is a slow process, having them distributed through different cores could really improve the overall performance of the algorithm.
- Increase the number of Technical Indicators used, which could be done simply by increasing the size of the chromosome and defining the score attribution rules for that indicator.
- Explore the reaction of CDS spreads to news using Sentiment Analysis and use that to define the investment strategy. The use of Fundamental Analysis could also be considered in order to find out if the company conditions are getting worse.
- Apply other AI techniques to perform the optimization of the technical rules.
- Apply the developed algorithm to other markets, such as Commodities, and more specifically to Sugar. Through the observation of its price variation it is possible to conclude that it has a similar behaviour to some CDS spreads, with some abrupt uptrends that have proven to be profitable when applying the algorithm.

References

[1] S. B. Achelis, “Technical Analysis from A to Z,” *Search*, vol. 77, pp. 33–4, 2000.

[2] J. J. Murphy, *Intermarket technical analysis: trading strategy for the Global Technical and Currency Markets*. 1991.

[3] M. Chui, “Derivatives markets , products and participants : an overview,” *IFC Bull.*, no. 35, pp. 3–11, 2008.

[4] N. Terzi and K. Uluçay, “The role of credit default swaps on financial market stability,” *Procedia - Soc. Behav. Sci.*, vol. 24, pp. 983–990, 2011.

[5] A. McKay, *The Big Short*. USA, 2015.

[6] D. Y. Tang and H. Yan, “Understanding transactions prices in the credit default swaps market,” *J. Financ. Mark.*, vol. 32, pp. 1–27, 2017.

[7] P. Augustin, “Credit Default Swaps: A Survey,” *Found. Trends® Financ.*, vol. 9, no. 1–2, pp. 1–196, 2014.

[8] R. McDonald and A. Paulson, “AIG in Hindsight,” No. 21108, 2015.

[9] R. M. Stulz, “Credit Default Swaps and the Credit Crisis,” *J. Econ. Perspect.*, vol. 24, no. 1, pp. 73–92, 2010.

[10] L. Norden, “Information in CDS spreads,” *J. Bank. Financ.*, vol. 75, pp. 118–135, 2017.

[11] J. J. Murphy, *Technical Analysis Of The Financial Markets*. 1999.

[12] P. Fernández-Blanco, D. J. Bodas-Sagi, F. J. Soltero, and J. I. Hidalgo, “Technical market indicators optimization using evolutionary algorithms,” *Proc. 2008 GECCO Conf. companion Genet. Evol. Comput. - GECCO ’08*, p. 1851, 2008.

[13] A. Silva, R. Neves, and N. Horta, “A hybrid approach to portfolio composition based on fundamental and technical indicators,” *Expert Syst. Appl.*, vol. 42, no. 4, pp. 2036–2048, 2015.

[14] M. Lam, “Neural network techniques for financial performance prediction: Integrating fundamental and technical analysis,” *Decis. Support Syst.*, vol. 37, no. 4, pp. 567–581, 2004.

[15] L. B. Booker, D. E. Goldberg, and J. H. Holland, “Classifier systems and genetic algorithms,” *Artif. Intell.*, vol. 40, no. 1, pp. 235–282, 1989.

[16] K. F. Man, K. S. Tang, and S. Kwong, “Genetic algorithms: Concepts and applications,” *IEEE Trans. Ind. Electron.*, vol. 43, no. 5, pp. 519–534, 1996.

[17] DEAP Project, “One Max Problem,” 2017. [Online]. Available: https://deap.readthedocs.io/en/master/examples/ga_onemax.html. [Accessed: 11-Sep-2017].

[18] M. R. Noraini and J. Geraghty, “Genetic algorithm performance with different selection strategies in solving TSP,” vol. 70, no. 12, pp. 5–10, 2011.

[19] G. Syswerda, “Uniform Crossover in Genetic Algorithms,” in *Proceedings of the 3rd International Conference on Genetic Algorithms*, 1989, pp. 2–9.

[20] D. Project, “Operators and Algorithms,” 2017. [Online]. Available: http://deap.readthedocs.io/en/master/tutorials/basic/part2.html. [Accessed: 05-Sep-2017].

[21] R. Blanco, R. Blanco, S. Brennan, S. Brennan, I. W. Marsh, and I. W. Marsh, “An empirical analysis of the dynamic relationship between investment-grade bonds and credit default swaps,” *J. Finance*, vol. 60, no. 5, pp. 2255–2281, 2004.

[22] J. Hull, M. Predescu, and A. White, “The relationship between credit default swap spreads, bond yields, and credit rating announcements,” *J. Bank. Financ.*, vol. 28, no. 11, pp. 2789–2811, 2004.

[23] F. Longstaff, S. Mithal, and E. Neis, “Corporate Yield Spreads: Default Risk of Liquiidty? New evidence form the credit default market,” *J. Finance*, vol. 60, no. 5, pp. 2213–2253, 2005.

[24] H.-G. Fung, G. E. Sierra, J. Yau, and G. Zhang, “Are the U.S. stock market and credit default swap market related? Evidence from the CDX indices.,” *J. Altern. Investments*, vol. 11, pp. 43–61, 2008.

[25] J. Ericsson, K. Jacobs, and R. Oviedo, “The Determinants of Credit Default Swap Premia,” *J. Financ. Quant. Anal.*, vol. 44, no. 1, p. 109, 2009.

[26] D. Avino and O. Nneji, “Are CDS spreads predictable? An analysis of linear and non-linear forecasting models,” *Int. Rev. Financ. Anal.*, vol. 34, pp. 262–274, 2014.

[27] I. Kaastra and M. Boyd, “Designing a neural network for forecasting financial and economic time series,” *Neurocomputing*, vol. 10, no. 3, pp. 215–236, 1996.

[28] S. Thawornwong, D. Enke, and C. Dagli, “Neural Networks as a Decision Maker for Stock Trading: A Technical Analysis Approach,” *Int. J. Smart Eng. Syst. Des.*, vol. 5, no. 4, pp. 313–325, 2003.

[29] K. Kim, “Financial time series forecasting using support vector machines,” *Neurocomputing*, vol. 55, no. 1–2, pp. 307–319, 2003.

[30] F. E. H. Tay and L. J. Cao, “Support Vector Machine With Adaptive Parameters in Financial Time Series Forecasting,” vol. 14, no. 6, pp. 1506–1518, 2004.

[31] M. R. Hassan and B. Nath, “Stock market forecasting using hidden Markov model: a new approach,” *Intell. Syst. Des. Appl. 2005. ISDA’05. Proceedings. 5th Int. Conf.*, pp. 192–196, 2005.

[32] Z. Huang, H. Chen, C.-J. Hsu, W.-H. Chen, and S. Wu, “Credit rating analysis with support vector

machines and neural networks: a market comparative study," *Decis. Support Syst.*, vol. 37, no. 4, pp. 543–558, 2004.

[33] R. C. Cavalcante, R. C. Brasileiro, V. L. F. Souza, J. P. Nobrega, and A. L. I. Oliveira, "Computational Intelligence and Financial Markets: A Survey and Future Directions," *Expert Syst. Appl.*, vol. 55, pp. 194–211, 2016.

[34] J. Y. Potvin, P. Soriano, and V. Maxime, "Generating trading rules on the stock markets with genetic programming," *Comput. Oper. Res.*, vol. 31, no. 7, pp. 1033–1047, 2004.

[35] K. J. Oh, T. Y. Kim, and S. Min, "Using genetic algorithm to support portfolio optimization for index fund management," *Expert Syst. Appl.*, vol. 28, no. 2, pp. 371–379, 2005.

[36] A. Gorgulho, R. Neves, and N. Horta, "Applying a GA kernel on optimizing technical analysis rules for stock picking and portfolio composition," *Expert Syst. Appl.*, vol. 38, no. 11, pp. 14072–14085, 2011.

[37] K. Kim and I. Han, "Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index," *Expert Syst. Appl.*, vol. 19, no. 2, pp. 125–132, 2000.

[38] H. Kim and K. Shin, "A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets," *Appl. Soft Comput.*, vol. 7, no. 2, pp. 569–576, 2007.

[39] R. Choudhry and K. Garg, "A Hybrid Machine Learning System for Stock Market Forecasting," *World Acad. Sci. Eng. Technol.*, vol. 2, no. 15, pp. 315–318, 2008.

[40] Y. Li and H. Xu, "Credit Rating Analysis with Support Vector Machines Optimized by Genetic Algorithm," *4th Int. Conf. Wirel. Commun. Netw. Mob. Comput.*, pp. 1–4, 2008.

[41] J. J. Grefenstette, "Genetic Algorithms for Changing Environments," *Parallel Probl. Solving from Nat.*, vol. 2, no. 1992, 1992.

[42] H. G. Cobb, "An Investigation into the Use of Hypermutation as an Adaptive Operator in Genetic Algorithms Having Continuous, Time-Dependent Nonstationary Environments," *Tech. Rep.*, pp. 1–19, 1990.

[43] H. G. Cobb and J. J. Grefenstette, "Genetic algorithms for tracking changing environments," *Proc. 5th Int. Conf. Genet. Algorithms*, no. 1992, pp. 523–530, 1993.

[44] F.-A. Fortin, F.-M. De Rainville, M.-A. Gardner, M. Parizeau, and C. Gagné, "DEAP : Evolutionary Algorithms Made Easy," *J. Mach. Learn. Res.*, vol. 13, pp. 2171–2175, 2012.

[45] Erasmus Data Service Center of the Erasmus University Rotterdam, "Datastream - CDS Spreads." [Online]. Available: https://datateamoftheeur.files.wordpress.com/2016/08/cds-datastream.pdf. [Accessed: 02-Sep-2017].

[46] J. Bai and P. Collin-dufresne, "The CDS-Bond Basis The CDS-Bond Basis," no. November 2009, 2013.

[47] K. Jebari and M. Madiafi, “Selection Methods for Genetic Algorithms,” vol. 3, no. December, pp. 333–344, 2013.

[48] Y. Fang and J. Li, “A Review of Tournament Selection,” pp. 181–192, 2010.

[49] H. Cartwright, “Using artificial intelligence in chemistry and biology: A practical guide,” 2008.

[50] O. Kramer, “Genetic Algorithm Essentials,” vol. 679, pp. 11–20, 2017.

[51] Omni Calculator, “Normal Distribution.” [Online]. Available: https://www.omnicalculator.com/statistics/normal-distribution. [Accessed: 27-Sep-2017].

[52] R. L. Haupt, “Optimum population size and mutation rate for a simple real genetic algorithm that optimizes array factors,” in *Antennas and Propagation Society International Symposium, 2000. IEEE*, 2000, vol. 2, pp. 1034–1037.

[53] J. Ni and C. Zhang, “An efficient implementation of the backtesting of trading strategies,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3758 LNCS, pp. 126–131, 2005.

[54] A. E. Biondo, A. Pluchino, A. Rapisarda, and D. Helbing, “Are Random Trading Strategies More Successful than Technical Ones?,” *PLoS One*, vol. 8, no. 7, pp. 1–17, 2013.

[55] J. B. Maverick, “What are the most common periods used in creating Moving Average (MA) lines?” [Online]. Available: https://www.investopedia.com/ask/answers/122414/what-are-most-common-periods-used-creating-moving-average-ma-lines.asp. [Accessed: 13-Dec-2017].

[56] StockCharts.com, “Rate of Change,” 2017. [Online]. Available: http://stockcharts.com/school/doku.php?id=chart_school:technical_indicators:rate_of_change_roc_and_momentum. [Accessed: 13-Dec-2017].

[57] A. Ruttiens, “Relative strength index,” 2010. [Online]. Available: https://www.investopedia.com/terms/r/rsi.asp. [Accessed: 13-Dec-2017].

[58] J. Holland, “Genetic Algorithms,” *Sci. Am.*, 1989.

[59] The Street, “Stressed Out: Is This the Death Knell for Chesapeake Energy?,” 2016. [Online]. Available: https://www.thestreet.com/story/13451239/1/stressed-out-is-this-the-death-knell-for-chesapeake-energy.html. [Accessed: 14-Sep-2017].

[60] Reuters, “Alcadon Group AB (ALCA.ST),” 2017. [Online]. Available: http://www.reuters.com/finance/stocks/overview/ALCA.ST. [Accessed: 10-Oct-2017].

[61] Alcadon Group, “Historik,” 2017. [Online]. Available: http://alcadongroup.se/om-alcadon/historik/. [Accessed: 10-Oct-2017].