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

João Barata
Instituto de Telecomunicações
Instituto Superior Técnico
Lisboa, Portugal

***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 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.***

## I. INTRODUCTION

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", 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, 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.

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. 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". 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.

This paper is structured as follows: Section 1 – Introduction; Section 2 – Background and State-of-the-Art - presents fundamental concepts, regarding financial knowledge and genetic algorithms, and some relevant works developed in the recent years, in these fields; Section 3 – Proposed Architecture – contains the explanation of the proposed system architecture, explaining the main modules of the system and its behaviour; Section 4 – System Evaluation – shows the performed system validation through its application on the CDS market; Section 5 – Conclusions – summarizes the work and presents some suggestions of future work.

## II. BACKGROUND AND STATE-OF-THE-ART

As stated already, this paper 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, as well as some works in these fields.

### A. *Credit Default Swaps*

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 certain Company, in exchange for a fee or a premium, called the CDS spread,

over the maturity of the CDS contract [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 [2]. 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 [1]. 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.

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 [2]. Essentially CDOs repackage individual loans, like credit card debts, mortgages or corporate debt, and sell it to investors.

The default of a protection seller 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. 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 [1].

After the period of crisis in 2007, 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. 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, 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 [3], 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.

### *B. Technical Analysis*

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.

In his book, John J. Murphy [4] stated that the fundamentalist studies the cause of market movement, and the technician is more interested about the effect. In this work it will only be used the Technical Analysis, although, these two are not excusive, meaning that can be used together [4].

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 [4]. 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.

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.

The technical indicators used in this thesis will be presented in Section III.B.

### *C. Genetic Algorithms*

A Genetic Algorithm [5] 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, iteratively refined, until a 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, with the best ones being 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 their bit strings. On the other hand, mutation creates an alteration to one or more genes in a chromosome, preventing them to became too similar to each other, which promotes evolution and avoids the algorithm to converge to some local maxima/minima. 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 0.2 [6]. 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.

In Figure II-1 are represented the main stages of a GA.

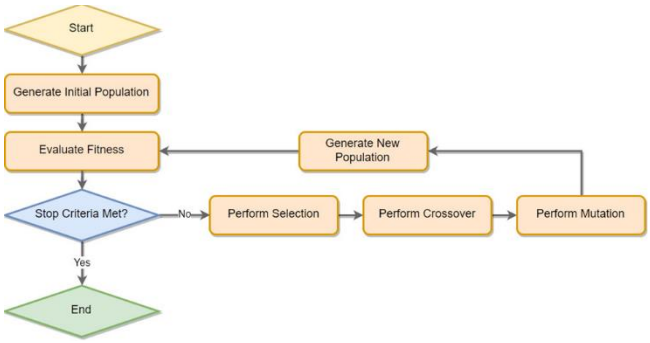


*Figure II-1: Genetic Algorithm flow chart.*

### D. State-of-the-Art

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. Although, to the author's knowledge, there are not works combining this instrument with GA, and for that reason this section will present works in these fields separately.

#### 1) Works on CDS

An application of the CDS market was presented by Hull et al. [7], 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.

Fung et al. [8] studied the relationship between the stock market and high yield and investment grades and the CDS markets in the USA 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 the stock and bond markets at the time.

Avino et al. [9] 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, MAR 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.

#### 2) Works on GA

After their presentation in 1989 [10], 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.

Considering the difficulty of choosing the appropriate parameter values for technical indicators, Fernández-Blanco et al. [11] 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.

Gorgulho et al. [12] 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 work.

It is also commonly found in the literature the application of GA with other AI techniques. Choudhry et al. [13] 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 [14] 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 if a 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. John J. Grefenstette along with Hellen Cobb introduced some new features, namely Hypermutation, which increases the mutation rate whenever it is triggered, and Random Immigrants, which replace a part of the population, usually the least fit in each generation, with randomly generated values. Those mechanisms were tested in [15] and proven to have advantages over the standard GA. In this work, the implementation of these mechanisms will be considered.

## III. Proposed Architecture

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, depending on each other as represented in Figure III-1.

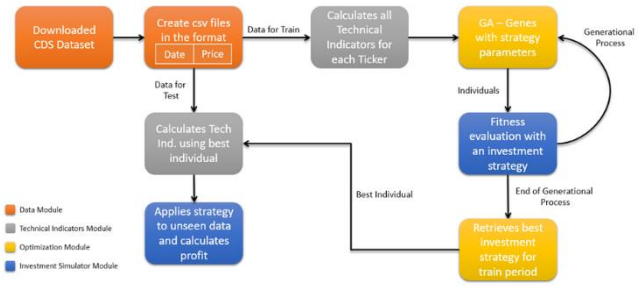


*Figure III-1: Main phases of the system.*

### A. Data Module

The CDS spreads data was obtained from two sources, CMA Datavision and Thomson Reuters CDS. 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.

The downloaded csv is 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 train and test periods will be defined, according to the available data of each ticker. 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.

### B. Technical Rules Module

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 Exponential Moving Average (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 parameters of the chosen technical indicators will be chosen by the GA, as well as the contribution (weight) of each for the final decision ($W_i$). A score system based on the one in [12] was also defined, depending on how much the indicators advise to buy or sell a CDS, this way four levels were defined:

- Very High – Strong buy signal, gives a score of 1 point;
- High – Buy signal, gives a score of 0.5 points;
- Low – Sell signal, gives a score of -0.5 points;
- Very Low – Strong sell signal, gives a score of -1 point;

These scores are attributed to each indicator daily. Then, based on the weight attributed to each indicator, the position will be calculated based on equation (1).

$$Position = \sum_{i=0}^{N} W_i \cdot Score(X, i) \qquad (1)$$

The following paragraphs present the selected indicators and the respective classification rules.

#### 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

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, when the price crosses bellow the EMA is a good time to sell. Based on this, it is possible to define the rules presented in the Table III-1.

*Table III-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 |

### 2) *Double Crossover Method*

This technical indicator uses two EMAs 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, is more sensible to market variations, reacting faster.

For trading using the Double Crossover Method, 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 III-2.

*Table III-2: Score attribution rules for the Double Crossover indicator.*

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

### 3) *Relative Strength Index*

The RSI 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.

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 III-3.

*Table III-3: 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 |

### 4) *Rate of Change*

ROC measures the speed at which prices are changing, i.e., the difference between the current price and the price from $n$ periods ago. 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 [4]. 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 III-4 for the ROC indicator.

*Table III-4: 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 |
| **Low Score** | ROC line is rising, price decreasing |
| **Very Low Score** | ROC line crosses below 0 |

### 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 III-5.

*Table III-5: 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 |

## C. *Optimization Module*

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.

### 1) *Chromosome Representation*

The chromosome is composed by a vector of floats with values in the interval [0,1]. Notwithstanding, some parameters of the technical indicators are integers, so 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 III-6. It is composed by the eight parameters of the technical indicators, 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 III-6: 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 |

*2) Selection*

The selection of the individuals will choose the individuals to mate or mutate, hoping the individuals of the subsequent generation have a higher fitness. From the different methods to perform selection it was chosen the Tournament. In the defined tournament, a set of 3 individuals are randomly chosen from the population, and compete against each other. The individual with the highest fitness wins and will be included as one of the next generation population.

*3) Crossover*

After the individuals are selected, the crossover is performed, forming a new offspring between two randomly selected "good parents". The technique used was two-point crossover, as it was designed to overcome some of the disadvantages of both uniform and one-point crossover, allowing the problem space to be searched more thoroughly. The crossover rate chosen was the most found on the literature, 50%.

*4) Mutation*

Mutation is also responsible for the manipulation of the selected individuals to form the next generation. The mutation method chosen was the Gaussian Mutation, as it stands as the most popular operator for real value floating point vectors. To the mutation rate was defined the value 20%.

*5) Improvemnts*

Sometimes, crossover and mutation might generate weaker offspring when compared to the parents, and good candidates might end being discarded. 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 $N$ best individuals that ever lived in the population. At each generation, the hall of fame is updated.

Another improvement attempt to the GA was made by using the approach proposed in [15], namely the implementation of Random Immigrants. In this thesis the replacement rate chosen was 5% of the entire population.

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

## D. 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. As a result, it retrieves the profit made for that ticker with that investment strategy, along with the detailed graphic of the ticker, containing the technical indicators and a set of markers drawn in the CDS price, indicating either it is a buy or a sell order.

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.

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.

# IV. System Evaluation

This section presents the validation tests submitted to the proposed algorithm, performed using data from the CDS market. According to what was previously stated, the dataset will be divided in train (80 %), to find the best strategy, and test (20%) to confirm its validity in unseen data.

## A. 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.

*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, as stated by (2):

$$ROI = \frac{Gain - Cost}{Cost} \tag{2}$$

Due to the fact that ROI is measured as a percentage, it can be used to compare the profitability of different strategies.

*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 with other strategies. Below is presented a list of the used parameters:

- Number of Trades – The total number of trades done in a specific period;
- Percentage of Profitable Trades - Number of trades that generated a positive return, after deduction all the trade commissions, divided by the total number of trades;
- Percentage of Non-Profitable Trades - 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;
- Greatest profit - Represents the most profitable trade done;
- Greatest loss - Represents the biggest loss.

### 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 – Simulates a random behaviour, 150 random individuals are generated and tested in the same period of the other strategies. The best performing was chosen.
- Literature - This strategy was implemented with the intent of compare a strategy based on the typical values found for the technical indicators with the ones retrieved by the evolutionary strategy.
- Standard GA – This strategy is based on the standard implementation of the GA, using the same chromosome structure, crossover and mutation probability and selection scheme of the proposed architecture, however without the improvements to the algorithm previously introduced.
- Improved GA - Aiming to improve the results of the previous strategy, the features previously defined, Elitism, Random Immigrants and Hypermutation, were implemented.

## B. *Case Studies*

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

### 1) *CDS Market*

This case study consists on the application of the strategies to the CDS Market, more specifically to 177 companies selected after the pre-processing done by the Data Module. Those train and test periods are explicit in Table IV-1 and are also used in the subsequent examples.

*Table IV-1: Train and test periods.*

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

The parameters of the GAs for all the performed tests are summarized in the Table IV-2.

The performed tests led to the graph of the Figure IV-1, where it is possible to observe that both the GA outperform significantly the remaining strategies. 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. Table IV-3 summarizes the performance of the applied strategies, according to the previously defined parameters, for this Case Study.

*Table IV-2: Parameters of Genetic Algorithms.*

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

For a better comparison between the two evolutionary strategies, the most relevant genes of the best individuals are presented. To facilitate the reading, the best individual was mapped from the values used during the execution of the algorithms, to the ones defined for the technical indicators intervals. Table IV-4 and Table IV-5 show the obtained parameters for the used technical indicators.

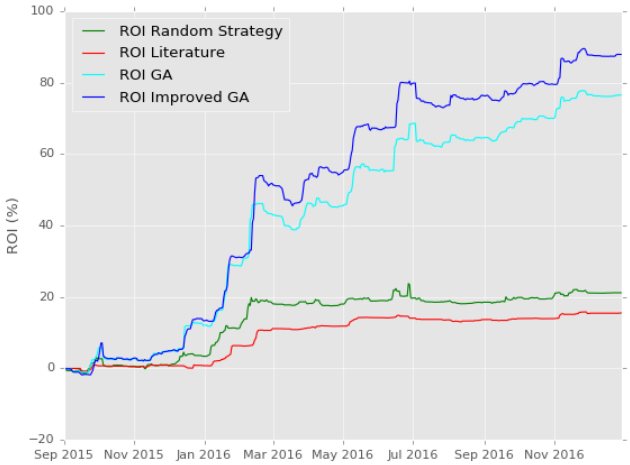


*Figure IV-1: Evolution of the Average ROI for the defined strategies.*

*Table IV-3: Classification Parameters for CS 1.*

| Parameter | Random | Literature | Standard GA | Improved GA |
|---|---|---|---|---|
| **Number of Trades** | 17462 | 2350 | 10387 | 10366 |
| **Profitable (%)** | 27.83 | 33.36 | 31.41 | 28.81 |
| **Non-Profitable (%)** | 72.17 | 66.64 | 68.59 | 71.19 |
| **Average Profit (%)** | 0.19 | 0.95 | 0.87 | 0.98 |
| **Most Profitable (%)** | 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 |

*Table IV-4: CS 1 - Indicators parameters for the GAs.*

| 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 |

*Table IV-5: CS 1 - Weights of indicators for GAs, 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 |

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 IV-5 shows that the most relevant indicator is the Percent Changes, in both cases, which confirms its suitability to volatile markets, as the CDS market. On the contrary, the RSI indicator is the least used.

The following examples will take on this case study and perform an analysis to three different scenarios, with the best and the worst performing companies and the one that achieved the median ROI.

*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, along with the fact that oil prices plunge to their lowest levels in a decade had contributed to a crash in their stock values and on the opposite side an explosive rise on their CDS spreads.

To have a clear visualization of this rise on the spread, Figure IV-2 presents the months that contain the referred period, along with the buy (green) and sell (red) points for the Improved GA strategy. It is possible to observe that the strategy is capable of take advantage of the situation to perform a trade with a return of almost 128%. Despite the three subsequent trades are not profitable due to the fall in the spread, it is able to continue to perform profitable trades when the spread rises.

Table IV-6 and Figure IV-3 show the result of the application of the different strategies to the entire dataset. It is clearly visible the rise in the spread during February in Figure IV-3. From the analysis of the Table IV-6 and the Figure IV-3, it is clear that the Improved GA strategy clearly outperforms the other strategies,

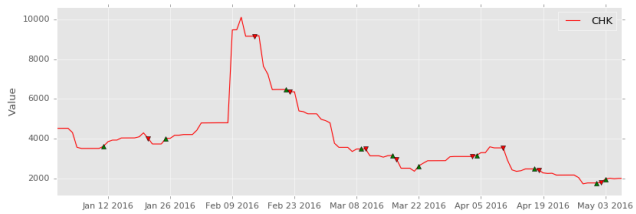


*Figure IV-2: Buy and sell points of the Improved GA strategy to CHK (Jan 2016 - Apr 2016).*

*Table IV-6: Classification Parameters for CHK example.*

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

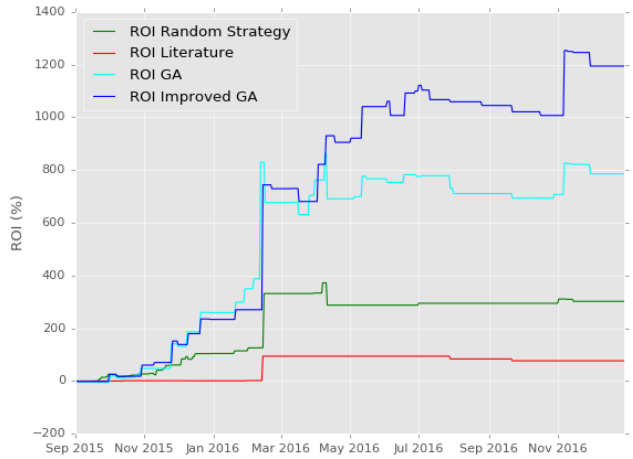


*Figure IV-3: Evolution of the ROI for the CHK example.*

achieving the highest ROI. The number of profitable trades is around 50% in all the strategies.

Table IV-7 presents the weights of the indicators for both GA strategies. Both strategies attribute the highest weight to the Percent Changes, which states its validity when it comes to trade in markets with this kind of volatility

*Table IV-7: CHK - Weights of indicators for both GAs.*

| 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 |

### 3) *ALCA – Alcadon Group*

This example focuses on the company to which the strategy performed worse. The Alcadon Group is a Swedish company involved in the electric components industry. Similarly to what has been done before, the defined strategies were applied to this data set, and the obtained results are summarized in the Table IV-8.

*Table IV-8: Classification Parameters for ALCA example.*

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

Analysing the Table IV-8 it is clear that none of the strategies can perform well during the test 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 presented in Figure IV-4, 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.

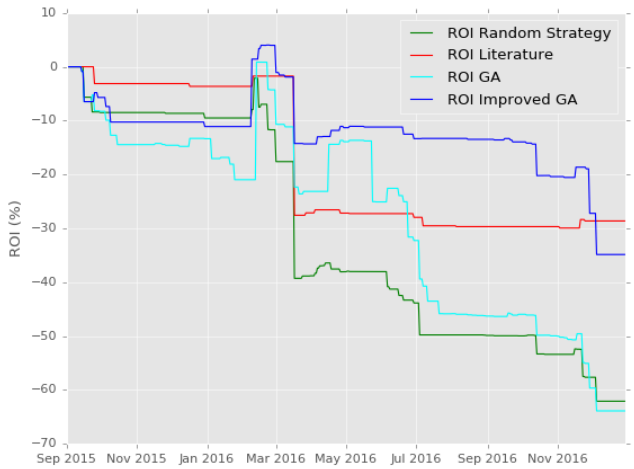


*Figure IV-4:Evolution of the ROI for the ALCA example.*

Figure IV-5 presents a part of the period in analysis, showing the buy and sell points defined by the Improved GA strategy. Here 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.

This period can partially explain why the strategy does not perform well in this company, since the spreads that compose the dataset are either falling or constant. A possible solution could be to sell protection (instead of buying it) when the spreads are high, which would make the contract much more valuable after the spreads fall, what is comparable to adopt a short position on the stock market.

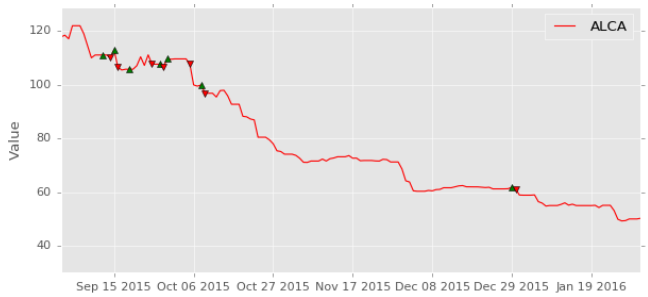


*Figure IV-5: Buy and sell points of the Improved GA strategy to ALCA (Sep 2015 – Jan 2016)*

### 4) *BKT – BlackRock Income Trust Inc.*

Aiming to describe a most frequent situation in the CDS market, this example will take on the company that achieved the median ROI during Case Study 1, more specifically BlackRock Income Trust. 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. Table IV-9 contains the obtained classification parameters for all the strategies applied to BKT, during the test period.

*Table IV-9: Classification Parameters for BKT example.*

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

Table IV-9 clearly states that both strategies based on evolutionary techniques perform much better than the two other strategies, achieving a considerable higher ROI. Figure IV-6 shows the evolution of the ROI for BKT during the period in analysis.

As observed in Figure IV-6, 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. However, when the spread falls abruptly as in the case of August 2016, they both experience higher losses. Figure IV-7 represents the buy and sell points of the first months of the test period.

This period reflects well what happens during the entire test period, with rises, falls and periods of more stability. Like in the previous examples, the strategy can perform well when the spreads rise, but not so well when they fall. The implementation of a short strategy could solve this issue. It is also important to notice that the spreads of this example are far from the ones of the CHK, which could be 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

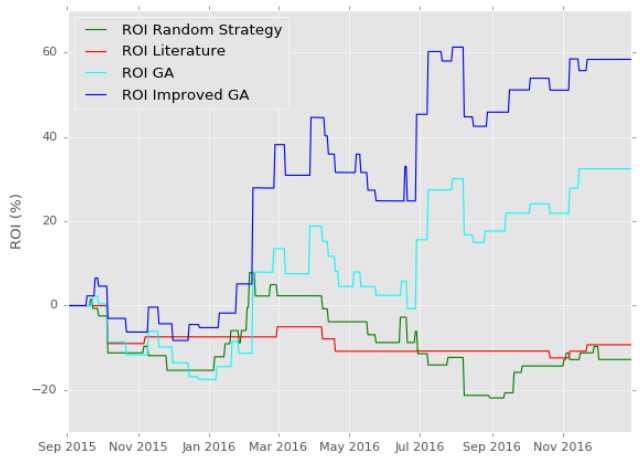


*Figure IV-6: Evolution of the Return on the Investment for the BKT example.*

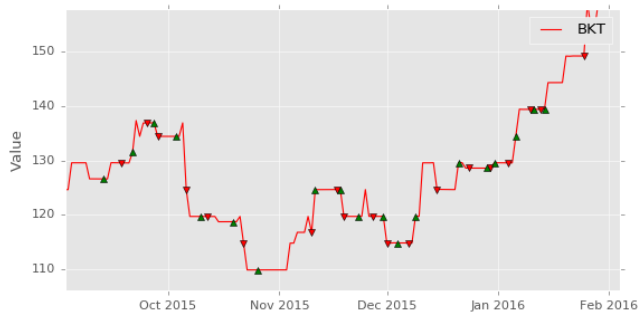


*Figure IV-7: Buy and sell points resulting of the application of the Improved GA strategy to BKT (Sep 2015 - Jan 2016).*

## V. 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 considering 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 profitable, but also 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.

## References


[1] N. Terzi and K. Uluçay, "The role of credit default swaps on financial market stability", *Procedia - Social and Behavioral Sciences*, Vol. 24, 2011, pp. 983–990.

[2] P. Augustin, "Credit Default Swaps: A Survey", *Foundations and Trends® in Finance*, Vol. 9, No. 1–2, 2014, pp. 1–196.

[3] L. Norden, "Information in CDS spreads", *Journal of Banking and Finance*, Vol. 75, 2017, pp. 118–135

[4] J.J. Murphy.

[5] L.B. Booker, D.E. Goldberg, and J.H. Holland, "Classifier systems and genetic algorithms", *Artificial intelligence*, Vol. 40, No. 1, 1989, pp. 235–282.

[6] K.F. Man, K.S. Tang, and S. Kwong, "Genetic algorithms: Concepts and applications", *IEEE Transactions on Industrial Electronics*, Vol. 43, No. 5, 1996, pp. 519–534.

[7] J. Hull, M. Predescu, and A. White, "The relationship between credit default swap spreads, bond yields, and credit rating announcements", *Journal of Banking and Finance*, Vol. 28, No. 11, 2004, pp. 2789–2811.

[8] 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", *Journal of Alternative Investments*, Vol. 11, No. 1, 2008, pp. 43–61.

[9] D. Avino and O. Nneji, "Are CDS spreads predictable? An analysis of linear and non-linear forecasting models", *International Review of Financial Analysis*, Vol. 34, 2014, pp. 262–274

[10] J. Holland, "Genetic Algorithms", *Scientific American*, 1989.

[11] P. Fernández-Blanco, D.J. Bodas-Sagi, F.J. Soltero, and J.I. Hidalgo, "Technical market indicators optimization using EA", *Proceedings of the 2008 GECCO conference companion on Genetic and evolutionary computation - GECCO '08*, 2008, pp. 1851

[12] A. Gorgulho, R. Neves, and N. Horta, "Applying a GA kernel on optimizing technical analysis rules for stock picking and portfolio composition", *Expert Systems with Applications*, Vol. 38, No. 11, 2011, pp. 14072–14085.

[13] R. Choudhry and K. Garg, "A Hybrid Machine Learning System for Stock Market Forecasting", *World Academy of Science, Engineering and Technology*, Vol. 2, No. 15, 2008, pp. 315–318.

[14] Y. Li and H. Xu.

[15] H.G. Cobb and J.J. Grefenstette, "Genetic algorithms for tracking changing environments", *Proceedings of the 5th International Conference on Genetic Algorithms*, No. 1992, 1993, pp. 523–530