Innovative Options Data Signal for Trading with Technical Indicators and VIX optimized by a GA

José Bernardo Lira Gomes de Almeida Belo

Thesis to obtain the Master of Science Degree in

**Electrical and Computer Engineering**

Supervisor: Prof. Rui Fuentecilla Maia Ferreira Neves

**Examination Committee**

Chairperson: Prof. Teresa Maria Sá Ferreira Vazão Vasques

Supervisor: Prof. Rui Fuentecilla Maia Ferreira Neves

Members of the Committee: Prof. João Filipe Pereira Fernandes

**June 2019**
Declaration:

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.
Abstract

In this thesis it is proposed the use of options data, organized in an innovative form, in a mechanism made by an investment simulator led by technical and VIX indicators using trailing stoplosses and optimized by a genetic algorithm with the objective of discover a rule that can be applied in the options market, then some tests were made in the SP500 Index options market for results comparing and analyzing. Daily options data from 2011 to 2017 on the SP500 index were used for train with a rolling window and a return of investment of 325 % was achieved in test in a time period of 1200 days.

Keywords: Data Processing, Genetic Algorithm, Options Market, Technical Indicators, Trailing Stoplosses, VIX
Resumo

Nesta tese é proposto o uso de dados relativos a opções, organizados de um modo inovador, num mecânismo criado por um simulador de investimento, conduzido por indicadores técnicos e pelo indicador VIX com trailing stoplosses e optimizado por um algoritmo genético com o objectivo de encontrar uma regra que seja possível aplicar no mercado de opções com a finalidade de obter lucros e, de seguida, são realizados alguns testes no ‘SP500 Index options market’ de modo a comparar e analisar resultados. Os dados usados correspondem aos valores diárias das opções do SP500 durante o período de 2011 até 2017 e são usados para treinar um sistema de janela deslizante conseguindo atingir um resultado de 325% de retorno num período temporal de 1200 dias.

Palavras Chave: Algoritmo Genético, Indicadores Técnicos, Mercado de Opções, Processamento de Data, Trailing Stoplosses, VIX
Acknowledgements

It is an incredible feeling to live this enthusiastic moment that is writing this message. Studying in Técnico was an incredible part of my life where it was given to me a chance of improving all my character in every ways. Técnico obviously gave me a lot of work and challenging experiences but besides that, it was a path that made me grow in several life dimensions and brought so much good things. I’m sure that I developed enough skills both professional and personal. It taught me to be resilient, to face challenges and find solutions to address them, to work in group, to listen and give attention to different points of view and recognize that results and objectives are only achieved with hard work. I’m so thankful for this experience that made me grow so much, for the professors for all they taught me and their willingness to help over the years. I must also thank to my family who always supported me and accompanied me throughout the all course, to my colleagues and friends that shared with me so many good moments, achievements and challenges. To Catarina that was always a solid support and shared with me every step of this way, thank you.

For all of that, it is with a great feeling and proud that I write this message of acknowledgement to all who were present in my journey.

Sincerely,

José Bernardo Belo
# Contents

Abstract iii

Resumo v

Acknowledgements vii

List of Figures xii

List of Tables xv

Acronym List xvi

Software List xvii

1 Introduction 1
   1.1 Work’s Purpose ........................................ 2
   1.2 Document Structure .................................... 2
   1.3 Contributions ........................................... 2

2 Background and State-of-the-Art 4
   2.1 Overview .................................................. 5
   2.2 Background ............................................... 5
      2.2.1 Options ............................................... 5
         2.2.1.1 American and European Options .................... 5
         2.2.1.2 Call and Put Options ............................ 6
         2.2.1.3 In the Money and Out of the Money options ........... 6
         2.2.1.4 Auction System ................................ 6
         2.2.1.5 Some Examples ................................ 7
         2.2.1.6 Option Root .................................. 7
      2.2.2 Black-Scholes Pricing of Options .................. 7
         2.2.2.1 Options Pricing Variables ........................ 7
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.2.2 The Model</td>
<td>8</td>
</tr>
<tr>
<td>2.2.2.3 The Greeks</td>
<td>9</td>
</tr>
<tr>
<td>2.2.3 Volatility and VIX</td>
<td>10</td>
</tr>
<tr>
<td>2.2.4 Technical Indicators</td>
<td>11</td>
</tr>
<tr>
<td>2.2.5 Genetic Algorithm</td>
<td>12</td>
</tr>
<tr>
<td>2.2.5.1 Initialization</td>
<td>13</td>
</tr>
<tr>
<td>2.2.5.2 Fitness Function and Selection</td>
<td>14</td>
</tr>
<tr>
<td>2.2.5.3 Crossover</td>
<td>15</td>
</tr>
<tr>
<td>2.2.5.4 Mutation</td>
<td>16</td>
</tr>
<tr>
<td>2.3 State-of-the-Art</td>
<td>16</td>
</tr>
<tr>
<td>2.3.1 Option Market Research</td>
<td>16</td>
</tr>
<tr>
<td>2.3.2 Algorithms and Technical Indicators Research</td>
<td>18</td>
</tr>
<tr>
<td>2.3.3 Genetic Algorithms and Technical Indicators Research</td>
<td>19</td>
</tr>
<tr>
<td>2.3.4 Stop Loss Research</td>
<td>20</td>
</tr>
<tr>
<td>2.4 Chapter Conclusion</td>
<td>22</td>
</tr>
<tr>
<td>3 Proposed Architecture</td>
<td>24</td>
</tr>
<tr>
<td>3.1 Overview</td>
<td>25</td>
</tr>
<tr>
<td>3.2 Structure Overview</td>
<td>25</td>
</tr>
<tr>
<td>3.3 Data Layer</td>
<td>26</td>
</tr>
<tr>
<td>3.3.1 Data Base</td>
<td>26</td>
</tr>
<tr>
<td>3.3.2 Options Plot</td>
<td>26</td>
</tr>
<tr>
<td>3.3.3 Window Plot</td>
<td>27</td>
</tr>
<tr>
<td>3.4 Indicators Layer</td>
<td>32</td>
</tr>
<tr>
<td>3.4.1 Technical Indicators</td>
<td>32</td>
</tr>
<tr>
<td>3.4.1.1 Relative Strength Index</td>
<td>33</td>
</tr>
<tr>
<td>3.4.1.2 Rate of Change</td>
<td>34</td>
</tr>
<tr>
<td>3.4.1.3 Bollinger Bands</td>
<td>35</td>
</tr>
<tr>
<td>3.4.1.4 Moving Average Convergence Divergence</td>
<td>36</td>
</tr>
<tr>
<td>3.4.2 Economic Indicators</td>
<td>37</td>
</tr>
<tr>
<td>3.4.2.1 Volatility Index</td>
<td>37</td>
</tr>
<tr>
<td>3.5 Investment Simulator Layer</td>
<td>38</td>
</tr>
<tr>
<td>3.5.1 Bear and Bull Spreads</td>
<td>39</td>
</tr>
<tr>
<td>3.5.2 Orders</td>
<td>40</td>
</tr>
<tr>
<td>3.5.3 Trailing Stop Losses</td>
<td>40</td>
</tr>
<tr>
<td>3.6 Optimization Layer</td>
<td>41</td>
</tr>
</tbody>
</table>
List of Figures

2.1 Option Price ................................................................. 8
2.2 The Greeks Map ........................................................... 9
2.3 VIX Indicator with SP500 Stock Price ............................... 11
2.4 Genetic Algorithm Flowchart ........................................... 13
2.5 Genetic Algorithm Initialization ....................................... 14
2.6 Elitism Selection Example ............................................... 15
2.7 Genetic Algorithm Uniform Crossover ............................... 15
2.8 Genetic Algorithm Mutation ............................................ 16

3.1 System Structure .......................................................... 26
3.2 Windows with the chosen, cropped and normalized put options (A) .... 29
3.3 Windows with the chosen and cropped put options (B) ............... 30
3.4 Underlying Asset Value (SP500) ....................................... 30
3.5 Strike Bands (C) ............................................................ 30
3.6 Data Layer ................................................................. 31
3.7 Selection of an option to a window .................................... 32
3.8 RSI of VIX with put marks with thresholds 30 and 70 ............... 34
3.9 ROC of VIX with put type marks and derivative parameter 3 .......... 35
3.10 Bollinger Bands of VIX with put type marks ......................... 36
3.11 MACD of VIX with put type marks ................................... 37
3.12 Investment Simulator Layer Fluxogram ............................... 38
3.13 Bull Put Spread .......................................................... 39
3.14 Bear Put Spread .......................................................... 40
3.15 Chromosome ............................................................ 42

4.1 Data split scheme for the rolling window ............................ 46
4.2 ROI (%) for the two Strategies ......................................... 50
4.3 Opened and Closed positions along test time ......................... 51
4.4 VIX with BB and VIX Signal in 2015 ........................................ 51
4.5 ROI (%) for the two Strategies .............................................. 53
4.6 VIX with BB and VIX Signal in 2015 ........................................ 54
4.7 Opened and Closed positions along test time .......................... 54
4.8 ROI (%) along the 3rd case study ........................................... 55
4.9 Positions opened and closed along the 3rd case study .................. 57
4.10 VIX the 3rd case study ...................................................... 57
List of Tables

2.1 Call Data Examples .................................................. 7
2.2 Put Data Examples .................................................. 7
2.3 Technical Indicator Types and Examples ............................ 11
2.4 Main Scientific Papers used as inspiration for this work ............ 23
3.1 Option Values through time .......................................... 27
3.2 RSI Marks Rules - Put Contracts .................................. 33
3.3 Rate of Change Marks Rules - Put Contracts ....................... 34
3.4 Bollinger Bands Marks Rules - Put Contracts ...................... 36
3.5 Moving Average Convergence Divergence Marks Rules - Put Contracts ............... 37
3.6 GA chromosome values range ........................................ 42
4.1 Data train and test division dates ..................................... 46
4.2 Strategies characterization ........................................... 48
4.3 Data Layer First Case Study Parameters .......................... 49
4.4 Evaluation Metrics for the First Study Case ......................... 50
4.5 Best Chromosome for each rolling window for the best improved GA result ............... 50
4.6 Data Layer Case Study Parameters .................................. 52
4.7 Evaluation Metrics for Study Case .................................. 53
4.8 Best Chromosome for each rolling window for the 2nd Case Study ....................... 53
4.9 Evaluation Metrics for the 3rd Study Case .......................... 55
4.10 Best Chromosome for each rolling window for the third case study ....................... 56
Acronym List

ANN - Artificial Neural Networks
BB - Bollinger Bands
CBOE - Chicago Board Options Exchange
DC - Donchian Channels
EMA - Exponential Moving Average
GA - Genetic Algorithm
HMA - Hull Moving Average
HOF - Hall of Fame
ITM - In the money
MACD - Moving Average Convergence Divergence
OBV - On balance Volume
OTM - Out of the money
PVT - Price Volume Trend
ROC - Rate of Change
ROCP - Rate of Change Percentage
RSI - Relative Strength Index
ROI - Return on Investment
SMA - Simple Moving Average
SVM - Support Vector Machine
TSI - True Strength Index
UAP - Underlying Asset Price
VIX - Volatility Index
Software List

Overleaf - Platform for writing projects
Jupyter Notebook - Platform to Python development
Plotly - Platform to plot graphics
Draw.io - Platform to make figures and schemes
Chapter 1

Introduction
1.1 Work’s Purpose

Nowadays it is very important to have financial knowledge. Knowing when and how to invest is an art that requires many years of experience and knowledge but it can be for everyone. Whether it’s an engineer, a doctor or a postman, knowing how to invest is a quality that should be studied by everyone and, since every worker gets paid in the country we are, knowing how to reinvest can make the incomes of everyone much greater. If all the Portuguese population knew how to invest and succeed on it, certainly the country would get richer and get better conditions for all.

The main objective of this thesis is to develop a model that studies the options market and their patterns and that discovers a firm and stable rule, which regularly settles positive transactions and gives returns in the future.

For the objective purposed it will be used data from the Chicago Board Options Exchange from 2011 to 2017 of the option contracts and the volatility index. The model will be trained and tested for results analysis with a rolling window that will roll on the time period described.

1.2 Document Structure

This document is organized in five main chapters described below:

- Introduction - in this chapter the works purpose is explained along with the document structure.
- Background and State-of-the-Art - in this chapter, it is made a view on the main themes that this work approaches and then all the research made is scrutinized and resumed.
- Proposed Architecture - in this chapter, it is projected the architecture for the model developed and all the characteristics that were built.
- System Evaluation - along this chapter are exposed the study cases developed for the architecture testing, while a set of comments is made on the results achieved.
- Conclusion and Future Work - In the last chapter it is made a conclusion on the work and improvements that might be done in future works are suggested.

1.3 Contributions

Bellow are the main contributions made along this thesis:

- Use of option spreads to reduce risk in the short term but profiting in a long term.
- Creation of an Innovative signal made of different options with a self designed data organizing layer.
- Using VIX with Bollinger Bands to avoid major losses on the fund and prevent big draw downs.
• Use of an adaptive algorithm (GA) along with technical indicators (BB, RSI) so that parameters get adapted to market dynamics.

• Use of the Hall of Fame on the GA architecture.
Chapter 2

Background and State-of-the-Art
2.1 Overview

In this section, firstly, it is introduced a background on the themes that will be studied along this thesis. It is presented a view on the Options Market and its main concepts along with a brief explanation of the Technical Indicators analysis and the Genetic Algorithm structure. In a second moment, it is shown and explored all the research made along this work being presented the past scientific papers and investigations which supported this project.

2.2 Background

'The financial markets are a zero sum game where every dollar won by one investor is lost by another. Knowledge and trading tools are the differentiating factors that determine whether an investor lands on the winning or losing side' [1].

2.2.1 Options

An option is a financial derivative that represents a legal contract sold by one party (writer) to another party (holder). An option is a contract that gives the holder the right, but not the obligation, to buy or sell an underlying asset from or to the writer at a specified price for a specified time period [2, 3, 4].

The specified price, known as strike price or execute price, is the price which the holder can exercise the contract. The specified time period is given by the expiration date, when the contract ends, and it begins at the moment the holder signs it. Every contract is sold by a premium price which varies with the parameters referenced above: strike price, expiration date and underlying asset behaviours. This premium price is an immediate negative cash flow for the holder who believes that can make a profit if the underlying asset price rises or decreases above or under the exercise price by an amount exceeding the premium price. When the expiration date is reached the contract loses all its effects and the holder loses his right to execute it losing the premium price payed for the option.

An option holder can also sell an already bought option contract too another entity if maturity is not reached yet and so an option can change its holder throughout its lifetime.

2.2.1.1 American and European Options

There are two types of option contracts [2, 3] based on the expiration date and period of exercising, European which can only be executed at the expiration date and American which can be executed at any time between the signing date and the expiration date. Since American options can be executed between a time period and not only on a specified date, their price is superior than European options. As said in [2] most of the options traded nowadays are American options and so on along this document, a special attention to the American Options will be given.
2.2.1.2 Call and Put Options

An option is called a ‘Call’ when the holder as the right, but not the obligation, to buy an underlying asset from the writer at a specified price and time period and it is called a ‘Put’ when the holder as the right, but not the obligation, to sell an underlying asset to the writer at the specified price and time period [2, 3, 4]. On one hand, a Call is a bet on the underlying asset price increase and the holder can make profit by buying an asset cheaper than the real price since the contract is executable at the strike price. If the asset price doesn’t get lower than the strike price during the contract the holder will only lose the premium price payed for the signing. Therefore a Call option has a limited possible lose which is the premium price and an infinite possible profit for the holder since assets can theoretically grow infinitely. On the other hand, a Put is a bet on the underlying asset price decrease and the holder can make profit by selling an asset more expensive than the real price since the contract is executable at the strike price. A Put option can be seen as an insurance. If the asset price doesn’t go below the strike price during the contract the holder won’t execute the contract and the premium price will be lost. Therefore a Put option has a limited loss which is the premium price and a finite possible profit since asset prices are limited at zero.

2.2.1.3 In the Money and Out of the Money options

Besides executing an option contract during the contract period an investor can also sell an already bought contract in the middle of its lifetime by a bigger premium price than the one paid to the writer because of the variations on the assets price. In its lifetime an option can be in-the-money or out-of-the-money [4]. An option is in-the-money if, in a moment, the strike price is smaller than the asset price in case of a call and if the strike price is greater than the asset price in case of a put. So on, and summarizing an option is in or out of the money depending if its intrinsic value is, respectively, positive or negative.

\[
IntrinsicValue(Call) = UnderlyingAssetPrice - Strike
\]

\[
IntrinsicValue(Put) = Strike - UnderlyingAssetPrice
\]

2.2.1.4 Auction System

To buy or sell an option it is essential to understand how the system works and how these contracts are traded. The system used for options trading is an auction based system. The ask price is fixed by the sellers and is the minimum value in which some seller is willing to sell an option [5]. The bid price is fixed by the buyers and is the maximum value offered by the community of investors to buy a determined option. Both ask and bid prices will be oscillating because of the market members will and competition. At last, we have the last price which is the last option price accepted by a seller to write a contract to a
2.2.1.5 Some Examples

On tables 2.1 and 2.2 it is possible to see some examples of real traded option contracts:

<table>
<thead>
<tr>
<th>Call</th>
<th>UAP</th>
<th>Company</th>
<th>Traded Date</th>
<th>Expiration Date</th>
<th>Strike</th>
<th>Last</th>
<th>Bid</th>
<th>Ask</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>604.35</td>
<td>Google</td>
<td>03/01/2011</td>
<td>22/01/2011</td>
<td>230</td>
<td>382</td>
<td>372.2</td>
<td>375.4</td>
</tr>
<tr>
<td></td>
<td>560.54</td>
<td>Apple</td>
<td>12/12/2013</td>
<td>13/12/2013</td>
<td>375</td>
<td>0</td>
<td>184.3</td>
<td>186.9</td>
</tr>
<tr>
<td></td>
<td>51.83</td>
<td>Facebook</td>
<td>12/12/2013</td>
<td>03/01/2014</td>
<td>55</td>
<td>0.68</td>
<td>0.69</td>
<td>0.72</td>
</tr>
</tbody>
</table>

Table 2.1: Call Data Examples

<table>
<thead>
<tr>
<th>Put</th>
<th>UAP</th>
<th>Company</th>
<th>Traded Date</th>
<th>Expiration Date</th>
<th>Strike</th>
<th>Last</th>
<th>Bid</th>
<th>Ask</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>923</td>
<td>Google</td>
<td>12/07/2013</td>
<td>17/08/2013</td>
<td>1055</td>
<td>0</td>
<td>132.4</td>
<td>135.6</td>
</tr>
<tr>
<td></td>
<td>426.51</td>
<td>Apple</td>
<td>12/07/2013</td>
<td>12/07/2013</td>
<td>485</td>
<td>66.85</td>
<td>57.6</td>
<td>60.4</td>
</tr>
<tr>
<td></td>
<td>51.83</td>
<td>Facebook</td>
<td>12/12/2013</td>
<td>01/15/2016</td>
<td>20</td>
<td>0.81</td>
<td>0.65</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Table 2.2: Put Data Examples

For example, in the first entry of table 2.1, the Google Call option with strike 230, in 03/01/2011 was last traded with the premium price 382 and by the end of the day sellers were willing to sell a new option at the price 375.4 and buyers were willing to buy at 372.2.

2.2.1.6 Option Root

The Option Root is a code that fully represents an option contract and it consists of merging the asset company or type, the expiration date, the type of the option (call or put) and eight digits for the strike price [3]. For example, in the first line of the table 2.2 we would get GOOG130817P01055000 as option root known that Google symbol is 'GOOG'.

2.2.2 Black-Scholes Pricing of Options

The Black-Scholes formula (also called Black-Scholes-Merton) was the first widely used model for option pricing [7, 8, 9]. It’s used to calculate the theoretical value of an option and, invented in the early 70’s by Myron Scholes and Robert Merton, this formula was worth the 1997 Economy Nobel Prize [6], being listed Fischer Black as a contributor, though he was ineligible for the prize as he had passed away before it was awarded.

2.2.2.1 Options Pricing Variables

Black, Scholes and Merton made a formula which was capable of giving a premium price to an option contract having in consideration the variables which characterize it. In the figure 2.1 its possible to see the variables that directly change the value of an option.
Figure 2.1: Option Price

Each single variable contributes for an increasing or decreasing in the final premium option value. This model decides the amount of money a buyer must pay for an option in the present date.

### 2.2.2.2 The Model

The Model for call and put pricing consists, as described by [2], on the equations below:

\[ C = S_0 N(d_1) - X e^{-rt} N(d_2) \]  
\[ P = X e^{-rt} N(-d_2) - S_0 N(-d_1) \]

(2.3)

(2.4)

With,

\[ d_1 = \frac{ln\left(\frac{S_0}{X}\right) + (r + \frac{\sigma^2}{2})T}{\sigma\sqrt{T}} \]  
\[ d_2 = d_1 - \sigma\sqrt{T} \]

(2.5)

(2.6)

Listed below are the parameters used in the formulas:

- \( S_0 \) - Actual Asset Price
- \( X \) - Contract Strike Price
- \( T \) - Time to contract expiration date
- \( r \) - Risk-Free Interest Rate
- \( \sigma \) - Volatility

Where \( C \) and \( P \) are the premium prices for a call and a put option and it is calculated by subtracting two parcels. In formula (2.3) the first parcel is the actual asset price multiplied by the cumulative standard normal distribution of \( d_1 \) and the second parcel is the Strike Price multiplied by the time discount made by the risk-free interest rate and the cumulative standard normal distribution of \( d_2 \). Summarizing, the
first parcel of the equation (2.3) is the benefit that an holder will get from the asset since it’s used the actual price while the second parcel corresponds to the writers loss, as the exercising is made at the strike price, discounted in time. Analogously the same interpretation is applied in (2.4). In formulas (2.5) and (2.6) the parameters $d_1$ and $d_2$ are calculated with the parameters shown in figure 2.1 (dividends are not used) which affect the options premium value calculated.

2.2.2.3 The Greeks

The Greeks are simply sensitivities to options risk characteristics [3]. Each one of the Greeks variables depend on different options parameters and in figure 2.2 it is displayed each Greek and which parameter sensitivity reflects:

- The Delta parameter is the degree of change an option premium price will move based on a 1 dollar change on the UAP. Delta values fluctuate between 0 and 1 for calls and between -1 and 0 for puts and, for example, if an option has a delta of 1 it means that its premium price will oscillate in the same rate of change as the UAP. An ITM call will approach a delta of 1 as it gets closer to expiry and, on the contrary, an OTM call will approach a delta of 0 as it gets closer to expiry. Analogously an ITM put will approach a delta of -1 as it gets closer to expiry and an OTM put will approach a delta of 0 as it gets close to expiry.

![Figure 2.2: The Greeks Map](image)
• The Gamma parameter is the rate of change the premium price will move based on the next increments of 1 dollar after the first one in the UAP. Gamma is the 'acceleration' of Delta throughout successive increments on the UAP. Option contracts with highest Gammas are more responsive to changes in the UAP.

• The Vega parameter is the amount of money the premium price will move based on a 1 percent move in the implied volatility of the UAP. Vega is normally the same for Calls and Puts.

• The Theta or time decay is always negative and is the enemy number one of an investor because it is the amount of money the premium price will decrease every single day. Since options are wasting assets their premium price decays over the time and it is expected that options with more time to expiration have less time decay than others with less days to expiration.

• The Rho parameter is the amount of money the premium price will move based on a 1 percent move in the interest rate.

2.2.3 Volatility and VIX

Volatility refers to the amount of uncertainty related to the size of changes in an asset value. A higher volatility means that an asset value is spread out over a larger range of values. This means that the price of the asset has changed more over a time period in either direction. A lower volatility means that an asset value does not fluctuate so much and so it is steadier. Volatility on the assets directly impact the options premium prices because options on assets with higher volatility are much more risky to the writers than assets with lower volatility.

The VIX® Index was developed by CBOE [10] and is a ‘benchmark index to measure the market’s expectation of future volatility. The VIX Index is based on options of the SP500® Index, considered the leading indicator of the broad U.S. stock market’.

In figure 2.3, it is represented the VIX indicator along the years of 2015 and 2016 with the representation of the SP500 stock market price.
2.2.4 Technical Indicators

In finance, technical analysis is a security analysis discipline for forecasting the direction of prices through the study of past market data, primarily price and volume [11]. Technicians say that market prices reflect all relevant information, so the analysis is made by looking at the history of a security’s trading pattern rather than external drivers such as economic, fundamental and news events. Price action also tends to repeat itself because investors collectively tend toward patterned behaviors so technicians focus on identifiable trends and conditions. In table 2.3 it is shown some types of indicators and examples.

<table>
<thead>
<tr>
<th>Trend Indicators</th>
<th>MA, MACD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Momentum Indicators</td>
<td>RSI, ROC, ROCP</td>
</tr>
<tr>
<td>Volatility Indicators</td>
<td>BB</td>
</tr>
<tr>
<td>Volume Indicators</td>
<td>OBV</td>
</tr>
</tbody>
</table>

Table 2.3: Technical Indicator Types and Examples

While the trend indicators measure the strength and direction of a trend, the momentum indicators measure the velocity of price changes as opposed to the actual price levels themselves [12]. So on, the Volatility Indicators measure the rate of price movements and can help identifying direction changing moments. To finish, the Volume indicators measure, with volume based data, and confirm the strength of new trends. There are a lot of different technical indicators and so on it is important to choose and set them carefully. A full description on each of the technical indicators used is presented in the following sections.
2.2.5 Genetic Algorithm

It is needed to go back to the 70’s where for the first time John Holland [14], Professor of electrical engineering and computer science at the University of Michigan, developed and built this new concept that merges Darwin’s ideas [13] into process optimization. In his book [14], John Holland describes adaptation, which retracts the biological response whereby organisms evolve by rearranging genetic material to survive in environments confronting them, and overpasses this analogy to a practical mechanism. Later on, David A Coley [16], in his book, defines the Genetic Algorithms as ‘...numerical optimization algorithms inspired by both natural selection and natural genetics’ recalling to Darwin’s Theory [13] and, as Melanie Mitchel [15], describes GAs in four main blocks:

1. A population of members, each one with a solution to the problem;

2. A way of calculating how good or bad each member of a population is and a selection method to develop future populations;

3. A method for mixing fragments of the selected members inside a population to form a new and better one;

4. A mutation operator that gives diversity in members inside new populations.

So on, the GAs begin by Initializing a population with n members, also called chromosomes. Each chromosome can be seen as a vector of m genes or features that represents a solution in a vast solution space. Then, this first generation is analyzed in a Fitness Function where it is created a rank in how well each chromosome fits the problem. After ranking a generation, a Selection method is applied to find the best solutions inside the population and those who will be the parents of the following generation. So on, these parents are used in the Crossover Stage where it is created a new generation of chromosomes, a new population, made of those who were considered the best by the Fitness Function in the last analyzed generation. The Crossover gathers two or more parents successively and randomly mixes them forming offsprings that will be joined by new random chromosomes and used on the next population. A Mutation operator is then applied in the new elements, which consists of changing some of the features inside the offsprings so that it is guaranteed diversity along the GA and avoiding that it falls to a local maximum. So after ranking, selecting the best chromosomes, crossover the parents and mutate the offsprings a new generation is ready for ranking again. This process is repeated until a chromosome fits so well that reaches the stop criteria.

The method is a general one, capable of being applied to an extremely wide range of big featured non-linear problems and both Melanie Mitchel [15] and David A Coley [16] give a huge amount of examples on processes which can be optimized by GAs such as image processing, medicine, facial recognition and
training and designing artificial intelligence systems such as artificial neural networks. In figure 2.4 it is
the flowchart of a Genetic Algorithm.

![Genetic Algorithm Flowchart](image)

Figure 2.4: Genetic Algorithm Flowchart

Now, an exploration on each block that makes a GA will be done.

### 2.2.5.1 Initialization

This first step in a GA is the creation of the first population from where the algorithm will grow. This population is made of n members and each member, also called as chromosome, is a possible solution to the process being optimized. Each chromosome is a vector of genes or features with length m, equal to the solutions space.

It is mandatory that each chromosome has identical characteristics so that all of them can be logical applied to the Fitness Function and to the process itself. In figure 2.5 it is possible to see the initialization of a population with the characteristics described above.
2.2.5.2 Fitness Function and Selection

The Fitness Function is a block that gets surpassed by every chromosome and it will attribute a mark (a fitness measure) to each one making a rank of evaluations of all the n members inside a population. This Fitness Function is one thing to have in mind since it is this function that, inside the GA, measures if a chromosome is worse or better for the purpose. If a chromosome achieves what it is called the stop criterion, that is, the desired optimized result, the GA will end and show the chromosome with the best fitness measure found. In formulas 2.7 and 2.8 it is defined the fitness function.

\[ f(x_1, x_2, ..., x_m) = y \]  
\[ \mathbb{R}^m \Rightarrow \mathbb{R} \]  

After grading each chromosome and if the stop criteria is not achieved the GA continues to find better solutions and goes for the selection of those chromosomes who will enter on the next population. In this process it is important to see that good chromosomes are needed so that the GA converges to an optimal solution and that bad chromosomes can be useful for adding variety in the new population. So it is important to achieve a balance between chosen chromosomes.

In Talbi’s analysis about GAs [17] there are many selection mechanisms. One of them, also studied by David A Coley [16], is the Roulette Wheel Selection in which each member will have a probability of being selected to the next population proportional to its mark, so a best ranked solution is more likely to be chosen than a less ranked one. Another selection method studied by Talbi is Elitism. This method, consists in selecting the best individuals out of each population to formulate the next one. As Talbi says, ‘This approach leads to a faster convergence and a premature convergence could occur. Sometimes,
selecting bad individuals is necessary to avoid the sampling error problem' so a flexible and balanced selection method should be used. In figure 2.6 it is represented an elitism strategy based by selecting the first two most well classified solutions plus one random.

![Figure 2.6: Elitism Selection Example](image)

### 2.2.5.3 Crossover

The Crossover consists in getting two or more previously selected chromosomes called parents and joining them to create an offspring. Crossover happens with the chromosomes that were chosen by the Selection block with the purpose of fulfill and improve the next population. In his book, Talbi [17] shows many ways to realize the crossover. One way that is looked it is called as Uniform Crossover where each gene in the offspring is discovered from a lottery between the same gene but on the parents chromosomes. Below in figure 2.7 there is an example of uniform crossover where the offspring chromosome gets the second and fifth gene from the parent chromosome 1 and the first, the third and the forth from the parent chromosome 2.

![Figure 2.7: Genetic Algorithm Uniform Crossover](image)
2.2.5.4 Mutation

The Mutation operator described by Melanie Mitchel [15] 'can occur at each' gene 'position (...) with some probability, usually very small' and makes the algorithm variety increases avoiding that it gets stuck in a local maximum. In this block every gene in every offspring might be changed to a random new value so that the algorithm keeps looking for new and better chromosomes. In figure 2.8 there is an example of the mutation of a chromosome where only the forth gene was mutated.

![Mutation Diagram](image)

Figure 2.8: Genetic Algorithm Mutation

2.3 State-of-the-Art

After giving the background about the key points reviewed along this thesis, in this section it will be explored many scientific documents which were gathered and object of analyze during the research of this work. Firstly it is presented the documentation about Option Contracts and their role in the markets. In a second moment it is introduced the research made on the advances that machine learning brought to the financial and option markets including the use of technical indicators and stop-losses strategies. Finally it is explored the researched on the GAs and its improvements along time.

2.3.1 Option Market Research

The long history of the theory of option pricing began in 1900 when the French mathematician Louis Bachelier deduced an option pricing formula based on the assumption that stock prices follow a Brownian motion with zero drift [18].

Studies on options analysis have been made along the decades by the scientific community and step by step the options theories grew up and proved their values in the real world.

In 1964, Kruizeng [19] explored the combinations of different position types to create strategies with varying degrees of risk and return using for representation, a vector notation. Every position type, for example a call or a long, was described by a two-dimension vector in which the first entry represented...
the return in case of a unit price increasing and the second, the return in case of a unit price decreasing, on the asset. For example a short would be (-1, 1).

Following this, in 1969, Hans. R. Stoll [20], developed his work to show that, in theory, an arbitrage mechanism exists which ought to keep put and call prices in line with each other irrespective of the demands of buyers of options. Using Kruizeng notation [19], Stoll firstly settles that both call and put prices, in equilibrium, are separated by the interest rate and that if the call price increases an equal decrease in the put price happens than justifies why puts and calls are linked by a parity relation and with an empirical examination through a set of Companies concludes that a rise in the call price won’t imply that the expected value of the stock price change is greater than before but only that the probability distribution of price changes has widened. Later in [9], Black and Scholes confirmed Stoll theory.

After these studies, Merton [9] in 1973, proved that the only reason that an American put will sell for a premium over its European counterpart is that there is a positive probability of exercising prior to expiration. There was loads of works trying to formulate an option pricing formula like Sprenkle [21], Ayres [22], Boness [23], Samuelson [24] and Chen [25], but it was also in 1973 that, Black and Scholes [7], helped by the works of Merton [9], developed a formula for option pricing and realized that option buyers were paying and supporting prices that were consistently higher than those predicted by the formula. They also concluded that the market appeared to underestimate the effect of differences in variance rate on the value of an option since the difference between the price paid by option buyers and the value given by the formula was greater for options on low-risk stocks than for options on high-risk stocks.

Along the following years, many researchers have empirically investigated the relation between options and equity markets. Anthony (1988) [26] uses daily closing prices to conclude that options lead stocks. Later Stephan and Whaley (1990) [27] conclude the exactly opposite relation finding that stocks lead option prices. This discussion is also made by Manaster and Rendleman (1982) [28], Bhattacharya (1987) [29], Vijh (1988, 1990) [30], Conrad (1989) [31], DeTemple and Jorion (1990) [32], Damodoran and Lim (1991) [33], Sheikh and Ronn (1994) [34], Kumar, Sarin and Shastri (1995) [35] and however the evidence on market interrelationships is inconclusive as to which of the two markets reflects new information earlier.

In 1998, David Easley, Maureen O’Hara and P. S. Srinivas [36], present a “pooling” model where a mixed solution about the markets it is used. With an intraday option data this model chooses whether to invest inside the option market or in the stock market or even in both and it is concluded that in many cases the stocks lead the option market but surprisingly in some particular option volumes the option market leads the stock market. It is also settled that how volume is correlated with information and what implies for price movements is surely important.

Jun Pan [37] in 2006, presents strong evidence that option trading volume contains information about future stock prices. With his work, Jun settles that it takes several weeks for stock prices to adjust fully to
the information embedded in option volume and also refers that the predictability of options gets higher with the concentration of informed traders and the leverage of option contracts.

2.3.2 Algorithms and Technical Indicators Research

The crucial problem is not creating new jobs. The crucial problem is creating new jobs that humans perform better than algorithms [38]. Many were the works on algorithmic trading in the era of computation that we live in. With different implementations algorithms such as Neural Networks, Support Vector Machine, Genetic Algorithms, Decision Trees or even Hybrid Algorithms were studied and revealed a great potential for this application.

We go back to 1998 when Jingtao Yao, Chew Lim Tan and Hean-Lee Poh, in their work [39], compared the capabilities of a neural network agaisnt the past conventional ARIMA models for the Kuala Lumpur Composite Index forecasting. In their analysis data is processed by technical indicators (RSI, MA’s, Momentum, stochastic and MA’s of the stochastic) and set in the neural network for training. With a normalized mean squared error as performace metric their model outpermed the ARIMA model with a return of 26%.

In 2003, David Enke and C.H.Dagli exposed in their work [40], three types of neural networks joined with technical indicators for predicting future stocks behavior. In their publication indicators like Relative Strength Index, Money Flow Index, Moving Average, Stochastic Oscillator and Moving Average Convergence Divergence are used and using as evaluation metrics the return and the risk, it is settled that The overall results indicate that the proportion of correct predictions and the probability of stock trading guided by these neural networks are higher than those guided by their benchmarks (traditional technical indicators strategies). Also in 2003, L. J. Cao and Francis E. H. Tay [41] test, in their work, the capabilities of SVM and Adaptive SVM against back-propagation and radial basis function neural networks for financial time series forecasting. With the help of technical indicators such as Exponential Moving Averages and Relative Differences in Percentage and performance metrics like normalized mean squared error, mean absolute error, and directional symmetry, five data sets are studied and it is concluded that the SVM and the Adaptive SVM, with even less support vectors than SVM, outperform their back-propagation, weighted back-propagation and radial function basis neural networks strategies generally.

A few years later, in 2010, Binoy Nair, N. R. Sakhivel and V. P. Mohandas [42] developed on their work, to predict the next days trend, an hybrid system where a set of features, extracted from SENSEX historical data by technical indicators, were chosen by relevance and a decision tree and then presented to a rough set based system to induce rules from the extracted features. In their work volume based indicators such as RSI, PVT, OBV, price based indicators like MACD, Momentum, William’s and overlay indicators like BB and MA’s were used. This hybrid model, with 90.22% of accuracy, outperformed their neural network (77.66%), their Naive Bais based prediction system (72.36%) and their rough set based
trend prediction stand alone (88.18%).

To conclude, in 2017, Anton Aguilar Rivera and Manuel Valenzuela-Rendon [43] gathered and united an algorithm trading method based on evolutionary algorithms and portfolio theory to develop a mechanism which could make investment decisions which consider dynamic restrictions like transaction costs, portfolio unbalance, and inflation. Comparing some different test strategies a final return of 88.62% is achieved in the best case settling that the experiments showed this approach had better performance than buy-and-holds and single-period portfolios for the proposed metrics of performance.

2.3.3 Genetic Algorithms and Technical Indicators Research

In 2008, Rohit Choudhry and Kumkum Garg [44], developed a model in which a genetic algorithm and a support vector machine Algorithm together in a hybrid model outperformed a simpler SVM model. In the work Rohit and Kumkum use technical indicators (PVT, Stochastic, Disparity, ROC, Williams and Momentum) of not only the stock price in case study but also the correlated stock price companies as features for the algorithms and show that the GA-SVM algorithm outperforms the SVM stand alone in three different stocks, TCS with a hit ratio of 61.732% against 58.09%, Infosys with 60.285% against 56.748% and in Reliance with 59.534% against 55.643%.

Also in 2008, Hyunchul Ahn and Kyoung-jae Kim [45], presented in their work a new hybrid model for effective corporate bankruptcy prediction. To the best prediction model made for the purpose until their time, the Case-based reasoning (CBR) they merged a genetic algorithm to optimization of feature weighting and the instance selection, improving its qualities. With experimental results it is settled that this hybrid model can achieve greater results than the CBR alone and offer a new solution for bank agencies.

Some years later, in 2011, Gorgulho [46] developed a model merging a GA with a technical indicator analysis. In his work, with the use of indicators as EMA, HMA, DC, ROC, RSI, OBV, TSI and MACD and a 2003-2009 data set it is settled that this model results are promising since the approach clearly beats the remaining approaches during the recent market crash and a ROI of 62.95% was achieved by the best GA.

On the following year, Kyoung-jae Kim and Hyunchul Ahn [47], introduced their study in financial forecasting in which is presented a hybrid solution that consists on a neural network optimized by a genetic algorithm. With seven different models the authors test the capabilities of a genetic algorithm to maximize a neural network ability for prediction, by defining its architecture and weights to trade in the Koren Composite Stock Price Index. Also in this academic study technical indicators like Relative Strength Index, Stochastic Oscillators, Momentum, Price rate of Change and Disparity are applied. It is concluded that GA’s simultaneously optimizes multiple architectural factors and connection weights in ANN, and then enhances the prediction accuracy and the generalization ability of the classifier based on
Later, in 2013, Mariela Nogueira, Carlos Cotta and Antonio J. Fernandez-Leiva, exposed their work [48], with the objective of achieve wining strategies in the RobotWars (a two-player real time strategy (RTS) game developed in the University of Malaga for research purposes) game, in which the main goal is testing different approaches in order to implement the concept of HOF (Hall-of-Fame) as part of the self learning mechanism in competitive coevolutionary algorithms. In their thesis, six different types of hall of fame are implemented proving to be a feature that can be useful for an evolutionary algorithm. The implementation of the HOF with the best results is the diversity mechanism which consists of measuring the entropy of each solution in the HOF so those with the less entropy measure can be taken out without corrupting the quality of the entire HOF set. It is settled that the motivation of this proposal is to maintain certain diversity among the members of the HOF, and at the same time to reduce (or maintain an acceptable value for) the size of the memory. With this idea, we assume that the deleted individuals will not affect the quality of the found solutions.

Lastly, we go back to the year of 1993 where Helen G. Cobb and John J. Grefenstette [49], expose three new GA features that might improve the overall results of it. In this study, the first presented solution is a high mutation rate which provides good tracking performance for environments that change continuously through translation, but with this overall increase in mutation, the average performance deteriorates. The second feature consists in what it is called the Hypermutation, where an adaptive changing on the mutation happens whenever the time-averaged best performance of the population deteriorates [50] and is settled by Cobb and Grefenstette that for certain classes of environmental change, the Hypermutation GA adaptively introduces diversity when needed. To conclude, the last feature proposed by these academics is the introducing of The Random Immigrants inside each new population before the fitness evaluation but this approach increases the probability of losing information that may match small incremental changes in the environment, as shown by the relatively poor performance on the translating environments case.

2.3.4 Stop Loss Research

In 1998, Terrance Odean [51] tested the tendency of investors to hold losing investments too long and sell winning investments too soon, by analyzing trading records for 10,000 accounts at a large discount brokerage house. In his study Terrance settles that individual investors demonstrate a significant preference for selling winners and holding losers.

Some years later, in 2007, Kathryn M. Kaminski and Andrew W. Lo [52], using monthly returns data from January 1950 to December 2004, found that certain stop-loss rules could add 50 to 100 basis points per month to the buy-and-hold portfolio during stop-out periods. In their study it is presented the idea that using stop losses is playing safe and it is concluded that as difficult as it may be to accept,
for the millions of investors who lamented after the bursting of the Technology Bubble in 2000 that “if I only got out earlier, I wouldn’t have lost so much”, they may have been correct.

Following these work, in 2009, Adam Y.C. Lei and Huihua Lib [53], published a paper on stop losses and their role on the investment world. In their presentation it is used two different stop loss approaches. The first, the traditional stop loss with a fixed stop price and a second approach, the trailing stop loss with a stop price that adjusts upwards automatically with the security price but not downwards. In these study it is affirmed that the first method is able to reduce losses in some cases but not in others unlike the trailing stop loss that shows the effect of reducing investment risk rather than reducing investment losses.
2.4 Chapter Conclusion

In this chapter, in a first moment, it was introduced the theory on the themes that this thesis works on. Themes like option definitions and types, followed by options pricing and option market operation were deeply viewed. It was also exposed analysis on topics like volatility, market technical indicators and the structure and behavior of a genetic algorithm.

In a second moment, scientific documents and investigations were discussed and presented to the reader to share the inspirational scenario that was used to this thesis. Along this sub chapter, it is presented the research made and the approaches used by scientific community on themes like the option market correlation with the stock market, stop losses mechanisms, algorithms used for trading in markets with technical indicators and a primarily view on genetic algorithms studies was done.

In the table 2.4, it is exposed the main scientific documents that inspired this thesis.
<table>
<thead>
<tr>
<th>Paper</th>
<th>Year</th>
<th>Area</th>
<th>Financial Market</th>
<th>Methodology</th>
<th>Indicators</th>
<th>Performance</th>
<th>Dataset</th>
<th>Return</th>
</tr>
</thead>
<tbody>
<tr>
<td>39</td>
<td>1998</td>
<td>Stock</td>
<td>KLCI</td>
<td>NN</td>
<td>RSI, MA’s, Momentum, Stochastic</td>
<td>NMSE</td>
<td>3/1/1984-16/10/1991</td>
<td>26%</td>
</tr>
<tr>
<td>42</td>
<td>2010</td>
<td>Stock</td>
<td>Sensex</td>
<td>Decision Tree</td>
<td>RSI, PVT, MACD, BB and more</td>
<td>Accuracy</td>
<td>3/9/2003-7/3/2010</td>
<td>90.22%</td>
</tr>
<tr>
<td>44</td>
<td>2008</td>
<td>Futures</td>
<td>Chicago MM</td>
<td>GA-SVM</td>
<td>PVT, Stochastic, DISP, ROC, Momentum</td>
<td>NMSE</td>
<td>5 Contracts</td>
<td>61.732%</td>
</tr>
<tr>
<td>45</td>
<td>2008</td>
<td>Bankruptcy</td>
<td></td>
<td>GA-CBR</td>
<td>-</td>
<td>-</td>
<td>5 Contracts</td>
<td>-</td>
</tr>
<tr>
<td>46</td>
<td>2011</td>
<td>Stock</td>
<td>All stocks from DJI</td>
<td>GA</td>
<td>EMA, HMA, ROC, RSI, MACD, TSI, OBV</td>
<td>ROI</td>
<td>01/01/03-31/06/09</td>
<td>62.95%</td>
</tr>
<tr>
<td>47</td>
<td>2012</td>
<td>Stock</td>
<td>KCSI</td>
<td>GA-NN</td>
<td>RSI, Stochastic, Momentum, ROC, DISP</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>48</td>
<td>2013</td>
<td>Robot Wars</td>
<td></td>
<td>GA-HOF</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>43</td>
<td>2017</td>
<td>Stock</td>
<td>Mexican IPC</td>
<td>EA</td>
<td>-</td>
<td>ROI</td>
<td>2012-2014</td>
<td>88.62%</td>
</tr>
</tbody>
</table>

Table 2.4: Main Scientific Papers used as inspiration for this work
Chapter 3

Proposed Architecture
3.1 Overview

In this section, firstly, it is introduced the global idea of the proposed structure architecture that was explored along this work and together the summarizing fluxogram is presented and explained. Then, in a second moment, a full description is made on each main block that form the model presented. Each main block is introduced with its inputs, outputs and functions with a detailed explanation on how blocks link with each other.

The architecture described below was implemented and tested with python, since it is used for its simplicity and the vast number of libraries available.

3.2 Structure Overview

Along this work it is used an architecture based on four main blocks, the Data Layer, the Indicators Layer, the Investment Simulator Layer and the Optimization Layer. Each layer has its own mechanisms and together, the four layers, form an optimized trading system made to invest inside the options market. Below are presented the main layers with a small introduction to its characteristics and role inside the full process.

1. The Data Layer is the beginning of the full process and organizes the raw options data to be presented to the next layers. This process is personalized to options data and explored deeply in the Data Layer Section.

2. The Indicators Layer was based on Gorgulho’s [46] system that gathers technical indicators and a grading mechanism to evaluate (daily, weekly or monthly) the best moments to long or short the stock markets. This system was used, adapted for options markets, using options market data from the Data Layer instead of the stock markets data as Gorgulho did and improved with new features that are reviewed below.

3. The Investment Simulator Layer receives the grades from the Indicators Layer and the organized data from the Data Layer and simulates investments on the options market. Both the Investment Simulator Layer and the Indicators Layer will be part of a process that will be optimized by the last layer.

4. The Optimization layer is where the Genetic Algorithm will be applied. The genetic algorithm will try to maximize the results of our process experiencing and evaluating lots of parameters combinations that are inputs in both the Indicators Layer and Investment Simulator Layer. This optimization pretends to find a rule which is applied inside the options market and profitable for the future.
In the figure (3.1) it is possible to see the fluxogram that describes the proposed architecture. Following this quick introduction, each layer is explored and detailed.

![Figure 3.1: System Structure](image)

**3.3 Data Layer**

Options data can give relevant information about the future underlying asset behaviors as explored in the state-of-the-art and so on, the data layer is a major part of this work since it is used the options raw data to develop and create a plot of the contracts value along time which might give information about the underlying asset behaviors. Call and Put options can have different lifetimes and different strikes inside the same underlying asset and in this section we will explore how the options data can be settled and organized to be used for the other layers.

**3.3.1 Data Base**

The Data Base is charged with options daily data from .csv files, gave by the Supervisor, from 2011 to 2017, from deltaneutral.com and historicaloptiondata.com. Each entry represents an option contract and has 13 fields: Company Symbol, Underlying Asset Price, Exchange, Option Root, Option Extension, Type, End Date, Begin Date, Strike Price, Last Price, Bid Price, Ask Price and Volume. All this parameters were explored and presented during the Background section and will be used along the project.

**3.3.2 Options Plot**

An option plot is made of the values that a determined contract has along its lifetime and begins on the contract signing moment until its end. This plot is achieved by searching and gathering the values of consecutive options contracts, that only differ from each other on the beginning date. The options lose value along their lifetime due to time decay and so on it is expectable that the last values of an option lifetime plot are around zero. Also it is important to understand that the underlying asset price
is inversely proportional to the put option prices as it is seen in 3.2 and proportional to call option prices which means that if the underlying asset price decreases, the value of put contracts increases and in reverse for call contracts. In the table 3.1 there is an example of the options that would represent the option SPX140621P01800000 time value from 01/23/2014 to its maturity, 06/21/2014. The Last Values are united in a plot that describes the value of an option contract along its life.

<table>
<thead>
<tr>
<th>Desired Option</th>
<th>Company</th>
<th>Type</th>
<th>Begin Date</th>
<th>End Date</th>
<th>Strike</th>
<th>Last Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Option 1 Values</td>
<td>SPX</td>
<td>Put</td>
<td>01/23/2014</td>
<td>06/21/2014</td>
<td>1800.0</td>
<td>57</td>
</tr>
<tr>
<td>Option 2 Values</td>
<td>SPX</td>
<td>Put</td>
<td>01/25/2014</td>
<td>06/21/2014</td>
<td>1800.0</td>
<td>78.5</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Option n Values</td>
<td>SPX</td>
<td>Put</td>
<td>06/21/2014</td>
<td>06/21/2014</td>
<td>1800.0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3.1: Option Values through time

3.3.3 Window Plot

The strategy used for the called window plot is represented in figure 3.6 with legends A, B and C that are represented in real examples in figures [3.2], [3.3] and [3.5] correspondingly and as it is possible to see the plot is the result of many temporal windows put together. This strategy begins with the cropping of the full time period in windows with the same size defined as window length size. Each window corresponds to a time interval (for example the first window of the year 2011 with twenty days window length is from the 1st of January to the 21st of January). After this first step, it will be chosen, for each window, from the data base, the option contract that best fits inside it. For this selection some options parameters, that will reproduce different plots, will be used further ahead and explained below in the inputs list. When the selection is complete, all the chosen contracts will be represented in the respective window and then united in two different ways:

1. The first plot (A), is the union of each window with its option inside normalized to the last one which means that one by one, starting from the first to the last, windows are added to combine the final plot and each new window added is normalized to the last point of the last window assembled changing its scale. To conclude, when all the windows are joined and normalized, the last value of the final graphic is set to 10 changing also proportionally all the plot. As it is seen, option values decay over time and so this graphic, as being the union of many option values along time is tendentiously decreasing. This plot will be presented to the investment layer and it will give an idea of the market tendencies rather than the absolute values of the contracts which mean that if the plot is rising or decreasing the value of the underlying asset price is also moving in the same way for a call price or in reverse way for a put price as it is possible to see in the figure 3.6 where puts were used. With different inputs this plot will be represented with different options and so on.
with bigger or smaller oscillations due to options characteristics.

2. The second plot (B), is only the union of each window with no normalization and will be used for the investment simulator layer to open and close orders of option contracts by their real values.

Now, that the options for the plots (A) and (B) are selected and built, a third plot will be done. New options that only differ from the previously selected by a strike price difference will be searched and plotted as (B) was plotted (without any normalization). This plot can be built with different strike prices differences and it is usually parallel to the (B) plot because their contract values differences are proportional to their strike prices differences. This plot was called a strike band (C), and will also be used in the investment layer for trading. For example in figure 3.5, to form the green line, a search is made for the options that differ in the strike price, from those selected for the windows of the graphs A and B (yellow line), from 1 to 30 and then the ones with the highest volume are selected and plotted.

This data organizing system will be used for different tests since it is possible to represent the price of puts or calls values along time with contracts of different conditions. For example, options with a smaller time for maturity will have in module a shorter Delta value and then will oscillate less its price than options further to maturity which respond more to the underlying asset price oscillations. Another example, options with longer time duration will have a smaller Theta or time decay and so will be losing value along time less than options with smaller duration. This inputs will affect the final results of the plots and so of the all process, and can be manipulated for different study cases. So on, some inputs must be given to the data layer for plotting and are explained above along with figure 3.7 where the mechanism for choosing an option to its window is schematized.

1. Time Period:

   • The Time Period is the temporal gap in which the data will be represented. The time period is gave by a begin and an end date and it is chosen depending on the test wished.

2. Window Size:

   • The Window Size measures each window time gap. Bigger windows will be filled with bigger contracts.

3. Options Specifications:

   • With the Time Period and window length size defined we will search for the best options that fit each window by a set of parameters.

(a) Options Company

(b) Options Type - Call or Put
(c) Options Contracts Duration - minimum and maximum of contracts duration on the options searched.

(d) Option Contracts Intrinsic Value - minimum and maximum of intrinsic values on the options searched.

(e) Maturity Time Remaining - minimum remaining lifetime options must have after window end.

(f) Strike Bands Differences - minimum and maximum strike differences for each strike band representation.

(g) Volume Maximization - in addition to the characteristics previously mentioned, if many options contracts fit the parameters and thus a window, the one with the highest volume will be selected.

In the figure 3.3 it is possible to see the plot of the year 2014 with window length 60 days. It is also possible to see that none of the windows last value is zero because those options have still some days to maturity. Comparing this figure made of put contracts to figure 3.4 it is observable that the price oscillates in the reverse direction of the underlying asset price. The figure 3.2 is the result of the figure 3.3 normalization as explained previously.

In the figure 3.5, the yellow line represents the (B) plot with ITM put contracts and the other lines are different strike bands with different strike differences. As the yellow line represents tangential ITM contracts, the rest of the other lines, gather lower strike contracts than the previous selected for (B), and so, cheaper contracts, because they are OTM. In figure 3.6, all the Data Layer process is resumed. First the time period is divided in windows, secondly for each window an option is selected, then plots are formed and outputted for other layers, as explained above.
Figure 3.3: Windows with the chosen and cropped put options (B)

Figure 3.4: Underlying Asset Value (SP500)

Figure 3.5: Strike Bands (C)
Figure 3.6: Data Layer
3.4 Indicators Layer

The Indicators layer is another block of the described model and in this part data will be reviewed, analyzed and graded. In this phase, the already prepared (A) plot data will be submitted to a set of tests that will try to figure which are the best moments for entering and exiting the market. There are many indicators, that can be used for data analysis, technical indicators, economic indicators or sentiment indicators, for example, but it is fundamental to choose those which are appropriate. As said in [3.3.3], our data is a normalization of an options contracts set and as viewed in [2.2.3] options contracts, due to time decay, are tendentiously decreasing its value through time. Along with this, since options offer a big leverage to investors, big price oscillations will be seen so on indicators should be choose for this features. The Technical Indicators layer is inspired on Gorgulho’s work [46] and improved features were added. Bellow it is explained the use of the chosen indicators which will evaluate the prepared data.

3.4.1 Technical Indicators

The Technical Analysis is one of the main tools that investors use to study and understand better timings for market entering and as mentioned previously [2.2.4], technicians believe that markets price reflects all relevant information for future investments. Gorgulho used the Relative Strength Index, the Moving Average Convergence Divergence and the Rate of Change with positive results and so those will be adapted for our problem and will be set together with another improvement indicator, the bollinger bands. Each of the below described technical indicators will require some input values for their construction and will sweep the normalized window plot giving a mark at each and every moment. During this inspection,
each indicator will grade these data giving:

- **High Grade**: 1 point for a possible short situation (asset is going to decrease)
- **Low Grade**: -1 point for a possible long situation (asset is going to increase)
- **0 c.c.

This evaluation mechanism was developed with inspirations on Gorgulho's work. In this thesis the data gave to the indicators layer is naturally decreasing and so on momentum indicators which can *measure the velocity of price changes* will be used to measure changing moments in the market. If the stock market value increases, the call options values will increase contrary to put options values that decrease, as seen before, so what is a high grade for a call is a low grade for a put. In the following subsections each technical indicator is explored and reviewed.

### 3.4.1.1 Relative Strength Index

The Relative Strength Index (RSI) is a momentum indicator which measures the level of recent price changes to evaluate overbought or oversold conditions in an asset. The RSI is an oscillating indicator with values between 0 and 100. Typically RSI values under 30 are considered to be a reflex of oversold conditions, while values above 70 reflect overbought moments for stock markets prices. In the equation (3.1) it is possible to see the RSI formula, where the Average Gain with period $n$ is the quotient between the average percentage of the increases in the last $n$ positions and the number of times an increase occurred in the past $n$ values and the Average Loss with period $n$ is the quotient between the average percentage of the decreases in the last $n$ positions and the number of times a decrease occurred in the past $n$ values. In table 3.8, it is represented the moments where a put price will have a high and a low mark.

$$RSI(n) = 100 - \frac{100}{1 + R(n)} \quad (3.1)$$

$$R(n) = \frac{AverageGain(n)}{AverageLoss(n)} \quad (3.2)$$

<table>
<thead>
<tr>
<th>Marks</th>
<th>Relative Strength Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>when the indicator is above the min limit parameter</td>
</tr>
<tr>
<td>Low</td>
<td>when the indicator is over the max limit parameter</td>
</tr>
</tbody>
</table>

*Table 3.2: RSI Marks Rules - Put Contracts*
The RSI will need three parameters, a period, a maximum and a minimum thresholds to provide the evaluation.

### 3.4.1.2 Rate of Change

The Price Rate of Change (ROC) is a momentum indicator that measures the percentage change in price between the current price and the price $n$ periods ago. The ROC indicator is plotted centered in zero, with the indicator moving upwards into positive values if prices are increasing, and moving into negative values if prices stop increasing and begin to fall.

$$ROC(n) = 100 \cdot \frac{AssetValue(t) - AssetValue(t - n)}{AssetValue(t - n)}, t \geq n$$  \hspace{1cm} (3.3)

<table>
<thead>
<tr>
<th>Marks</th>
<th>Rate of Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>when the indicator crosses the 0 from above</td>
</tr>
<tr>
<td>Low</td>
<td>when the indicator crosses the 0 from below</td>
</tr>
</tbody>
</table>

Table 3.3: Rate of Change Marks Rules - Put Contracts

So on, when the ROC chart is moving from the forth quadrant to the first, in other words is coming from negative values to positive, the price of an asset is beginning to increase. On the contrary if the ROC chart crosses from above the zero line the asset price is beginning to decrease. In these parameter, Gorgulho uses a point and its previous to see if a zero interception occurs in the indicator and so on.
grade it with a good or bad mark. In the architecture designed in this thesis a derivative parameter is also implemented so that, to see if a zero interception was made in the actual point, a comparison to its value in another moment, some points back, is made. This derivative parameter can be one and fully adopt Gorgulho’s work strategy or it can be another value and so, more possible different analysis are contemplated and will be analyzed. To conclude, two parameters are needed for the ROC, a period and a derivative parameter.

3.4.1.3 Bollinger Bands

The Bollinger Bands (BB) are a momentum indicator defined by a set of lines plotted some standard deviations away from a simple moving average. One of the lines is set positively to the Simple Moving Average (SMA) and the other negatively. The standard deviation is proportional to volatility so when the markets become more volatile the bands widen and in reverse, they contract. Bollinger Bands also try to figure out when oversold or overbought conditions occur and it is defined that when the upperband is below the signal, overbought conditions exist and when the lowerband is above the signal, oversold conditions happen.

\[
BB_{middleband}(n, k_1, k_2) = SMA(n) = \frac{\sum_{i=t-n}^{t} PlotValue(i)}{n}, t \geq n \tag{3.4}
\]

\[
BB_{upperband}(n, k_1, k_2) = BB_{middleband}(n, k_1, k_2) + k_1 \cdot \sigma(n) \tag{3.5}
\]
\[ BBlowerband(n, k_1, k_2) = BBmiddleband(n, k_1, k_2) - k_2 \cdot \sigma(n) \] (3.6)

Figure 3.10: Bollinger Bands of VIX with put type marks

<table>
<thead>
<tr>
<th>Marks</th>
<th>Bollinger Bands</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>middle band is above the lower band</td>
</tr>
<tr>
<td>Low</td>
<td>middle band is over the upper band</td>
</tr>
</tbody>
</table>

Table 3.4: Bollinger Bands Marks Rules - Put Contracts

This indicator will need, as parameters, a time period and two constants to manipulate the upper and lower bands distance to the SMA or middleband \((k_1\) and \(k_2\)).

3.4.1.4 Moving Average Convergence Divergence

The Moving Average Convergence Divergence (MACD) is a trend-following momentum indicator that shows the relationship between two moving averages. The MACD is usually calculated by subtracting a slow-period Exponential Moving Average (EMA) from a fast-period EMA. The result of that calculation is called the MACD line. Then an EMA of the MACD line, called the signal line, is plotted which works as a trigger for buy and sell moments. The Histogram is the subtraction between the signal line and the MACD line. Traders may buy the stock asset when the MACD line crosses above its signal line and sell when the MACD line crosses below the signal line.

\[ EMA(n, t) = PlotValue(t) \cdot \frac{2}{n + 1} - EMA(n, t - 1)(1 - \frac{2}{n + 1}) \] (3.7)
Table 3.5: Moving Average Convergence Divergence Marks Rules - Put Contracts

<table>
<thead>
<tr>
<th>Marks</th>
<th>Moving Average Convergence Divergence</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>MACD line crosses below the signal line</td>
</tr>
<tr>
<td>Low</td>
<td>MACD line crosses above the signal line</td>
</tr>
</tbody>
</table>

To conclude, the MACD indicator will need a fast and a slow period for the EMA’s, a period for the signal line and a derivative parameter to be used as improvement like in ROC [3.4.1.2] to see if a cross between the MACD signal and the signal line occurs.

3.4.2 Economic Indicators

3.4.2.1 Volatility Index

The VIX Index is explained above and it will be used in the architecture of the system because it can be useful for detecting moments where the market is highly volatile and so dividing moments for go short or go long. Since options contracts are used, a fall in the stock market can result in great options prices oscillations and so on it is very important to avoid high volatile moments because if an order goes wrong losses can be deadly for an investor. The VIX is used with bollinger bands and when the VIX Index value is above the upper band it triggers the investment layer only to short avoiding big stock crashes, by closing long positions, and trying to profit from an asset decrease.
3.5 Investment Simulator Layer

The Investment Simulator Layer receives the signal to trade (B) and the correspondent marks gave by the Indicators Layer. In this layer, all the four technical indicators individual grades will be weighted, summed and applied in an investment simulation that trades the trade signal which has the price of option contracts along time. The following rule is applied in this process as Gorgulho did:

$$\sum w_i = 1, \quad i = 1, ..., 4$$

(3.8)

For this Layer a fund of 1M is ready for simulation and the total risk can only be more than 1M if the past transactions achieved profit increasing the money the fund has to invest. Both long and short positions are positions that provide income, the first is by the increasing of the asset value and the second by the decreasing. Many were the studies seen with long and short positions strategy but the problem is that this types of orders alone cannot be risk measured, so as we use options data along this thesis, instead of using long and short positioning bear and bull option spreads are used. Option spreads can be risk measured unlike an unique option order, since they have a maximum win and a maximum lost possible calculable even before orders opening. On figure 3.12, it is presented the fluxogram of this block. If the final point of a window is achieved, all the positions are closed. A mechanism of trailing stoplosses is set so that if a crash on the market happens, some of the profits are saved. To conclude, positions are opened based on the technical indicators analysis on the signal made of the options chosen and the VIX. Below bear and bull spreads are described.

![Investment Simulator Layer Fluxogram](image-url)
3.5.1 Bear and Bull Spreads

A Bull Spread consists in buying and selling, at the same time, two option contracts in the same asset that only have different strikes and this position gets an income through the UAP value increasing. On the contrary, a Bear Spread consists in buying and selling, at the same time, two options contracts in the same asset that have only different strikes and this position gets an income through the UAP value decreasing.

In figures 3.13 and 3.14 examples of a bull and a bear put spread are shown.

A Bull Put Spread consists in buying a lower strike put and sell a higher strike put. Both this contracts have the same expiration date, both were bought at the same time and both can be ITM, OTM or the lower strike OTM and the higher ITM and so the price of the bought put must be smaller than the sold one so that a positive spread is received. On one hand, if the UAP falls down the smaller strike, during the contracts period, the lost is equal to the different of both strike plus the spread won at the signing date. On the other hand, if the UAP rises above the higher strike, the income is equal to the spread won at the signing date. In figure 3.13, a bull spread is made at 1430 and it is possible to see the lost or the return according to the oscillation of the underlying asset price.

A Bear Put Spread consists in buying a higher strike put and sell a lower strike put. Both this contracts have the same expiration date, both were bought at the same time and both can be ITM, OTM or the lower strike ITM and the higher OTM and so, the price of the bought put must be higher than the sold one so that a negative spread is received. On one hand, if the UAP falls down the lower strike, during the contracts period, the income is equal to the different of both strike minus the spread lost at the signing date. On the other hand, if the UAP rises above the higher strike, the lost is equal to the spread made at the signing date.
In figure 3.14, a bear spread is made at 1370 and it is possible to see the lost or return according to the oscillation of the underlying asset price.

Both the examples in figures 3.13 and 3.14 are orders that have a possible maximum profit of 20 and a possible maximum lost of 30 resulting in a 66 % return on the investment made.

For the investment layer it is needed the spreads data and that’s why the strike bands are calculated.

3.5.2 Orders

It is important in this moment to define how an order is processed and it is opened. Since we are using option spreads, this mechanism will look for moments with a grade under a minimum threshold to do a bull spread in the market and above a maximum threshold to make a bear spread in the market. This grade mechanism will be helped by the VIX which will say when volatility is too high and so a crash in market is predictable closing bull positions and making only bear spreads until volatility return to lower values. Since a window plot will be used, at the end of each window all the orders will be closed because another contract is ready for trading. Besides this way and the VIX there is only one case in which an order can be closed, a stoploss. The orders risk is also defined as input. All this parameters shall be given to the investment layer and different parameters will generate different results.

3.5.3 Trailing Stop Losses

As discussed in [2], traders have seriously difficulties in knowing when to stop a loss and in Adam Y.C. Lei and Huihua Lib [53] studies trailing stop losses showed very good results and so on they will be used in this system. Trailing stop losses are adaptive stop losses which grow if the income from a determined position increases to a new maximum. If an order keeps increasing its profit, the stoploss will also be increasing along time avoiding to lose the profit if a crash happens. If the income from a position
falls, there wont be a stop loss update and if it matches the stop loss the position is closed, avoiding more risks.

3.6 Optimization Layer

The Optimization Layer is, as it is called, an optimization system and so on requires a process to optimize. This process is the union of both the indicators layer and the investment layer and, as presented above, some inputs are needed for their working. Using a personalized chromosome for the process designed and explained above the GA will search for an optimum solution to the problem in case.

3.6.1 Genetic Algorithm

The Genetic Algorithm has the objective of improve the process results and find optimum solutions. In the section [2], it is done a review on how the mutation, crossover and selection works and the optimization layer was implemented with that characteristics plus some improvements, so it will only be presented here the Chromosome of the algorithm and the improvements made (the use of Hypermutation, Hall of Fame and Immigrant Elements). Along the algorithm the fitness function will maximize the total amount of money the fund has with the objective of maximizing the profit along the train period.

3.6.1.1 Chromosome

For the GA, it is needed to define the structure of the chromosome, which means, it is necessary to define the variables that the GA will give as inputs to the process. The process developed is on the indicators and investment simulator layers so on, for the first layer, it is needed the inputs for the technical and economic indicators, for example, the number of days the RSI is going to check the window plot values or the period for the VIX bollinger bands. For the second layer, it is needed inputs to the thresholds that define the market entrance moments (mark1 and mark2), the weights which define the relative importance of each technical indicator grade (which the sum is 1), the net risk value of each order, the stop-loss measure and the strike band used for the spreads. The chromosome defined then, has a length of 24 elements and each of those elements can oscillate its value inside a predefined value range. In figure 3.15 the full chromosome of the GA is represented and in table 3.6 the range in which the chromosome values will be generated.

3.6.1.2 Hall of Fame

The Hall of Fame mechanism is an improvement in the algorithm and it is popped when a determined number of generations occur without the appearance of a new maximum. The Hall of Fame basically saves the best solutions found along the algorithm and when called, it provides the parents for the new
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Range</th>
<th>Parameters</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight 1</td>
<td>[1, 25]</td>
<td>BB VIX period</td>
<td>[15, 50]</td>
</tr>
<tr>
<td>Weight 2</td>
<td>[1, 25]</td>
<td>BB VIX std1</td>
<td>[0.5, 2.5]</td>
</tr>
<tr>
<td>Weight 3</td>
<td>[1, 25]</td>
<td>BB VIX std2</td>
<td>[0.5, 2.5]</td>
</tr>
<tr>
<td>Weight 4</td>
<td>[1, 25]</td>
<td>MACD period1</td>
<td>[16, 30]</td>
</tr>
<tr>
<td>RSI period</td>
<td>[5, 30]</td>
<td>MACD period2</td>
<td>[5, 15]</td>
</tr>
<tr>
<td>RSI max threshold</td>
<td>[65, 90]</td>
<td>MACD period3</td>
<td>[2, 8]</td>
</tr>
<tr>
<td>RSI min threshold</td>
<td>[10, 35]</td>
<td>MACD derivative</td>
<td>[1, 4]</td>
</tr>
<tr>
<td>ROC period</td>
<td>[5, 30]</td>
<td>Strike Band</td>
<td>[1, 4]</td>
</tr>
<tr>
<td>ROC derivative</td>
<td>[1, 4]</td>
<td>Stoploss</td>
<td>[10, 50]</td>
</tr>
<tr>
<td>BB period</td>
<td>[5, 30]</td>
<td>Order Value</td>
<td>[1, 150000]</td>
</tr>
<tr>
<td>BB std1</td>
<td>[0.5, 2.5]</td>
<td>MARK1</td>
<td>[0, 1]</td>
</tr>
<tr>
<td>BB std2</td>
<td>[0.5, 2.5]</td>
<td>MARK2</td>
<td>[-1, 0]</td>
</tr>
</tbody>
</table>

Figure 3.15: Chromosome

Table 3.6: GA chromosome values range

offspring avoiding step backs and losses of memory in the system. This mechanism permits a bigger
hypermutation and random immigrants ratio because the best solutions will always be saved and will
never be lost.

3.6.1.3 Hypermutation

The Hypermutation mechanism *adaptively introduces diversity when needed* [49] and it will be used
inside the improved GA. Along the generations if a better solution doesn’t get popped the mutation
rate will increase until it reaches a threshold. If the threshold is reached the Hall of Fame mechanism is
popped and the initial mutation ratio is again applied. This mechanism will introduce more possibilities
for the algorithm to find new solutions and it prevents that the mechanism doesn’t get stuck in a local
maximum as it makes more random values get in the process.
3.6.1.4 Random Immigrants

Along with Hypermutation, this mechanism introduces new random chromosomes during the optimization making the algorithm more capable of discovering new optimum solutions avoiding local minima and reproducing new unseen solutions for the process.
3.7 Chapter Conclusion

In this chapter, in a first moment, it is given the big picture of the developed system. All the four main layers, Data Layer, Indicators Layer, Investment Simulator Layer and Optimization Layer were presented and summarized.

In a second moment, all the layers were explored in detail one by one. The Data Layer was examined and it was presented the mechanism of data organizing which prepares the raw data for the process that will be optimized. Then, the Indicators Layer was presented and the indicators used were justified and characterized. Following this, the Investment Simulator Layer was presented and it was explained how the bull and bear spreads could take profit from increases and decreases in the stock market. Concluding, the Chromosome of the Optimization Layer is described and the improvements like Hypermutation, Hall of Fame and Immigrant elements on the GA were explored.
Chapter 4

System Evaluation
4.1 Overview

In this section it is tested the implementation described on the last chapter. In a first moment, it is explained how data was divided and settled for the model developed, then some evaluation metrics for the results are described for evaluation of the different tests established. Following this steps, the two strategies conceived are presented along with the characteristics they own. Then, to conclude, all the three case studies are characterized and an analysis is made on the results.

4.2 Data Split

For the case studies, a rolling window mechanism will be used, which means that options data from 2011 to the end of 2017 will be divided in three different sets, each with 75% train data and 25% test data. Since the data used is the plot made of the union of windowed options, the rolling window measure used will be a multiple of the options window length so that each train and test sets end in a contract option window end. Since the option windows length will be 100 days the rolling window measure will be 1200 days for the train data and 400 days for the test data. In table 4.2 it is possible to see how data was divided and in figure 4.1 it is represented the scheme for the rolling window used for train and test data. The optimization was done with the GA on the train data and then the final parameters are applied on the test data to achieve results for comparison.

Along the data a media between the ask and the bid price will be used since it represents the median point between the minimum value a seller is willing to sell a contract and the maximum value a buyer is willing to buy a contract. An explanation on the ask, bid and last price is made on section 2.

<table>
<thead>
<tr>
<th>Train</th>
<th>Nº Train days</th>
<th>Test</th>
<th>Nº Test days</th>
</tr>
</thead>
<tbody>
<tr>
<td>06/06/2011 - 17/09/2014</td>
<td>1200</td>
<td>18/09/2014 - 22/10/2015</td>
<td>400</td>
</tr>
<tr>
<td>10/07/2012 - 22/10/2015</td>
<td>1200</td>
<td>23/10/2015 - 25/11/2016</td>
<td>400</td>
</tr>
<tr>
<td>14/08/2013 - 25/11/2016</td>
<td>1200</td>
<td>26/11/2016 - 31/12/2017</td>
<td>400</td>
</tr>
</tbody>
</table>

Table 4.1: Data train and test division dates

Figure 4.1: Data split scheme for the rolling window
4.3 Evaluation Metrics

Along the research made for this work it was noticed that many academics used the Return on Investment as a main metric for results comparison and so on it will be used. The ROI metric is a percentage that indicates the gain or loss generated on an investment relative to the amount of money invested and it is calculated as in (4.1).

\begin{equation}
ROI = \frac{Gain - Cost}{Cost}
\end{equation}  

Along with ROI, listed below, are also other evaluation metrics used for study cases analysis:

1. Number of Spreads - It indicates the number of spreads made along the time period. The number of spreads opened can be a measure of each spread individual risk since the approach designed conceives an initial fund with limited money.

2. Percentage of Successful Spreads - It indicates the percentage of successful spreads and it is important to measure the accuracy of the used model since, even with positive results, a test might have a low successful order percentage.

3. Percentage of Unsuccessful Spreads - It indicates the percentage of unsuccessful spreads and, as the last feature, it can show if the model is steady or unsteady.

4. Average Spread Profit - It indicates how well are spreads being made. This feature is the profit made divided by the numbers of order and the highest, the better the results are.

5. Biggest Spread Profit - It indicates the best spread made on a specific test. This can be a good feature to compare in which conditions different tests get good returns.

6. Biggest Spread Lost - It indicates the worst spread made on a specific test. This can be a good feature to see how and when the mechanism fails an order.

7. Biggest Draw Down - It indicates the maximum fall that happens along the investment in the portfolio made of the money in the fund and the actives owned.

8. Number of Stoplosses - It indicates the number of positions closed by a stoploss

4.4 Used Strategies

In this thesis, three different strategies were developed with three different GA architectures for results comparison and for the study cases all the strategies will be done five times in each of the three windows that result of the data split for accuracy measure. Below are described the three optimization processes used:

2. Improved Genetic Algorithm 1 - Based on the background description plus the State-of-the-Art improvements described:
   - Random Immigrants with 10% rate.
   - Hypermutation which means an oscillating mutation rate with increment steps of 7% capped at 28%.

3. Improved Genetic Algorithm 2 - Based on the background description plus an additional State-of-the-Art improvement described:
   - Random Immigrants with 10% rate.
   - Hypermutation which means an oscillating mutation rate with increment steps of 7% capped at 28%.
   - Hall-of-Fame implementation with the best results along the generations.

In table 4.2 it is presented the characteristics of the strategies used.

<table>
<thead>
<tr>
<th>Standard GA</th>
<th>Improved GA 1</th>
<th>Improved GA 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 Elements</td>
<td>100 Elements</td>
<td>100 Elements</td>
</tr>
<tr>
<td>50 Generations</td>
<td>50 Generations</td>
<td>50 Generations</td>
</tr>
<tr>
<td>Elitism Selection</td>
<td>Elitism Selection</td>
<td>Elitism Selection</td>
</tr>
<tr>
<td>50% Crossover</td>
<td>50% Crossover</td>
<td>50% Crossover</td>
</tr>
<tr>
<td>10% Mutation</td>
<td>10% Random Immigrants</td>
<td>10% Random Immigrants</td>
</tr>
<tr>
<td></td>
<td>Hypermutation</td>
<td>Hypermutation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HOF</td>
</tr>
</tbody>
</table>

Table 4.2: Strategies characterization

4.5 Case Studies

Along this thesis it was created a model to plot option contracts values along time by selecting options for certain parameters in the data layer. This implementation was done and for both, the first and the second case studies, it was developed different put options alignments. For the last study case it was used a different mechanism for the trailing stoploss in the investment simulator layer. In all the case studies usually OTM put options were used, even for the strike bands of the options chosen. So on, it is important to define how put contract time to maturity reacts along the lifetime of a put option.

1. Delta and Time to Maturity - the delta of an OTM put option approaches 0 as the option approaches expiration and for ITM put options approach -1.
2. Time Decay and Time to Maturity - The further a put option is from the expiration, the smaller is the time decay.

In the first case study, the options window plot will be made with options closer to maturity compared to the options used for the second case study which are further to maturity. With this maturity difference, the chosen options for the first study case will overall be less exposed to delta and more exposed to time decay than those used on the second study case and this gap on maturity will influence all the optimization process. For the third case study, options used for the first case study will be reused but now with a trailing stoploss on the option prices rather than on the stock prices. For the first two study cases an absolute trailing stoploss was used on the stock price and for the third a percentage trailing stoploss is used closing positions when a loss is seen in a value range of 10% to 20%. All this three case studies will use the same strike band margins used for the plot 3.5 and will result in some conclusions when compared.

4.5.1 Study Case 1 - Options closer to maturity

For the first case study, the Data Layer is set with the parameters on table 4.3.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Window Size</td>
<td>100</td>
</tr>
<tr>
<td>Strike Difference Minimum</td>
<td>-20</td>
</tr>
<tr>
<td>Strike Difference Maximum</td>
<td>0</td>
</tr>
<tr>
<td>Duration Contracts Minimum</td>
<td>0</td>
</tr>
<tr>
<td>Duration Contracts Maximum</td>
<td>300</td>
</tr>
<tr>
<td>Maturity</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 4.3: Data Layer First Case Study Parameters

For this data layer inputs set, contracts are searched and organized to form the previously explained plots (Section 3). Then the Data layer will send this signal for the process that it will be optimized (technical indicators layer and investment simulator layer) by both the used strategies above described, the Standard GA and the Improved GAs. On figure 4.2, it is possible to see the ROI along time of the best results developed for each strategy.

In table 4.4 are exposed the evaluation metrics for the best results represented in figure 4.2.

From table 4.4 and figure 4.2, it is possible to conclude that the Improved GAs outperform the Standard GA, achieving a bigger average ROI and a percentage of successful spreads slightly bigger for the best results achieved which can be a signal that improvements made like Random Immigrants, Hypermutation and the use of an Hall of Fame can be good features to set in a GA for a better performance. This conclusion might be taken since many tests were made and in figure 4.2 it is represented the average of the best results and so a tendency is seen for the improved GAs to have better results that the Standard one. The best result of all was achieved by the Improved GA 2.
### Evaluation Metrics

<table>
<thead>
<tr>
<th>Evaluation Metrics</th>
<th>Improved GA</th>
<th>Standard GA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Spreads</td>
<td>277</td>
<td>267</td>
</tr>
<tr>
<td>Percentage of Successful Spreads</td>
<td>53.3</td>
<td>52.4</td>
</tr>
<tr>
<td>Percentage of Unsuccessful Spreads</td>
<td>46.7</td>
<td>47.6</td>
</tr>
<tr>
<td>Average Spread Profit</td>
<td>0.98</td>
<td>0.92</td>
</tr>
<tr>
<td>Biggest Spread Profit</td>
<td>20.4</td>
<td>20.9</td>
</tr>
<tr>
<td>Biggest Spread Lost</td>
<td>-3.9</td>
<td>-6.2</td>
</tr>
<tr>
<td>Biggest Drawdown</td>
<td>-31.6</td>
<td>-28</td>
</tr>
<tr>
<td>Number of Stoplosses</td>
<td>142</td>
<td>124</td>
</tr>
<tr>
<td>ROI</td>
<td>263.69</td>
<td>245.4</td>
</tr>
</tbody>
</table>

Table 4.4: Evaluation Metrics for the First Study Case

![Figure 4.2: ROI (%) for the two Strategies](image)

In table 4.5 it is presented the Chromosome of the best result obtained with the Improved GA for each data set. Each line represents the Chromosome that was found along the optimization process which results in the best profit on the test data for the three data sets defined.

<table>
<thead>
<tr>
<th>-</th>
<th>W_RSI</th>
<th>W_ROC</th>
<th>W_BB</th>
<th>W_MACD</th>
<th>RSI_p</th>
<th>RSI_max</th>
<th>RSI_min</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.35</td>
<td>0.03</td>
<td>0.36</td>
<td>0.26</td>
<td>23</td>
<td>65</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>0.48</td>
<td>0.05</td>
<td>0.32</td>
<td>0.16</td>
<td>24</td>
<td>90</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>0.31</td>
<td>0.13</td>
<td>0.53</td>
<td>0.03</td>
<td>11</td>
<td>67</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>-</th>
<th>ROC_p</th>
<th>ROC_d</th>
<th>BB_p</th>
<th>BB_s1</th>
<th>BB_s2</th>
<th>BB_pVIX</th>
<th>BB_s1VIX</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>23</td>
<td>4</td>
<td>17</td>
<td>0.63</td>
<td>2.41</td>
<td>35</td>
<td>2.19</td>
</tr>
<tr>
<td>2</td>
<td>27</td>
<td>2</td>
<td>24</td>
<td>0.82</td>
<td>1.58</td>
<td>19</td>
<td>2.41</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>4</td>
<td>24</td>
<td>0.59</td>
<td>1.57</td>
<td>17</td>
<td>2.47</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>-</th>
<th>BB_s2VIX</th>
<th>MACD_p1</th>
<th>MACD_p2</th>
<th>MACD_p3</th>
<th>MACD_d</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.41</td>
<td>26</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>1.51</td>
<td>17</td>
<td>13</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>1.71</td>
<td>22</td>
<td>15</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>-</th>
<th>Stoploss</th>
<th>Order Value</th>
<th>Mark1</th>
<th>Mark2</th>
<th>Strike Band</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>42</td>
<td>145538</td>
<td>0.98</td>
<td>-0.03</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>37</td>
<td>149977</td>
<td>0.47</td>
<td>-0.01</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>37</td>
<td>147254</td>
<td>0.11</td>
<td>-0.03</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4.5: Best Chromosome for each rolling window for the best improved GA result
From table 4.5, it is seen that the bollinger bands indicator is the heaviest technical indicator used for grading the signal outputted from the data layer and can be analyzed as a successful bet since it was rare to see, during the research the use of this technical indicator. An improvement made on Gorgulho’s model was the implementation of the derivative parameters and it is observable that this strategy is chosen for the best chromosomes. Also it is seen that the strike band chose for all the best results was the closest to the fixed one and that the order value is set to the maximum so that a maximum ROI is obtained.

Figure 4.3: Opened and Closed positions along test time

Figure 4.4: VIX with BB and VIX Signal in 2015

Since its importance along the trading mechanism, the VIX is represented in the figure 4.6 with
the Bollinger Bands that work as a trigger for close bull or bear spreads. When the signal is above the upper band bull spreads are closed and only bear spreads can be made and when its not, bear spreads are closed and only bull spreads can be used. In figure 4.3, it is represented the number of opened and closed orders through time for the best result found.

From figure 4.3, it is seen that most of the positions opened are bull spreads since drawdowns in the asset are very quick and only cut by the VIX indicator by a small time period which takes the opportunity of the model to take profit from bear moments but consecutively profit from bull spreads. Looking to figure 4.3 and 4.6, it is concluded that many asset drawdowns, in 2015, were avoided but in the last represented, on September of 2015, the VIX is to slow to avoid a big drawdown that is observable in the ROI chart.

4.5.2 Study Case 2 - Options further to maturity

For the second case study it is used options further to maturity and the data layer inputs are on table 4.6. In this case options have, after each window, more than 200 days to maturity. This study case will be used for comparison to the first one seen before and as first comment it is imaginable that, since options further to maturity are less exposed to time decay, the profits for this case shall be smaller. What should be really seen is the percentage of successful spreads done by each strategy since even that the options traded are different, when a fall on the underlying asset happens, all the put contracts will increase their values and in reverse, decrease.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Window Size</td>
<td>100</td>
</tr>
<tr>
<td>Strike Difference Minimum</td>
<td>-20</td>
</tr>
<tr>
<td>Strike Difference Maximum</td>
<td>0</td>
</tr>
<tr>
<td>Duration Contracts Minimum</td>
<td>0</td>
</tr>
<tr>
<td>Duration Contracts Maximum</td>
<td>500</td>
</tr>
<tr>
<td>Maturity</td>
<td>200</td>
</tr>
</tbody>
</table>

Table 4.6: Data Layer Case Study Parameters

In figure 4.5, it is possible to see the ROI along time achieved on the best results for both the strategies used and following, the evaluation metrics are shown in the table 4.7 for these case.

Looking for the ROI result, again, the Improved GAs outperform the Standard one and the best result achieved comes again from the improved GA 2, but now some conclusions might be taken comparing the first case study to the second.

For the best solutions, the ROI and the average spread profit of the first study case is bigger than on the second case because options traded in the first case suffer more the effect of time decay and so for bull positions the spread is decreasing and the overall profit is bigger but the percentage of successful
Evaluation Metrics

<table>
<thead>
<tr>
<th>Evaluation Metric</th>
<th>Improved GA</th>
<th>Standard GA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Spreads</td>
<td>271</td>
<td>290</td>
</tr>
<tr>
<td>Percentage of Successful Spread</td>
<td>0.56</td>
<td>0.53</td>
</tr>
<tr>
<td>Percentage of Unsuccessful Spreads</td>
<td>0.44</td>
<td>0.47</td>
</tr>
<tr>
<td>Average Spread Profit</td>
<td>0.79</td>
<td>0.6</td>
</tr>
<tr>
<td>Biggest Spread Profit</td>
<td>9.6</td>
<td>9.4</td>
</tr>
<tr>
<td>Biggest Spread Lost</td>
<td>-4.85</td>
<td>-4.81</td>
</tr>
<tr>
<td>Biggest Drawdown</td>
<td>-14</td>
<td>-15.7</td>
</tr>
<tr>
<td>Number of Stoplosses</td>
<td>132</td>
<td>124</td>
</tr>
<tr>
<td>ROI</td>
<td>213</td>
<td>173.9</td>
</tr>
</tbody>
</table>

Table 4.7: Evaluation Metrics for Study Case

Figure 4.5: ROI (%) for the two Strategies

There are a little bit bigger for the second case study which might indicate that options used for the second study case, with more time to maturity, can make the model create a better rule for the purpose.

Below, in Table 4.8, it is the chromosomes chosen for the best results.

<table>
<thead>
<tr>
<th>-</th>
<th>W_RSI</th>
<th>W_ROC</th>
<th>W_BB</th>
<th>W_MACD</th>
<th>RSI_p</th>
<th>RSI_max</th>
<th>RSI_min</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.11</td>
<td>0.14</td>
<td>0.5</td>
<td>0.25</td>
<td>14</td>
<td>73</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>0.42</td>
<td>0.04</td>
<td>0.35</td>
<td>0.2</td>
<td>19</td>
<td>90</td>
<td>31</td>
</tr>
<tr>
<td>3</td>
<td>0.22</td>
<td>0.07</td>
<td>0.41</td>
<td>0.3</td>
<td>6</td>
<td>75</td>
<td>30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>-</th>
<th>ROC_p</th>
<th>ROC_d</th>
<th>BB_p</th>
<th>BB_s1</th>
<th>BB_s2</th>
<th>BB_pVIX</th>
<th>BB_s1VIX</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>28</td>
<td>4</td>
<td>7</td>
<td>0.64</td>
<td>2.29</td>
<td>16</td>
<td>2.14</td>
</tr>
<tr>
<td>2</td>
<td>29</td>
<td>2</td>
<td>28</td>
<td>1.12</td>
<td>2.34</td>
<td>15</td>
<td>2.29</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>4</td>
<td>22</td>
<td>0.56</td>
<td>1.87</td>
<td>17</td>
<td>2.46</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>-</th>
<th>BB_s2VIX</th>
<th>MACD_p1</th>
<th>MACD_p2</th>
<th>MACD_p3</th>
<th>MACD_d</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.7</td>
<td>17</td>
<td>12</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>1.18</td>
<td>24</td>
<td>12</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>1.46</td>
<td>20</td>
<td>15</td>
<td>8</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>-</th>
<th>Stoploss</th>
<th>Order Value</th>
<th>Mark1</th>
<th>Mark2</th>
<th>Strike Band</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>34</td>
<td>149834</td>
<td>0.47</td>
<td>-0.05</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>46</td>
<td>147523</td>
<td>0.6</td>
<td>-0.01</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>44</td>
<td>145469</td>
<td>0.44</td>
<td>-0.13</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4.8: Best Chromosome for each rolling window for the 2nd Case Study

In this second study case it is possible to see that the VIX Bollinger Bands indicator has now smaller
periods even that their standart deviations are approximately the same. Also it is seen that the bollinger bands continue to be the most heavy indicator for grading the options signal, and now even heavier, followed by the RSI, but now that we have a more stable grading signal since options further to maturity are less exposed to delta and time maturity and it is seen that the periods used for the heaviest indicators are now equal or smaller from the ones used on the first study case and both the margins of the BB are tendentiously more squeezed which can be the response of the system to the new grading signal which is has a smaller delta and oscillates less.

Figure 4.6: VIX with BB and VIX Signal in 2015

Figure 4.7: Opened and Closed positions along test time

In figure 4.7, it is exposed the opened and closed orders along the test time done on the best solution
found and again, the VIX is not fast enough to close positions in some cases, for example in September 2015, where the VIX trigger is done but at a local maximum making a big drawdown in the portfolio. Both the first and the second case study were done for comparison of the results where only it were used different options and regardless the ROI being lower for the second study case, it is seen that it appears to be a better result with a smaller maximum drawdown but with a much smaller biggest spread profit. However it is relevant that the second case study achieved a better percentage of successful spreads.

4.5.3 Study Case 3 - Trailing Stoploss on the spreads

For the third case, the trailing stop loss, instead of being used in the underlying asset price, it is used in the option spread prices with a value which oscillates in the gap [0.8;0.9], closing positions that lose 10% to 20% of their value. The parameters for the Data Layer, for this case, are the same used for the first study case so that the result can be compared.

In figure 4.8, it is possible to see the ROI for the best Improved GA 2 result in the third case study and in table 4.9, the evaluation metrics are exposed for examination.

<table>
<thead>
<tr>
<th>Evaluation Metrics</th>
<th>Improved GA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Spreads</td>
<td>287</td>
</tr>
<tr>
<td>Percentage of Successful Spreads</td>
<td>0.6</td>
</tr>
<tr>
<td>Percentage of Unsuccessful Spreads</td>
<td>0.4</td>
</tr>
<tr>
<td>Average Spread Profit</td>
<td>1.13</td>
</tr>
<tr>
<td>Biggest Spread Profit</td>
<td>19</td>
</tr>
<tr>
<td>Biggest Spread Profit Lost</td>
<td>-3.12</td>
</tr>
<tr>
<td>Biggest Drawdown</td>
<td>-20.3</td>
</tr>
<tr>
<td>Number of Stoplosses</td>
<td>164</td>
</tr>
<tr>
<td>ROI (%)</td>
<td>325.23</td>
</tr>
</tbody>
</table>

Table 4.9: Evaluation Metrics for the 3rd Study Case

Figure 4.8: ROI (%) along the 3rd case study
In this case study, the ROI obtained was the biggest and comparing this result to the first case, the use of trailing stoplosses on the options values can be an important feature that offers the model the chance to avoid the big drawdowns seen in figure 4.

<table>
<thead>
<tr>
<th></th>
<th>W_RSI</th>
<th>W_ROC</th>
<th>W_BB</th>
<th>W_MACD</th>
<th>RSI_p</th>
<th>RSI_max</th>
<th>RSI_min</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1</td>
<td>0.09</td>
<td>0.47</td>
<td>0.34</td>
<td>22</td>
<td>84</td>
<td>31</td>
</tr>
<tr>
<td>2</td>
<td>0.13</td>
<td>0.17</td>
<td>0.52</td>
<td>0.18</td>
<td>30</td>
<td>83</td>
<td>35</td>
</tr>
<tr>
<td>3</td>
<td>0.29</td>
<td>0.26</td>
<td>0.43</td>
<td>0.02</td>
<td>18</td>
<td>83</td>
<td>20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>ROC_p</th>
<th>ROC_d</th>
<th>BB_p</th>
<th>BB_s1</th>
<th>BB_s2</th>
<th>BB_pVIX</th>
<th>BB_s1VIX</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>2</td>
<td>27</td>
<td>0.56</td>
<td>2.23</td>
<td>29</td>
<td>2.48</td>
</tr>
<tr>
<td>2</td>
<td>27</td>
<td>3</td>
<td>17</td>
<td>0.61</td>
<td>2.15</td>
<td>22</td>
<td>2.49</td>
</tr>
<tr>
<td>3</td>
<td>27</td>
<td>3</td>
<td>17</td>
<td>0.57</td>
<td>1.58</td>
<td>26</td>
<td>2.32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>BB_s2VIX</th>
<th>MACD_p1</th>
<th>MACD_p2</th>
<th>MACD_p3</th>
<th>MACD_d</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.95</td>
<td>30</td>
<td>8</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>1.01</td>
<td>29</td>
<td>14</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>2.08</td>
<td>25</td>
<td>14</td>
<td>8</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Stoploss</th>
<th>Order Value</th>
<th>Mark1</th>
<th>Mark2</th>
<th>Strike Band</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.86</td>
<td>146927</td>
<td>0.44</td>
<td>-0.02</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>0.85</td>
<td>142842</td>
<td>0.49</td>
<td>-0.08</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>0.81</td>
<td>149862</td>
<td>0.1</td>
<td>-0.02</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4.10: Best Chromosome for each rolling window for the third case study

From table 4.10, it is seen that the trailing stoploss used in the best results are around 16% and regardless it has occurred more times than for the example showed in the first case study the ROI remains bigger along the time period which can indicate that the fact that the trailing stop loss was set on option contracts can give more elasticity for the model to find a better rule with the technical indicators. This trailing stoploss type, together with the BB on the VIX indicator, that now has bigger periods, show to be the best strategy since the average spread profit is higher than in first case study and the maximum drawdown and maximum spread loss both smaller. For this case study it is also possible to see that the VIX makes bear spreads able to happen. In figure 4.9 and 4.10 it is seen that, because of the technical indicators, positions are not opened in the stock drawdown of September 2015 even that VIX does not triggers a sufficient fast signal for closing positions, which can be a signal that with stoplosses on the option values can facilitate the analysis of the technical indicators, which make the ROI result much more stable and with less oscillations when compared to the first study case. It is also noticed that the strike bands now changed for value two which can mean that with the stoplosses on the options contract values, the model looks for bigger spreads.
Figure 4.9: Positions opened and closed along the 3rd case study

Figure 4.10: VIX the 3rd case study
4.6 Chapter Conclusions

In this chapter, it was made the evaluation of the model developed along this project. Three case studies were presented with the correspondent return on investment and evaluation metric. For each of the case studies, different GA architectures were used concluding that the improved GA with the implementation of the HOF achieves generically better results. Even seeing the ROI that was achieved it is always important to refer and understand that option contracts are a very risky asset where a drawdown can be even bigger that the underlying asset drawdowns. The first study case presented a set of options with the main feature of being closer to maturity compared to the options used for the second study case making possible the comparison between the use of different contracts in the model. With this implementation it is settled that, for the tests made, the options further to maturity achieve a better percentage of successful transactions which might indicate that they fit better the model. In the third case study it is presented a new mechanism for the trailing stoploss, being achieved the best result on this work with a 325% ROI over 1200 days.
Chapter 5

Conclusion and Future Work
5.1 Final Conclusion

In this project it was designed a model to create a rule to trade in the options market in train data that then revealed positive results on the test data. The model is made of an investment simulator layer where technical indicators, VIX and trailing stoplosses trigger entering and exiting moments for bull and bear positions. Both the investment simulator layer and the technical indicators layer are optimized by a genetic algorithm that thrives to discover the best input parameters with the objective of discover patterns that happen along data and fulfill the desired objective of finding a rule that shows to be profitable.

The use of technical indicators revealed to be an acceptable strategy to optimize a signal value like the option contract values through time and the use of RSI, ROC, MACD and specially Bollinger Bands gave interesting results that must be take in account.

The genetic algorithm was set in three different strategies, a standard one, described in many papers and scientific documents read and viewed along the research made and two improved ones with upgrades like Hypermutation, Random Immigrants and Hall of Fame that consistently showed that could overpass the Standard genetic algorithm version.

To conclude, and to be clear, the options market proved to be a risky asset where great return might be achieved but great losses were also seen. Even though the last result seems to be a very profitable one I do advise that for those who desire investing in the option markets, many cautions should be taken and every movement inside the market should be done in a responsible and informed way.

5.2 Future Work

With this work stays the opportunity of improving the already model done and here it is referenced some future arrangements that might gave better results for the mechanism projected.

One big improvement could be the settling of the data layer inside the optimization layer making the genetic algorithm heavier but searching for also the best option parameters to forecast the markets.

A second suggestion would be the use of other technical indicators that respond faster to prices oscillations which could result in a more accurate rule with greater results.

To conclude, another improvement that might done is related with the fitness functions inside the GA where many other parameters might be taken in consideration, like maximum drawdowns and number of successful transactions which will set the order value risk to another values and not maximize it.
Bibliography


62


[37] The Information in Option Volume for Future Stock Prices, Jun Pan and Allen M. Poteshman, 2004

[38] Harari, Yuval Noah, Homo Deus: A Brief History of Tomorrow, 2015


[43] Aguilar-Rivera, Anton and Valenzuela-Rendón, Manuel A new multi-period investment strategies method based on evolutionary algorithms, 2017


Appendices
Chromosomes for the tests done: (test window - study case - strategy)

1 - 1 - IGA

<table>
<thead>
<tr>
<th>Chromosomes</th>
<th>Test Window</th>
<th>Study Case</th>
<th>Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1-IGA</td>
<td>0.125, 0.1875, 0.47916666666666674, 0.2083333333333337, 16.90, 21, 23, 2, 14, 0.91, 2.42, 33, 2, 0.09, 1.79, 24.5, 2, 3, 41, 1.44388, 0.53, -0.13, 1</td>
<td>0.35384615384615387, 0.03076923076923077, 0.35384615384615387, 0.26153846153846155, 23, 65, 15, 23, 4, 17, 0.63, 2.41, 35, 1, 91, 2.3, 3, 2, 42, 1.45538, 0.98, -0.03, 1</td>
<td>0.2857142857142857, 0.163265306122449, 0.42857142857142855, 0.12244897959183673, 26, 71, 16, 21, 4, 16, 0.71, 2.3, 3, 2, 5, 0.55, 30, 6, 2, 2, 40, 1.49562, 0.35, -0.07, 1</td>
</tr>
<tr>
<td>1-1-IGA</td>
<td>0.29850746268656714, 0.2850746268656714, 0.34328358208955223, 0.059701492537313425, 17, 65, 13, 16, 1, 20, 0.93, 2.14, 50, 2, 26, 1, 23, 27, 8, 2, 2, 40, 1.49710, 0.51, -0.04, 1</td>
<td>0.39285714285714285, 0.07142857142857142, 0.3214285714285714, 0.2142857142857145, 0.2142857142857142, 27, 78, 25, 2, 18, 0.54, 2.13, 42, 2, 32, 1.3, 33, 11, 3, 4, 33, 1.40475, 0.26, -0.02, 1</td>
<td>0.4, 0.0, 2.5454545454545454, 0.30909090909090909, 0.03663636363636363, 23, 85, 11, 16, 2, 23, 0.61, 1.55, 16, 2, 34, 1.9, 26, 11, 8, 4, 44, 1.45020, 0.01, -0.01, 1</td>
</tr>
<tr>
<td>1-1-IGA</td>
<td>0.24074074074074074, 0.14814814814814814, 0.4074074074074074, 0.20370370370370378, 80, 10, 24, 2, 23, 0.73, 1.31, 15, 2, 31, 2, 0.26, 14, 8, 4, 44, 1.46637, 0.93, -0.03, 1</td>
<td>0.12121212121212121, 0.15151515151515152, 0.5454545454545454, 0.18181818181818181, 19, 75, 14, 24, 2, 24, 0.81, 1.65, 16, 2, 38, 2, 35, 18, 14, 7, 4, 44, 1.45816, 0.1, -0.09, 1</td>
<td>0.0, 0.0, 25, 0.47727272727272727, 0.04545454545454545, 0.31818181818181818, 0.15909090909090909, 24, 90, 10, 27, 2, 24, 0.82, 1.58, 19, 2, 41, 1.51, 17, 13, 8, 4, 37, 1.49977, 0.47, -0.01, 1</td>
</tr>
<tr>
<td>1-1-IGA</td>
<td>0.06382978723404255, 0.2127659574468085, 0.3829787234042554, 0.3404255319148936, 6, 72, 33, 27, 3, 28, 0.53, 1.91, 17, 2, 47, 0.63, 26, 11, 8, 4, 31, 1.49639, 0.34, -0.01, 1</td>
<td>0.4705882352941176, 0.02941176470588235, 0.35294117647058826, 0.14705882352941177, 12, 66, 18, 13, 3, 25, 0.54, 1.95, 16, 2, 4, 2, 12, 23, 15, 6, 4, 44, 1.45873, 0.29, -0.05, 1</td>
<td>0.21212121212121212, 0.09090909090909091, 0.57575757575757578, 0.12121212121212122, 16, 67, 33, 11, 4, 26, 0.51, 1.89, 17, 2, 37, 0.8, 20, 15, 8, 4, 31, 1.49739, 0.25, -0.09, 1</td>
</tr>
</tbody>
</table>
| 1-1-IGA     | 0.3125, 0.125, 0.53125, 0.03125, 11, 67, 10, 7, 4, 24, 0.59, 1.57, 17, 2, 47, 1.71, 22, 15, 3, 3, 37, 1.47254, 0.11, -0.03, 1 | 0.3125, 0.125, 0.53125, 0.03125, 11, 67, 10, 7, 4, 24, 0.59, 1.57, 17, 2, 47, 1.71, 22, 15, 3, 3, 37, 1.47254, 0.11, -0.03, 1 | 1 - 1 - SGA

66
| [0.07272727272727272,0.2545454545454545,0.4545454545454545,0.21818181818181818,0.1818181818181818,0.16,0.1521739130431870,21,22,3,16,0.57,2.36,40.221,0.61,30,7,2,4,42,148680,0.85,-0.0,1] |
| [0.2895507246376810,0.30434782606895654,0.31884057971014493,0.08695652173913043,18.70,21,22,3,16,0.57,2.36,40.221,0.61,30,7,2,4,42,148680,0.85,-0.0,1] |
| [0.4255319148936170,0.3829787234042554,0.12765957446808510,0.063829787234402552.28,69,20,27,4.15,0.64,2.42,35,2,0.62,16.7,3,1,4.5,149041,0.82,-0.05,1] |
| [0.026315789473684210,0.34210526315789475,0.44736842105263161,0.18421052631578946,17.71,22,19,4,16,0.55,2.32,35,2,19,1,24,21,5,4,2,42,147271,0.74,-0.04,1] |
| [0.125,0.0625,0.1525,0.65625,16.69,20,15,3.12,0.55,1.72,42,1.3,2.25,24,12.2,3.49,135501,0.21,-0.06,1] |

2 - 1 - SGA

[0.10638297872340426,0.10638297872340426,0.5106382978723404,0.2765957446808511,0.13,8,30,24,2,22,0.77,1.57,16,2,39,1.07,18,14,7,4,15,2.44,4.5,142445,0.21,-0.04,1]  
[0.09090909090909090,0.12121212121212121,0.51515151515151515,0.27272727272727272,18,87,15,24,2.5,0.84,2.14,1.7,2.47,2.49,17,13,8,4.5,148517,0.59,-0.09,1]  
[0.05714285714285714,0.14285714285714285,0.6285714285714286,0.17142857142857143,0.01,8,1,22,0.84,1.27,1.57,16,2,33,1.5,4,2,42,148517,0.59,-0.07,1]  
[0.43636363636363634,0.03636363636363636,0.4,0.12572727272727272,16,87,17,25,2,21,0.81,1.51,16,2,37,0.65,19,12,8,4.4,148346,0.06,-0.01,1]  
[0.2,0.03333333333333333,0.5333333333333333,0.2333333333333333,0.2333333333333333,12,89,25,25,2,22,0.83,1,31,15,2,33,1.53,23,15,8,4.5,149499,0.45,-0.03,1]  

3 - 1 - SGA

[0.37837837837837838,0.054054054054054054,0.40540540540540543,0.16216216216216217,10,71,28,15,3,24,0.58,1.32,17,2.43,1.28,28,12,4.4,4.5,140957,0.02,-0.11,1]  
[0.4423076923076923,0.09615384615384617,0.4423076923076923,0.019230769230769232,12,68,32,10,4,24,0.59,1.36,16,2,39,2,12,29,5,4.5,149881,0.01,-0.08,1]  
[0.37209302325581395,0.023255813953488372,0.5813953488372093,0.023255813953488372,15,68,18,30,4,23,0.87,1.57,16,2.35,0.96,19,11,7,4.5,147188,0.22,-0.02,1]  
[0.23404255319148937,0.0851063829787234,0.44680851063829785,0.23404255319148937,11,66,13,12,4,24,0.61,1,6.1,15,2,2.81,0.12,8,10,8,4.37,148873,0.07,-0.04,1]  
[0.12,0.0,28,0.5,0.1,13,68,19,7,4,23,0.7,1,69,17,2.47,1.84,30,6.5,3,37,149767,0.12,-0.02,1]  

1 - 2 - IGA

[0.1111111111111111,0.13888888888888889,0.5,0.25,14,73,15,28,4,7,0.64,2,29,16,2,14,1,7,17,7,2,2,34,149834,0.47,-0.05,1]
| 2 - 2 - SGA |
| 0.2972972972972973,0.1576756756756759,0.2972972972972973,0.2297297297297294,6.65,10.27,4.23,0.73,2.02,17.2,47.1,35.24,13.7,4.45,19.6631,0.62,-0.0,1 |
| 0.2456140350877192,0.17543859649122806,0.43859649122807015,0.14035087719298245,18.85,29.25,3.21,0.81,2.16,17.2,49.1,46.23,13.8,4.43,14.9521,0.8,-0.08,1 |
| 0.389830584745763,0.1359322033898305,0.33898305084745756,0.1359322033898305,9.78,25.10,4.24,0.81,1.39,15.2,26.1,33.29,5.3,3.44,14.3411,0.22,-0.06,1 |
| 0.08,0.08,0.76,0.08,14.65,23.12,3.21,0.54,1.41,16.2,35.2,35.2,21.7,3.4,4.44,14.8853,0.84,-0.02,1 |
| 0.425531914893617,0.02127659574468085,0.3191489361702128,0.23404255319148937,15.86,18.25,2.21,0.83,1.38,17.2,4.22,1.12,21.15,7,4.44,14.9429,0.48,-0.02,1 |
| 3 - 2 - SGA |
| 0.08108108108108109,0.1891891891891891892,0.4864868468846865,0.24324324324324326,7.67,28.17,2.24,0.78,1.36,16.2,35.1,57.24,12.7,4.44,14.5834,0.1,-0.03,1 |
| 0.6285714285714286,0.05714285714285714,0.2571428571428571,0.0571428571428571,15.79,21.16,2.26,0.81,1.22,15.2,3.1,63.16,12.7,4.45,14.9015,0.04,-0.04,1 |
| 0.0740740740740740740,0.33333333333333326,0.5555555555555556,0.037037037037037035,15.65,22.11,4.24,0.97,1.21,16.2,34.0,62.30,12,3.2,4.46,14.8670,0.04,-0.03,1 |
| 0.2857142857142857,0.0816326530612245,0.5102040816326531,0.1224897959183673,28.87,18.16,1.27,0.83,1.18,16,2.42,1.35,24,11,6,4.45,14.9380,0.78,-0.05,1 |
| 0.225,0.025,0.525,0.225,14,82,11.9,1.27,0.83,1.29,2.41,0.8,23.13,8,4.33,14.6553,0.26,-0.14,1 |
| 1 - 3 - IGA |
| 0.09433962264150944,0.09433962264150944,0.4716981132075472,0.339622641509434,22.84,31.25,2.27,0.56,2.23,29.2,48.0.95,30.8,8,4,0.86,14.6927,0.44,-0.02,2 |
| 0.23728813559322035,0.423728813559322035,0.23728813559322035,0.1016949152542373,25.68,25.25,3.15,0.69,2.33,34,2.18,2.37,24.8,2,1.86,14.9623,0.79,-0.01,1 |
| 0.05128205128205128,0.3076923076923077,0.3846153846153847,0.2564102564102564,14,70,35.20,4.11,0.59,2.29,30.2,21,1.86,28,7,2,0.86,14.6602,0.54,-0.01,2 |