# Using a Genetic Algorithm with Options Data to Forecast Stocks

André Filipe Ferreira Duarte
andre.filipe.ferreira.duarte@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

October 2016

**Abstract**

The possibility of profit coming from financial markets motivates a lot of efforts in financial forecasting. Many researches try to find information from big amounts of data, because access to data are became increasingly easier. This paper presents an innovate solution to make market forecasting. The system built applies technical indicators to options data in order to predict future market movements in the stocks market. We use a two steps optimization to find the best trading rules that leads to higher returns. Each of this two steps has a Genetic Algorithm optimizing trading rules during a training period. After that, the rules are tested and compared to index performance and funds benchmarks. Our solution has better performances than all of them. In a year when index has 0.70 % of loss, our system achieves 3.52 % of return. **Keywords:** Genetic Algorithm, Options Markets, Financial Forecasting, Stocks Market, VIX.

## Introduction

Nowadays, the human being tries to understand and predict almost everything. To achieve that, people use computers to process models, create simulations, among others. The same thing happens in financial context. The investment returns and emotions given by risk, together with the desire of knowledge, characteristic of the human being, motivate people to try understand markets and improve investment techniques.

Some of those techniques are too complex to be processed by human brains. This is the reason why computers are used in order to help humans in the prediction of financial markets movements.

The devices are becoming more powerful while algorithms and programming languages are improving too. Nowadays, everyone can have a powerful computer at home and a fast access to information provided by Internet. So, we have the resources that we need to get the data from financial markets and explore its contained informations.

Our main objective is to have the highest profit with the lowest risk.

An important aspect to profit in financial markets is get the knowledge as early as possible. So, we should get the information from the place where it appears first, at least where we can get it first. Multiple studies proved that options markets have more information and ahead information than stocks markets, because of the higher leverage and the reduce of risk offered by such instruments. This fact attracts the most important and informed traders.

It is wise to think that one single method does not make good predictions in all economic environments, but our application should be adaptive enough to have good results in different financial environments. However, politic events, news, traders speculation and many more will make the applications work more difficult.

The main objective of this work is develop an application, as good as possible, that provides to investors accurate information about future market movements, in profitable stocks of Standard & Poor's 500 (S&P 500) index. The application should use an algorithm to search for the best trading model and the best stocks to invest.

Find good ways to discover information within options flows is an important aspect of this work, because this information will give us guidelines of future market's movements.

The document is organized as follows:

- Chapter 2: explains some basic financial ideas. Some approaches of evolutionary computation are provided. At the end of section, it is also given an overview about different researches with positive results.
- Chapter 3: illustrates the solution's architecture of the application. Each layer and module are explained deeply.
- Chapter 4: presents the evaluation method and discusses the relevant case studies.

- Chapter 5: summarizes the document with a set of conclusions and proposes future work improvements.

**Related Work**

In this section, it is presented the background concepts used in our system and the state-of-the-art of different machine learning approaches used in finance forecast.

Options

An option is a financial derivative that represents a legal contract sold by one party to another party. The option gives the right, but not the obligation, to buy or sell a financial product, as known as option underlying asset.

This contract specifies a price to exercise right to buy or sell the financial asset called strike price. Options contract has a period while it is valid, that period ends at expiration date. In case that, an option is not exercised, after expiration date the contract has no value and no longer exists.

An option can be a call option, which is a right to buy an asset, or a put option, which is a right to sell an asset, both can be executed by the strike price before or on expiration date. The buyer of a call option expects that the corresponding stock price goes up, because he keeps the right to buy the asset by a price below the market price. The buyer of a put option expects that corresponding stock price goes down, because he keeps the right to sell the asset by a price above the market price.

VIX

CBOE Volatility Index (VIX) is a benchmark for future expected volatility, based on the S&P 500 Index.

VIX is the volatility index of Chicago Board Options Exchange (CBOE), it is a volatility index comprised of options rather than stocks. It is particularly relevant for investors because it shows a strong negative correlation with S&P 500 stock returns [3].

Black-Scholes Model

The Black-Scholes model is perhaps the world's most well-known option pricing model. It is used to calculate the theoretical price of options.

Implied volatilities can be obtained by inverting the Black-Scholes formula [6].

Market Analysis

The methods used to analyze market and to make investment decisions fall into two approaches: fundamental analysis and technical analysis. They have the same goal, but fundamental analysis involves analyzing the characteristics of a company in order to estimate its value, instead of technical analysis that does not care a bit about the value of a company or a commodity. Technical analysts are only interested in the market price movements. Technical Analysis is based on three premises [8]:

- Market action discounts everything.
- Prices move in trends.
- History repeats itself.

**Technical Indicators**

Technical indicators are mathematical calculations applied in technical analysis. They are used to predict the future price levels, or simply the general price direction. They are based on historic prices, volume, open interest information, among others. Two examples of technical indicator will be explained bellow.

The Relative Strength Index (RSI) is technical momentum indicator that compares the magnitude of recent gains to the recent losses with the attempt of determine overbought and oversold conditions of an asset. RSI works best for options on individual stocks instead of indexes, because stocks demonstrate the overbought and oversold condition more frequently than indexes [11].

Moving Average (MA) is a widely used indicator in technical analysis [8]. It allows to remove random price fluctuations that helps smooth out price action. MA is a trend-following indicator because its based on past prices. Many indicators are based on MA like: Moving Average Convergence Divergence (MACD), Exponential Moving Average (EMA) and Bollinger Bands.

Genetic Algorithms

Genetic Algorithms are a field of artificial intelligence, where inspired by Darwin's evolution theory, an algorithm solves problems of optimization and search. This technique is all based on biological evolution mechanisms. The GA has a set of individuals called population, each individual can represent an investment strategy. The most used genetic operators are selection, crossover and mutation. The figure 1 illustrates the normal flow of an Genetic Algorithm (GA).

Existing Solutions

The studies indicate that options markets have information about future movements of stocks [6].

Black (1975) in [1] indicates that options has more information than stocks because that instruments offer higher leverage than stocks, what attracts more informed traders. Authors find that options signals constructed from deep out-of-the-money options, which are highly leverage contracts, exhibit the greatest level of predictability. Whereas the signals from contracts with low leverage provide

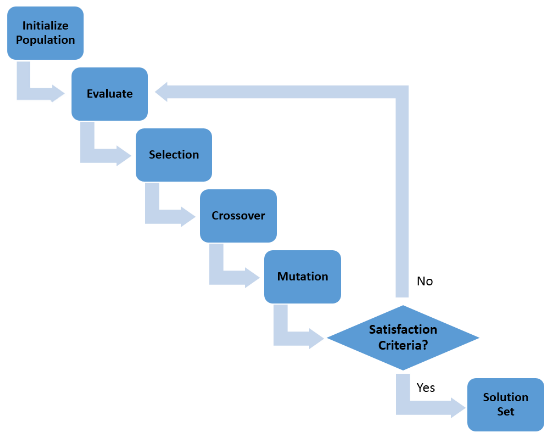

Figure 1: Genetic Algorithm normal flow.

very little, if any, predictability. Some years later, in [9], this assumption is proved too.

Several indicators can be taken out from options data and they provide important information about future trends. Financial studies proved that options volume has information about future stock prices [9]. In [9], ratios derived from options data like Put-Call ratios or Option-to-Stock-Volume ratio are used to prove that. Put-Call Ratio is calculated dividing put volume by call volume, this indicator is used to gauge market sentiment. Option-to-Stock-Volume ratio is calculated dividing total options volume by total stocks volume.

Another approach to find correlations between options market and stocks market is done in [5], where, instead of compare volume-to-volume or price-to-price, they compare option volume and stock prices. It was shown that the particular option volumes leads stock prices and that the negative options signals are stronger in significance than positive signals.

In solution described in [10], authors combined VIX with MACD and RSI, two technical indicators, as inputs of a Support Vector Machine (SVM) to forecast the S&P 500 index. Outputs of the SVM are the up or down movements expected to the following week and its degree of set membership, that represents degree of truth. The results obtained are compared with another SVM with all inputs except VIX and with Buy and Hold strategy. The model with VIX as input produce better results than the other two analyzed strategies, particularly in bearish movements.

In order to obtain as much information as possible from options data, in [2] is compared the predictability power of two option ratios across multiple time levels. The two options ratios analyzed are Put-Call Ratio and Option-to-Stock Volume Ratio. The volumes used in [2] to calculate the Put-Call ratio are buyer initiated put volume scaled by total option volume. Put options are used to hedge against market weakness or bet on a decline while Call options are used to hedge against market strength or bet on advance. So, their relation may have a correlation with market's future movements.

In [4] authors used another approach: Extended Classifier System (XCS). In this system, a rule based model is applied to forecast markets based on contrary sentiment indicators like VIX, Put-Call ratios and a trading indicator called Traders Index (TRIN). TRIN is a technical analysis indicator based on stocks that closed up, stocks that closed down and volumes of stocks. XCS is a new version of Learning Classifier System (LCS).

The next page starts with table 1. This tables shows keys aspects of some existing approaches.

## Solution's Architecture

In this section, it is provided a description of the solution implemented to answer the problem of forecasting stock market movements based in options data. The most remarkable part of the solution is the optimization made by two Genetic Algorithms solving each one a part of the problem. We will give more emphasis to the changes made to solutions explained in section 2. All indicators are based in options and VIX data and the simulation is made with the S&P 500 stocks market data.

### Overall Architecture

The application has three major layers, commonly used: Data Layer, Logic Layer and Presentation Layer. Each layer has well defined responsibilities and is composed by modules performing its job. The objective of such a strong organization is that we can change methods inside a layer without any more changes in other elements. The responsibilities and modules of each layer are presented below:

- Data Layer: Responsible in getting data either from the Internet or from local storage, clean files, calculate derived data and export information. Its modules are: Download/Storage Module, Data Clean and Fix Module and Indicator Calculation Module.
- Logic Layer: Responsible to find the best trading rules and test them. It is composed by the following modules: Optimization Module and Investment Simulation Module
- Presentation Layer: Responsible by the interface provided to the user. It allows to set the application with a text file and shows results and execution statistics. The two modules performing this tasks are Input Module and Output Module.

General architecture is present in the figure below, figure 2, with the three layers and respective

| Paper | Area of Research | Approach | Year | Data | Heuristics | Metrics | Return/ Confidance Rate/ Results |
|---|---|---|---|---|---|---|---|
| *4* | Engineering | Extended Classifier System | 2005 | S&P 500 Futures | VIX, Put/Call ratio, TRIN | Accumulated profit, Mean Profit, St. Dev. Win/Lose trades and Accuracy rate | 01/1997 to 12/2004 |
| *2* | Finance | Univariate and Multivariate tests | 2014 | Options | Put/Call ratio, Option-to-Stock Volume Ratio | Slope Coefficients and t-statistics | Daily - P/C ratio Weekly - O/S ratio Monthly - O/S ratio |
| 8 | Finance | Multiple time series regression | 1998 | CBOE Options | Option Volume and Stock Prices | Likelihood ratio test | Particular option volumes carry information |
| 7 | Finance | C4.5, Backwards Stepwise Logistic Regression and Neural Networks | 2004 | U. S. Options | Put/Call ratio and Strike Price | Percent of Positives | All three algorithms produce lift over random |

Table 1: Summary of existing solutions.

modules in evidence.

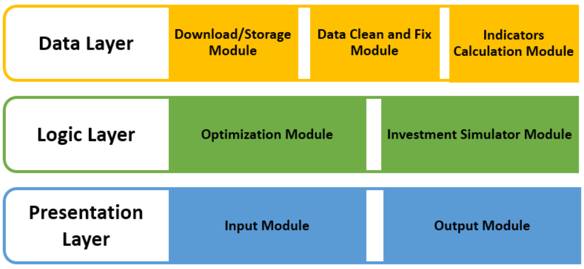


Figure 2: System Architecture divided by layers.

Data Layer
Data layer needs to get data from many sources because we need data about options, stocks and other small pieces of data to use in some formulas. The three modules of this layer will request and provide data between them. They communicate using .CSV files. Each module manipulates and creates new files that will be accessed by other modules even from other layers.

The Download/Storage Module, as the name suggests, downloads the information needed from web and stores the data locally. It was the first implemented module, because we need to collect data before any other task.

Our initial options dataset has end of day historical option prices for all U.S. Equity options including stocks, Indexes and ETFs. This dataset was huge. So, we create Data Clean and Fix Module to remove all options contracts except 99 companies belonging to S&P 500 index and index symbol. The ticker symbol for the S&P 500 index is GSPC. However, the ticker symbol for S&P 500 index in CBOE options exchange is SPX. When we started analyzing the reduced dataset, we found some errors. The most typical errors was duplicated lines and wrongs expiration dates. This modules fixes data mistakes and removes unneeded data.

The Indicators Calculation Module has a big importance in terms of data analysis of our application. This happens because here we define formulas to correlate information between options and stocks. We calculate some indicators learned from the research made inside financial forecasting studies and we create some new ones. Every indicator was calculated to 3 time spans. This is, all contracts were separated into one of the following time intervals. First time span contains all options that expires until next standard option expiration date, this is normally the third Friday of the month. Second time span contains all options that expires from the next option standard expiration date to the expiration date of the following month. Lastly, the third time interval includes options that expires after the expiration date of the following month. Within our application, we just use the first time span, because it provide short time forecasting and it is the time span with more volume. The indicators used are the following:

- Put-Call ratio: was cited in section 2. Here, it is presented in the way that it will be used. This simple ratio just uses the sum of all puts and the sum of all calls of each time span. The ratio value is the number of puts divided by the number of calls. Usually, high levels are associated to bearish markets, because traders are buying puts in order to buy protection.

- Option-to-Stock Volume Ratio: is another relation presented in section 2. This ratio relates stocks volumes with options volumes. It does not separate calls from puts. Option-to-Stock Volume Ratio are related to forecast bearish markets.

- Delta Price Ratio: the objective of this ratio is to compare how much traders are paying for protection or how much they are investing believing in a future rise. For this, we use the extrinsic option value, this value is attributable to time until its expiration. It adds potential value for the options based on future moves of the stock value [7]. We multiply extrinsic option value by volume to give importance to most traded contracts. It is calculated for calls and puts separately.

- Theoretical-Market Price Ratio: there is a formula created by financial researchers to price options, the Black-Scholes formula. It is referenced in section 2.3. If the difference between the theoretical option price and the value that the contract is being traded become too big, probably some information is embedded here. This ratio, like Delta Price Ratio, must be calculated separately for calls and puts.

An indicator value alone with just one day data can not be enough to do a good option market study. We decide to apply RSI and the 10-day EMA to each of the previous ratios in order to have indicator's information about the last days. RSI is a momentum indicator that compares recent gains and losses. EMA is a special case of a MA that gives more importance to recent prices.

VIX was explained in subsection 2.2. It is calculated using a range of options and it is known as the investor fear measure. We believe that VIX could contribute in our market analysis. We apply the same technical indicators to VIX, that we applied to ratios in order to retract its recent evolution

Logic Layer

Logic layer is responsible for optimization and simulation tasks. The better the modules from logic layer apply information, the better the overall application results. The optimization is made by two Genetic Algorithms and simulation is made using trading rules found by such algorithms. Information discovered in that layer is saved into .CSV files as before.

We separate the optimization into two stages: optimization of indicators and optimization of investment parameters. With this, we remove some degrees of freedom and we separate different parts of the trading strategy. The score of each chromosome is calculated in this way. The fitness function takes into consideration three aspects to evaluate each gene: next day returns, next week returns and next week volatility.

Besides the ratios calculated by Indicator Calculation Module, we have three more genes:

- Prediction Gap: represents the time needed to information from options market incorporates stocks market. It is the time between discovering in the options market and investment inside stocks market. This value is limited to a maximum value of 10.

- Similarity Level: is a value that represents maximum level of difference allowed between current indicator's level and the best values to invest. The best values to invest are the values found by first step of the GA for each indicator.

- Buy Minimum: is a limit value that must be reached to give an order to buy.

- Sell Minimum: is a limit value that must be reached to give an order to sell, if, we were in the market.

The following figures 3 and 4 represent the composition of the chromosome used in each step of the optimization.

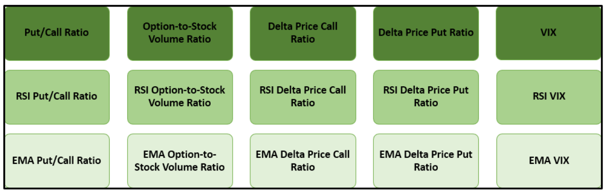


Figure 3: First step optimization chromosome.


Figure 4: Second step optimization chromosome.

We have some doubts about the importance of the gene Prediction Gap. So, we create a chromosome without the Prediction Gap gene to be evaluated later.

The Optimization Module applies basic genetic operators to create new generations. The crossover operator selects two individuals and randomly it choices two points to swapped them. The mutation is equivalent to change a specific gene. The method used to select individuals for breeding the next generation is known as tournament. In this selection method the selected individuals are the winners of each tournament.

The Investment Simulation Module, as the name suggests, tests the profit of our application in a given testing period. It uses a complete different time span than the training period, naturally. It uses the values obtained in optimization module. Commissions are applied in every trade. The main output is the average Return on Investment (ROI).

Presentation Layer

Presentation Layer provides the interaction between the user and the application. It allows to parameterize our solution with a configuration file and it shows the results of the execution to the user. This layer is composed by two modules, one that deals with input to application and other that deals with output.

Data Flow

In respect to the data flow inside our application, we have three sources of data to feed the system: options data, VIX data and S&P 500 stocks data. After that, we have the following steps:

- Raw Data - Options data and VIX data are inserted into Indicator Calculation Module. This module calculates all ratios and derivatives.

- Ratio Values - All ratios data are inserted into first GA. This algorithm finds the best value of each ratio. To perform that, the algorithm interacts with a fitness function that has all stocks data.

- Best Ratio Values - First GA sends best ratio values to second algorithm. This one will find the proximity needed between best values and current values to maximize profit. It uses also a fitness function that has all stocks data.

- Best Ratio Values and Invest Parameters - Best Ratio Values and Invest Parameters compose the best trading rule. In other words, it is composed by: best value of each ratio, similarity level, Buy Minimum and Sell Minimum. Trading rule is valid to buy or sell stocks.

- Comparison between Trading Rules - Best Trading Rule is compared to today's market situation. Today's market situation is calculated this way. We build ratios with today's data. After that, we compare each today's ratio with best ratio value, if the proximity is below the Similarity Level GA, we add its weight to today's grade.

- Buy/Sell Order - If market situation satisfies the best trading rule an order is given. This order can be sell, if we are in the market, or buy, if we are out of the market.

The figure 5 illustrates all data flow inside our application.

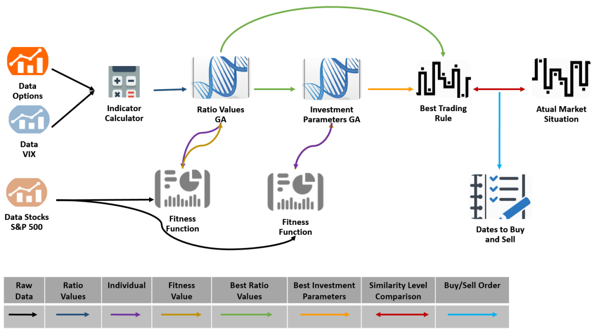


Figure 5: Information flow inside the application.

**Results**

In this section, it is described the validation approach to evaluate the defined system. This validation is done simulating trades during the defined testing period for all 100 stocks belonging to the list.

Algorithm Quality Measures

Our primary objective is get the highest profit. To measure that, we use ROI, because is one of the most important profitability ratios. ROI just represents the benefit obtained from an investment. The ROI formula is presented in equation 1:

$$ROI = \frac{Gain - Cost}{Cost} \tag{1}$$

where:
*Cost* is cost of the initial investment.
*Gain* is the winnings from the investment.

Another measure that will be used, in order to characterize a strategy, is Maximum Drawdown. It will be used to evaluate our system in terms of risk. The percentage of the Maximum Drawdown is given by equation 2:

$$MaximumDrawdown = \frac{Peak - Lowest}{Peak} \tag{2}$$

where:
*Peak* is the highest value before largest drop.
*Lowest* is the minimum value before new high be established.

Case Studies

In this section, we will present two relevant case studies from the several tests made. The first case study will compare different derivations of our system and will compare our system with the S&P 500 index and some fund benchmarks. We will append to results some explanations for the obtained performances. The second case study compares results obtained by systems with different genetic operations. Two mutation methods will be compared.

The following points must not be forget when reviewing the case studies:

- Every trade has commissions applied.
- We assume that all companies have the same weight, however the index value is calculated based on the market cap. Market cap is, basically, the number of shares multiplied by its current price.
- The dividends are completely ignored in our tests.
- We have a specific initial budget for each company and we can open a position to another company without close any position. So, the maximum value that we could have invested is 100 times the initial budget.
- When we open a position, we use all budget available from previous trade of the same company.
- The comparison charts represent average ROI evolution. It is calculated every day for companies with open positions.
- We close all opened positions in last day of test period. This could cause impact on our ROI, because we are selling without application's order.

### Case Study I - Training between 2011 and 2014, Test in 2015

The present case study shows the results of an simulation done with 4 years of training period and one year to test the trading rules obtained in that time. The training period starts on the 1st business day in 2011 and finishes on the last business day in 2014. The testing period starts on 1st business day in 2015 and finishes on the last business day in 2015. The system approaches that uses Prediction Gap do not make investments that exceeds the last business day of 2015.

Our first objective is find out if VIX helps us to get better results. The image 6 compares the ROI evolution during test period, between one derivation with VIX (blue) and another without (red).

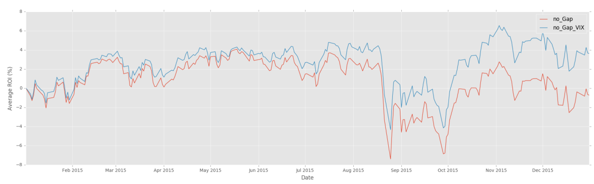

Figure 6: Comparison between a solution with VIX and other without VIX.

The derivation with better return is known as: No_Gap_VIX, blue line in the chart. This system uses glsVIX that helps to mitigate losses. In last period of the year, when the markets had a bearish time the solution with VIX reduced the number of positions more than other solution. While in the first part of the year the returns were similar, from mid-August until October, when the S&P 500 has big losses VIX made difference. Finally, due to reinvestment of capital, the solution without VIX can not achieve the same result. Due to its big losses, when the market is raising this strategy do not have same budget to recover.

As mentioned above, in section 2.2, VIX measures investor's feelings. When market has unstable periods, it raises in value and close positions is the better solution to avoid losses. The solution with glsVIX has a ROI of 3.52 %, on the other hand, the solution without VIX has a ROI of -0.83%.

Like we said before, we have some doubts about the importance of the gene Prediction Gap. So, we tested our system with and without this gene.

Prediction Gap gene measures time between get and apply information, it is explained in subsection 3.3. In figure 7, we compare two system derivations. The only difference between them is the use of chromosome in the step 2 of the GA, one approach has the Prediction Gap gene and other does not.

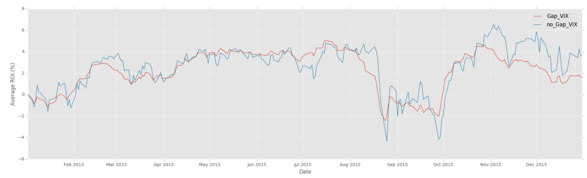

Figure 7: Comparison between a solution with Prediction Gap and other without Prediction Gap.

The solution called with Prediction Gap (red) has less return than the solution that does not have this gene (blue). The solution with Prediction Gap has, on the final period of the year, the ROI of 1.49 %. Even being a positive result, it is worst than the result obtained by the solution without Prediction Gap that have 3.52% of ROI.

The Genetic Algorithms have tendency to converge towards local maximums, that is, instead of find a consistent value to the problem that are solving, they find a local solution that fits good the training period. When we analyze the Prediction Gap values for each element of the 100 list, we find values from 0 to 10 (maximum value allowed). If, we carry that values to the finance field, we can not find a plausible reason for this fact. The Prediction Gap should exist, but end of the day data is not enough to find its value. This dataset limits us to integer days, when the real value could be hours or even minutes.

Now, we compare our best solution, with other benchmarks. The system is compared with the following investment benchmarks or indexes:

- S&P 500 index - This is the index. Index value is calculated based on market capitalization of each component. The dividends are ignored, because we ignore dividends in our application too.
- HFRX Absolute Return Index - This is designed to be representative of the overall composition of the hedge fund universe. Hedge funds explore unique investment strategies and seek to obtain absolute returns.
- iBoxx USD Treasuries - Index with minimum risk. The index represents US treasury bonds with a maturity of more than 10 years.

In figure 8, each curve represents the return of investment achieved by the respective investment:

Our application, named no_Gap_VIX, obtains better ROI when compared with funds and indexes. This result is achieved mainly due to the limited losses during the period between August and October. Except the Hedge Fund Absolute Return Index (HFRX) that has a consistent low rise during all year of 2015, all other indexes have more losses than our application in that period.

The return of our best system is always above the S&P 500 index return. This is important because it is the market where we are investing. As said before, we have commissions applied which decrease the performance of our system. The system performance during bear markets, even not being extraordinary, is very good, because the training period does not have similar periods. So, the system adapts well to new market situations.

The lower Maximum Drawdown is obtained by HFRX - Absolute return index. The index selects constituents which characteristically exhibit lower volatilities and lower correlations. The objective of this is reduce the risk. The 1.26 of Maximum Drawdown is far better than the value obtained in other benchmarks. Closely linked to this are the value of the ROI's Standard Deviation. This value is also low, because of the little volatility in index returns.

The table 2 summarizes information about benchmark ratings.

| **System/Index** | **ROI** | **ROI St. Dev.** | **Max. Drawdown** |
|---|---|---|---|
| *Our system* | 3.52 | 2.05 | 9.12 |
| *SP 500 index* | -0.70 | 2.67 | 12.35 |
| *HFRX Absolute Return* | 2.85 | 0.92 | 1.26 |
| *iBoxx USD Treasuries* | -1.00 | 3.15 | 13.79 |

Table 2: Comparison between our best solution and other benchmarks.

## Case Study II - Comparison between mutations methods

The objective of this case study is compare the results of our best system when we change the genetic operators. We opted by change the way how mutation is done inside optimization module.

Gaussian mutation is now used instead of the normal one, to compare the systems. Gaussian mutation adds a value to the gene who should be changed. The normal just change the gene to a random value. The mutation receives two values: the mean value and the Standard Deviation. This two values define the Gaussian Distribution. The value is obtained based on the probability density of the defined Gaussian Distribution.

The two systems will be evaluated just changing mutation methods. The training and testing periods will be different than in first case study. Now, they will be between 2011 and 2013 for training period and between 2014 and 2015 for testing period.

The figure 9 shows the evolution of ROI during the testing period.

The two systems have very similar results during the first half of the testing period. However, in the end of the period the system that has the Gaussian Mutation obtains more returns. Even the losses in bear market are similar.

The table 3 present the statistics of each system:

| **System** | **ROI Avg.** | **ROI St. Dev.** | **Max. Drawdown** |
|---|---|---|---|
| *No Gap VIX - Normal Mutation* | 7.29 | 3.48 | 9.53 |
| *No Gap VIX - Gaussian Mutation* | 8.86 | 3.59 | 8.71 |

Table 3: Execution statistics between system with different mutations.

The ROI achieved by the GA with the new mutation is about 1.5% higher than the system with the old mutation method.

Besides that, the system No_Gap_VIX - Gaussian Mutation has a smaller value of Maximum Drawdown, which represents lower risk involved. It is an expected result.

Gaussian Mutation tends to mutate genes to val-

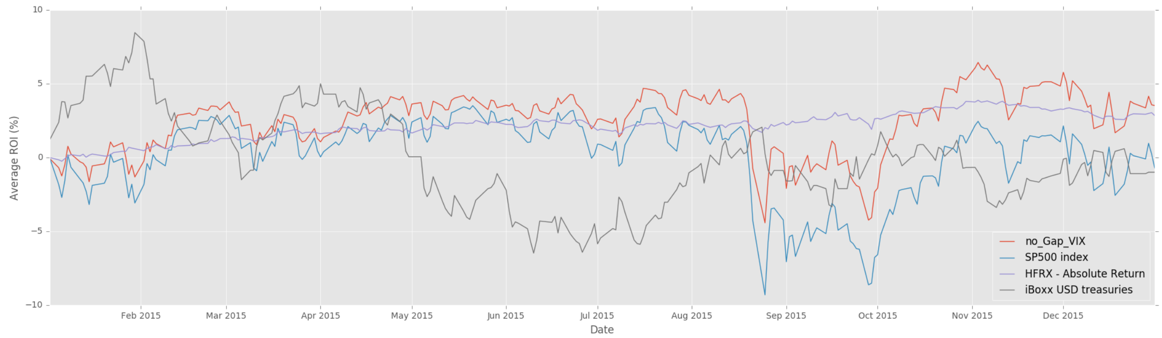


Figure 8: Comparison between our best solution and benchmarks.

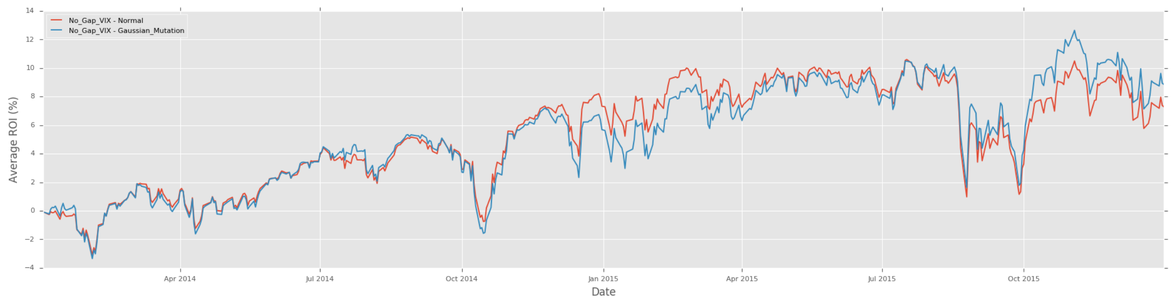


Figure 9: Comparison between different mutation methods.

ues near the mean, because the probabilities of mutate to values near the mean are higher. So, the chromosomes have more probabilities to have values near the mean. With values near the mean, the possibility of make big mistakes is lower and the Maximum Drawdown is lower two.

## Conclusions

The objective of the presented work was to forecast stocks based in options data. The system analyses a list of 100 companies and suggests when the user should buy or sell stocks. The objective was to overcome the returns of the index where we are investing. We compare our system with fund benchmarks. Our finds are exciting, but much more improvements can be made to get better the results.

The main conclusion of this application was that it is possible to have considerable returns from stocks markets using information contained in options markets. We build indicators to extract information from option data. They are useful to forecast stocks market movements.

We will present some suggestions for new features of the application. Besides that, the limitations that was not explored will be listed:

- The exploration of option data could be improved with more indicators. The bid and ask prices are not focused.
- The chromosome can have different genes. For instance, use different number of periods in EMA could tested.
- The GA has a fixed number of generations. The termination criteria, that stops the evolution process can be improved. Can be used a minimum fitness grade to stop the algorithm or the algorithm stops when no evolution were observed.
- We observed that different mutations lead to different returns. Other selection and crossover operations can be created to improve the evolution process.

## References


[1] F. Black. Fact and fantasy in the use of options. *Financial Analysts Journal*, 31:1–36, 61–72, 1975.

[2] N. Blau, B.M., Nguyen and R. Whitby. The information content of option ratios. *The Journal Of Finance*, 43:179–187, 2014.

[3] CBOE. The CBOE Volatility Index - VIX. Accessed 15-Sep-2016.

[4] A. Chen and Y. Chang. Using extended classifier system to forecast s&p futures based

on contrary sentiment indicators. *Evolutionary Computation, The 2005 IEEE Congress*, 3:2084–2090, 2005.

[5] M. Easley, D., O'Hara and S. Srinivas. Option volume and stock prices: Evidence on where informed traders trade. *The Journal Of Finance*, 53:431–465, 1998.

[6] R. Elliott, P. Chang, W. Jiang, and H. Zhang. A tale of two tails: Using options to forecast the movement of stocks. *Computational Intelligence and Software Engineering. IEEE*, 2010.

[7] G. Fontanills. *Trading Options For Dummies*. Wiley Publishing, Inc., 2008.

[8] J. Murphy. *Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications*. New York, Institute of Finance, 1999.

[9] J. Pan and A. Poteshman. The information in option volume for future. *Review of Financial Studies*, 19:871–908, 2006.

[10] R. Rosillo, J. Giner, and D. Fuente. The effectiveness of the combined use of vix and support vector machines on the prediction of sp500. *Neural Computing and Applications*, 25:321–332, 2014.

[11] S. Seth. The Top Technical Indicators For Options Trading. Accessed 2016-09-16.