Investment Strategy based on Volatility and Trends detection optimized by a Genetic Algorithm

Daniel Vieira Silva
Instituto Superior Técnico / IT
University of Lisbon
daniel.vieira.silva@tecnico.ulisboa.pt

***Abstract*— Financial markets regularly experience periods of high and low volatility, and while high volatility is associated with higher risk, it is not always negative to trade during such periods. However, if the goal is to have a safe and reliable strategy, it is advisable to avoid these more volatile periods. This thesis aims to create a trading strategy, applied to the stock market, based on volatility and market trends, being able to avoid the most volatile periods, generating high returns with low risk. The proposed system combined a strategy that identifies the quietest periods in the market, a Support Vector Machine (SVM) that identifies market trends and a Genetic Algorithm (GA) responsible for optimizing the entire system. The system uses GA to optimize and select technical indicators as well as to choose the hyperparameters used in SVM. In parallel, the GA is also responsible for setting the decision weights and margins of the volatility strategy. The proposed system was then evaluated in seven financial markets with different characteristics. The results show that this approach yields better results than the Buy&Hold strategy in most tested markets. In the S&P500 index, the proposed 2:1 leverage system achieves a return rate of 20.28% with a Maximum Drawdown (MDD) of 4.34%, while the Buy&Hold strategy achieved a return rate of 17.34% with an MDD of 15.07%.**

***Index Terms*— Financial markets, Volatility, Trading strategy, Support Vector Machine (SVM), Genetic Algorithm (GA)**

I. INTRODUCTION

The financial market, particularly the stock market, has great interest to investors because it allows the opportunity to invest capital and have a large margin of profit in the stock market. Monetary motivation has been one of the main drivers of all the research that has been carried out for several years by financial companies, the academic community or simply enthusiasts in the field.

Traditionally, the methods used to analyze stocks and make investment decisions fall into two main categories, fundamental analysis, and technical analysis. Fundamental analysis involves the study of various economic factors of a particular asset, such as income and expenses, market position, annual growth rates, among others [1]. On the other hand, technical analysis uses only the prices of an asset to study and identify patterns and fluctuations in its graphs and to calculate several variables that help predict future price movements [1].

The Efficient Market Hypothesis (EMH) [2] indicates that due to random market behavior it is impossible to predict the market. However, some research has shown that it is possible to predict future market behaviors using artificial intelligence (AI) models. These intelligent systems are designed to develop investment strategies that can analyze more data, generate more financial return and make investment decisions in a shorter timeframe than a human can achieve [3]. Many types of Machine Learning (ML) algorithms have been used to predict future market behaviors, such as Artificial Neural Network (ANN) [4], Decision Trees (DT) [5] and Support Vector Machine (SVM). [6] In addition to the financial return of a given strategy, it is necessary to analyze the risk relative with that strategy, and the risk is usually associated with the price volatility of a given financial asset. Following the 2008 global financial crisis, low volatility investments gained great importance within the investor community [7]. In the stock market, huge amounts of capital are traded per day and most investors seek huge returns, and the risk assessment of investments is often undervalued compared to potential returns. This way, they end up investing in more volatile stocks because they think that they get better returns in a shorter period. However, it is important to have realistic expectations about the performance of an investment strategy to avoid future disappointment when sharp market downturns occur, such as the 2008 crisis [8].

The main motivation of this master's thesis is to be able to create an automatic investment strategy, through machine learning and signal processing algorithms, able to detect and avoid being in the market when it becomes more volatile since volatility is associated with unpredictability and risk. To address this problem, an approach has been implemented that combines an SVM, which predicts market trends, and a GA, which identifies periods of low volatility and optimizes system-wide parameters. This system then aims to generate long positions by optimizing the risk-return ratio in the S&P500 index and some of its companies.

The main contributions for this work are:

- Signal characterization (prices), to identify the least volatile periods of the market through the creation of technical volatility indicators.
- Creation of a GA fitness function that combines return, drawdown and system accuracy.
- Process of assigning labels to SVM input data, by moving the average price, classifying market trends.

II. Background and State-of-the-art

This section provides crucial concepts about Stocks Market and an overview of existing solutions regarding several areas within Machine Learning and Evolutionary Computation.

A. Financial Concepts

The stock market is a broad term that includes a subset, on different stock exchanges, on which corporate stocks are publicly traded. There are many different stock exchanges, the most important and best known being the US National Association of Securities Dealers Automated Quotations (Nasdaq) and the New York Stock Exchange (NYSE). In addition to stock exchanges, there are also stock indexes. A stock market index measures the change in selected stock prices, allows you to describe the general financial market and compare the return on specific investments. The top three US indices are the Standard & Poor's 500 Index (SPX), the Dow Jones Industrial Average (DJIA) and the Nasdaq Composite Index (COMP). The Standard & Poor's 500 Index better known as the “S&P 500” is world renowned and brings together the top 500 US-based market capitalization companies.

B. Technical Analysis

Technical analysis is the tool that through the study of past market data, price and volume, aims to identify the future direction of prices of a financial asset. This analysis is performed using technical indicators that use past changes in asset prices to predict futures. However, the results obtained by the technical indicators do not always lead to the right decisions in the market.

Indicators are usually divided, according to the information they give us, into 4 types: Trend, *Momentum*, Volatility, and Volume. Trend indicators indicate whether there is a trend in the market, also showing its direction. It is easy to see that trend indicators are widely used as they make it possible to decide on good market entry points at the beginning of a trend. *Momentum* indicators are intended to characterize market trends, in other words, indicate when a trend is strong or weak. Otherwise, this type of indicator can be said to show when an asset is overvalued or undervalued. Volatility indicators use asset prices to indicate how much the price is changing over a given period. Volume indicators use volume changes to obtain market information such as the confirmation of trends.

C. Leverage

Leverage is a tool that allows you to improve return while maintaining your initial investment, but it is also highly risky. This is a strategy that uses borrowed capital as a source of finance to invest in financial assets [9]. For example, in the case of a 2:1 leverage, if you make an investment of 100,000 euros and after some time you get a 5% profit, without the financial return is 105,000 euros, while with leverage the return will be 110,000 euros. However, leverage carries a greater risk for investors. In this same example, if after some time instead of having the 5% profit, had 5% loss, with leverage of 2:1 the investor would owe 10,000 euros to the broker.

To try to mitigate this risk there are a few strategies that can be used such as using Stop Limits to avoid downfalls or as indicated in [10] having an investment portfolio that counteracts the risk of that investment.

D. Works in the Financial Markets

There are numerous studies developed that use ML algorithms in various financial areas, such as forecasting trends and market directions. In [11] a supervised feed-forward NN with backpropagation learning was used to learn patterns in Japanese candlesticks charts. In this study, two approaches were analyzed. The first approach was to use raw input data of Japanese candlesticks, such as open, high, low and close prices. In the second approach, the input dataset is more complex than the previous one, since reversal signals of Japanese candlestick were used, such as Morning star, Inverted hammer, Harami, among others. The model was tested with General Motors stock and yielded surprising results when compared to actual trends in both configurations. However, the results of the first approach are to some extent better.

In the study [12] a comparison was made between four forecasting models, which in this case predicted the direction of stocks and indexes of the Indian stock exchange. The four models compared were Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest (RF) and Naive-Bayes (NB) but in addition to these two approaches to handling input data in these models were also analyzed. The first approach for input data calculates ten technical indicators and only normalizes them between -1 and 1, while the second approach characterizes the values of all technical indicators as -1 and 1 according to the trend, where -1 represents a downtrend and 1 an uptrend. The evaluation was conducted over a 10-year period from 2003 to 2012 on two stocks, Reliance Industries, and Infosys Ltd. and two CNX Nifty and S&P Bombay Stock Exchange (BSE) Sensex indexes. The tests performed in this study show that the first approach to the input data gives the best results using an RF. It was also found that the performance of all these models increased when the second approach of the input data was used, obtaining very similar accuracies among the analyzed models, although the NB had better results.

E. Works with GA and SVM

In this subchapter, it focuses mainly on studies where GA's, SVM's or a system using both algorithms were used. In [13] they used SVM to predict the weekly direction of movement of the NIKKEI 225 index. In addition, this method was compared with a Linear Discriminant Analysis (LDA), a Quadratic Discriminant Analysis (QDA), an Elman Backpropagation Neural Networks. (EBNN) and a model that combines SVM with the other classification methods referred to. The data used are from January 1, 1990 to December 31, 2002. The data was divided into two parts, the first part consisting of 640 observations and is used to train the models, while the second part is used for the testing phase, however it has only 36 observations, making the results obtained insignificant because the testing period is very short. Finally, the authors conclude that the combined model has the highest hit rate, followed by the SVM that outperforms the other classification methods, obtaining respectively 75% and 73%.

In the study [14] they used a Multi-Objective Genetic Algorithm (MOGA) to predict future asset price trends. The investigated model was used to determine possible buy, sell or hold conditions on the stock exchanges to achieve high returns with the lowest possible risk. Model input data, composed of the volatility index (VIX) and other technical indicators, are optimized to find the best investment strategy. The proposed model was tested in various markets using data from key stock indexes such as: NASDAQ, S&P500, FTSE 100, DAX 30 and the NIKKEI 225. The results for the period 2006-2014 show that they clearly outperformed Buy&Hold and Sell & Hold strategies achieving a 10% higher annual return on the NASDAQ and DAX indices.

In [15] they used an SVM, but instead of using technical indicators they use some financial elements of companies to classify companies. The authors used a PCA to avoid direct use of complicated and highly dimensional financial ratios. With PCA, they lower data size, thus improving training accuracy and efficiency, while preserving initial data resources. In order to separate high return stocks from low return stocks, this article selects seven financial ratios from the 2009 and 2010 annual reports of A-share index of Shanghai Stock Exchange companies. The results show that norm-standardization PCA-SVM achieved an accuracy of 75.4464% in training and 61.7925% in test. For further analysis, the study presents a comparison between the cumulative return obtained by the proposed model and the A-share index of Shanghai Stock Exchange, where it can be seen that the PCA-SVM model has a higher cumulative return than the A-share index of Shanghai Stock Exchange.

In study [16], it uses a genetic algorithm to manage a portfolio of financial assets, a problem that has been widely studied in the financial world. The genetic algorithm uses a set of technical indicators to define when transactions should be made. Each technical indicator has a set of investment rules that rank each indicator at 4 different levels. The investment rules of the indicators are combined through a very interesting weighting strategy, which inspired one of the strategies implemented in the system proposed in this thesis. In order to evaluate the strategy implemented, tests were conducted in the period between 2003 and 2009, passing through the great financial crash of 2008. The strategy was compared with various investment methods such as the Buy&Hold strategy and a purely random strategy. The results showed that the genetic algorithm outperformed both the Buy&Hold strategy and the random strategy, and the best iteration of the genetic algorithm obtained, a ROI evaluation metric of 62.95%.

Regarding the importance of the feature selection process in the accuracy of SVM model classification, several studies have been conducted in this area, and part of them have been exploring the use of evolutionary algorithms to select the best features. One of the first works that used this approach was done by Huang and Wang in [17]. The objective of this work was to simultaneously optimize SVM kernel parameters and select features, to improve the SVM performance, and for this, an approach based on a genetic algorithm was presented. This approach was compared with the traditional parameter optimization method, Grid Search, across multiple data sets from various areas. It was concluded that the proposed GA-based approach significantly improves the accuracy of SVM classification by using fewer input features.

In [18] a system consisting of a GA and an SVM for trading in Forex markets was proposed. SVM was used to classify market type and GA to optimize an investment strategy based on technical indicators. The goal was to train three populations for the three types of markets and use them according to the type of market chosen by SVM, thus gaining the advantage of having a strategy appropriate to the type of market. In addition, a strategy has been created to define appropriate leverage, depending on the level of certainty of the forecast. In one of the case studies of this paper, the proposed strategy was compared to the Sell&Hold strategy from 01/02/2015 to 02/02/2016 in the EUR / USD currency pair. The average ROI obtained by the proposed System was 43.9% and the best result obtained by the developed System was 83.5, while Sell&Hold was 72.2%.

III. System Architecture

In this chapter, we will describe and analyze in detail the proposed solution specialized for trading on the S&P500 index.

A. *General Perspective*

The solution consists of two algorithms, an SVM, used to identify trends, and a genetic algorithm, used to optimize SVM parameters, optimize technical indicators, select the best features, and optimize parameters for low volatility period selection. The final decision of the proposed solution corresponds to the union between the SVM decisions and the low volatility period selection strategy decisions.

The developed system is divided into six main modules, the data module, the preprocessing module, the Genetic Algorithm module, the SVM module, the volatility strategy module, ending in the evaluation module.

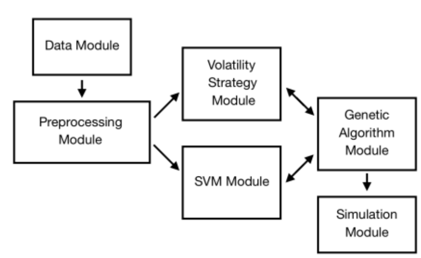

Fig. 1: Main stages of the proposed system.

B. Data Module

The Data Module is responsible for handling financial market data and calculating the technical indicators that will be used in the system.

In this module 10 technical indicators of Trend and *Momentum*, and 4 technical indicators of volatility were calculated using the financial data with hourly prices. The periods of the 10 technical trend and momentum indicators will be optimized by GA. The other 4 volatility indicators have fixed periods and will be used to identify calm, low volatility periods in the Volatility Strategy Module. The indicators used will be named later.

C. Preprocessing Module

Machine learning (ML) algorithms learn from data and so it is crucial to have quality data for the algorithms to perform well. However, most data contain noise, inconsistent, redundant, invalid values or values that are not on the same scale, which according to [19] makes it difficult to learn the algorithm in the training phase. There are several data preprocessing methods that, without damaging valid data, attempt to correct the data, eliminating these problems.

This module is responsible for eliminating invalid data from technical indicators because it uses past information, separating data into training, validation and testing sets and for normalizing technical indicators.

.

D. SVM Module

The SVM module is responsible for calculating classifiers capable of making predictions about the direction of market movements. This module operates in conjunction with the GA module. GA is responsible for finding out which technical indicators provide non-redundant information, their time periods, and optimizing SVM hyperparameters. Technical indicator data is properly labeled and used to train a classifier. This classifier is then used to assign labels to new samples, test sets, or validation sets. The following subsections discuss the SVM module in more detail.

1) Labeling – Vetor Y

When using supervised ML methods, it is necessary to label the input samples of the algorithms. The classification through SVM is no different, it is necessary to label the input samples in different classes. There are several types of labeling methods and system performance that varies depending on the method used. For example, Marcos Lopez de Prado [20] proposed another labeling strategy, called the Triple-Barrier Method. The idea of this method is based on three barriers, an upper barrier, a lower barrier and a vertical barrier. When the upper barrier is reached the assigned label is 1 ("Buy"), when the lower barrier is reached the assigned label is -1 ("Sell") and if none of the upper and lower barriers are reached until the vertical barrier, the label assigned is 0 because it means that the market does not have a definite trend.

In this thesis, another labeling method was developed that uses two classes, with the same importance or weight (unitary). The method used assigns label 1 when the future average price is higher than the current average price. On the other hand, when the future average price is lower than the current average price, label 0 is assigned to the sample associated with this time. This can be represented according to equation 1.

$$y_t = \begin{cases} 1, & se\ SMA24_{t+24} \geq SMA24_t \\ 0, & se\ SMA24_{t+24} < SMA24_t \end{cases} \quad (1)$$

In equation 1, t is the current time, y_t is the label assigned at time *t*, and *SMA24* represents an average of 24 hours (2 days) that is made to the price vector.

Before using this method we used the method that assigns the labels according to the hourly or next day price. The results obtained after the SVM classification were not satisfactory because the stocks have a practically random behavior.

Fig. 2: Difference between the two types of labeling tested.

In Figure 2 we can compare the two labeling methods tested, the next time hour based method (red line) and the SMA based method (green line). The blue line represents the price of an asset over a given period of time and the orange line represents the SMA of the price over a 24 hour period. From this figure it is possible to verify that the proposed labeling method classifies market trends more clearly, avoiding small market fluctuations.

2) Features - Matrix X

The purpose of the Data Module is to generate input features (matrix X) that together with the vector Y allow binary classifiers to train. The data set with Trend and *Momentum* indicators is normalized and then becomes matrix *X*. The technical indicators used in SVM were Simple Moving Average (SMA), Exponential Moving Average (EMA), Weighted Moving Average (WMA), Average Directional Index (ADX), Relative Strength Index (RSI), Triple Exponential Average (TRIX), Momentum (MOM), Williams' %R (WILLR), Commodity Channel Index (CCI) and AROON.

However, as explained later, GA will select the best features (indicators) for SVM so whenever a feature is considered by GA to be redundant, it will be dropped, or in other words the matrix column of that feature will be removed.

3) *Hyperparameters and Kernel*

To optimize SVM model performance, you must adjust its hyperparameters. Usually a search algorithm, Grid Search, is used to perform an exhaustive search of the best parameters, performing various combinations between the hyperparameters, but in this work the optimization of the hyperparameters was performed through a GA.

SVM uses three hyperparameters, where parameter C represents the trade-off between the training sample mismatch penalty and the decision boundary simplicity. Low values of C make the decision boundary simpler (closer to a straight line), while high values lead to a more complex decision surface. Gamma is a parameter for nonlinear hyperplanes. This parameter controls the influence of new samples on the decision margin. The larger the Gamma, the greater the influence of resources on the decision margin. The values of the hyperparameters used in this work are shown in table 1.

Table 1: Hyperparameters used in the SVM model.

Hyperparameters	Values
C	[0.1, 1, 10, 100]
Gamma	[0.001, 0.01, 0.1, 1]
Kernel	RBF

E. *Volatility Strategy Module*

As explained earlier, volatility plays a very important role in this thesis, since one of the main objectives is to understand if through the characterization of the input signals (prices) a strategy can be created to avoid volatile periods, trading with more stability, less risky and less unpredictable (mainly associated with more volatile periods).

In this module four technical indicators were used, one of which was one of the most used volatility indicators, the ATR. The other three indicators were developed using signals (candles) to obtain signals where volatility can be interpreted.

- Compare two Candles (CTC): This indicator starts by calculating the difference between the current candle price and the price of a past candle. When this difference is small it means that the market is not fluctuating too much. This indicator was intended, in addition to analyzing the difference, to understand mainly when the market has abrupt declines or when it is very unstable, and so, before averaging these differences, only the part in which the difference is negative, making zero the positive values.
- Candles Size Average (CSA): Normally when candle size is reduced it means the market is calm, not very volatile. In this sense, this indicator averages the sizes of the last candles. Then the final average is applied to the absolute average values of the last candle size.
- Candles Frequency (CF): When, for example, there is a considerable sequence of positive candles the market tends to be calmer. Therefore, this indicator aims to analyze the frequency with which each type of candle arises, that is, whenever the candle is positive (green) is assigned 1 and when the candle is negative (red) is assigned 0. When applying an average to these values give a value indicating which candles appeared most frequently in the previous period.

Using these four indicators rules were created that allow selecting the quieter periods of the asset being analyzed. Figure 3 shows the prices of an asset over a given period and Figures 4 e 5 show the CTC and CF indicator results when applied to the chart in Figure 3.

Fig. 3: Graph of the prices of an asset over a given period.

Fig. 4: CTC indicator chart applied to the chart in Fig. 3.

Fig. 5: CF indicator chart applied to the chart in Fig. 3.

In Figures 4 and 5, besides the normal results of the indicators are illustrated the rules that select the quietest periods of a given financial asset. As you can see in these figures, there is a line (red), representing the decision margin of each indicator. This margin is adjusted and optimized by GA for each indicator. If the ATR, CTC and CSA values below these margins (green) are considered as calm periods, then the decision is 1. If CF indicator values are higher than its margin, they are considered calm periods. In opposite cases the indicators classify the period with decision 0.

The GA also determines the weight of each technical indicator and a decision variable that allows the decisions of the various indicators to be combined to obtain the decisions for which periods are less volatile. The decision process consists of multiplying the weights of each indicator to the binary decisions generated above. If the sum of all these multiplications is greater than the decision variable, the volatility strategy understands that this is a calm period, giving the final binary decision of 1. If the sum is less than the decision variable, the volatility strategy understands, that this is a more unstable period, choosing not to trade (final binary decision of 0).

F. *Genetic Algorithm Module*

In this module, a GA was used which plays a very important role in the whole system developed. This is responsible for optimizing all required parameters, whether these are the parameters used in the SVM module, the volatility strategy module, or both strategies.

Next, the genes (parameters to optimize) of each chromosome will be presented and explained, which fitness function was used to classify the created chromosomes, the selection process of the best chromosomes, the crossover process applied to the selected chromosomes in the chromosome phase. selection and the mutation process.

1) *Chromosome Representation*

The chromosome encodes the information that is required to optimize the two strategies explained earlier. For market direction identification strategy (SVM), 22 genes are used, which represent the feature selection parameters, the time periods of these and the SVM C and γ (Gamma) hyperparameters. For the volatility strategy, are used 9 genes, representing the weights of the volatility indicators, the decision limits for each indicator and the variable of decision.

Table 2 shows the structure of a chromosome with the values of each gene. Note that the indicator periods referred to in Table 2 are multiplied by 12 (numbers of hours the market is open per day) before being used in the system.

Table 2: Genes that each chromosome contains.

Chromosome										
SMA		EMA		WMA		ADX		RSI		
Selector	Period	Selector	Period	Selector	Period	Selector	Period	Selector	Period	
0 or 1	5 to 30	0 or 1	5 to 30	0 or 1	5 to 30	0 or 1	5 to 30	0 or 1	5 to 30	
TRIX		MOM		WILLR		CCI		AROON		
Selector	Period	Selector	Period	Selector	Period	Selector	Period	Selector	Period	
0 or 1	5 to 30	0 or 1	5 to 30	0 or 1	5 to 30	0 or 1	5 to 30	0 or 1	5 to 30	
ATR		CTC		CSA		CF		Decision	C	γ
Margin	Weight	Margin	Weight	Margin	Weight	Margin	Weight			
0.1 to 0.2	0 to 1	0.1 to 0.2	0 to 1	0.1 to 0.2	0 to 1	-0.05 to 0.05	0 to 1	0.5 to 1	0.1 to 1000	0.001 to 0.1

2) *Fitness Function*

When strategies are combined, the fitness function aims to better classify those chromosomes that have the best risk-return ratio in the Volatility Strategy and the best accuracy in SVM. The fitness function of the solution that combines both strategies is represented in equation 1:

$$fitness = \frac{ROI}{Max.Drawdown + Average\ Drawdown} * accuracy \qquad (1)$$

3) *Selection*

Once the initial population is defined, the process begins with the selection of individuals, who will choose the individuals with the best fitness to evolve with crossover or mutation, hoping that individuals of the next generation will have better fitness. The selection method used was Elitist selection, as it ensures that in all generations the selected individuals are those who have the best fitness.

The selection process begins with 200 initial chromosomes, then the 100 most fit chromosomes are selected and advance to the next phase, while the 100 least fit chromosomes are eliminated.

4) *Crossover*

At this stage, two or more chromosomes are combined to generate new progeny. The genes of the selected parent chromosome selection phase are combined with the goal of finding offspring with better fitness.

The crossover process used was the two-point crossover, as it was designed to better weaken some of the other types of crossover. Regarding the crossover rate, the value used is 0.5, as it was the most found in the literature.

5) *Mutation*

This phase plays an important role in the functioning of GA, as although the mutation rate is low, in the case of this 0.2 system, it promotes diversity in the genetic population, in order to avoid easily converging to a local maximum, always trying to obtain the great solution.

This is the last phase of GA, and this is where the genes will be randomly modified. The mutation performed in this system is based on evenly distributed increments from a specific range of values for each gene. The decision margin genes of the volatility indicators are mutated individually within the range of [-0.1:0.1]. The decision variable gene of the volatility strategy is mutated in the range of [-0.1:0.1]. Regarding the weights of the volatility indicators, as their sum must always be equal to 1, a value between 0 and 0.2 is generated which is then added to two of the weights and subtracted from the other two weights thus keeping the sum of the four weights equal. a 1. The genes of the Trend and *Momentum* indicator periods individually mutate within the range of [-0.1:0.1]. The Trend and *Momentum* indicator selectors do not change. For SVM hyperparameters, they mutate according to their value, that is, the higher the value of C and Gamma the greater the range of mutation values.

IV. System Evaluation

To train and test the proposed system, the input data were minute prices (opening price, highest price, lowest price and closing price) from different markets. This data was resampled, turning the minute data into hourly data. This data set was then separated into two sets, the training data set and the test data set, with the following percentages, 70% and 30% respectively. This procedure was performed for the seven markets that were analyzed. The training periods, test periods and markets tested were as follows:

• S&P500 Index: The training set is from 23/11/2012 at 2:00pm until 28/03/2017 at 4:00pm and the test set runs from the end of the training set until 13/12/2018 at 2:00pm.

• Amazon.com Inc. stock (AMZN symbol): Training set is 10/25/2012 at 2:00pm until Jan. 31, 2017 at 2:00pm and the test set runs from the end of training set until 11/23/2018 at 5:00pm.

• Align Technology, Inc. stock (ALGN symbol): Training set is from 25/10/2012 at 2:00pm until 01/31/2017 at 2:00pm and the test set runs from the end of the training set until 11/23/2018 at 5:00pm.

• Target Corporation Inc. stock (TGT symbol): Training set is 18/10/2012 at 2:00pm until 03/07/2017 at 5:00pm and the test set runs from the end of the training set until 10/07/2019 at 8:00pm.

• Johnson & Johnson Stock (JNJ symbol): The training set is 18/10/2012 at 2:00pm until 03/07/2017 at 5:00pm and the test set runs from the end of the training set until 10/07/2019 at 8:00pm.

• Walmart, Inc. Stock (WMT symbol): Training set is 10/18/2012 at 2:00pm until 3/07/2017 at 5:00pm and the test set runs from the end of the training set until 10/07/2019 at 8:00pm.

• McDonald's Corporation Stock (MCD symbol): Training set is 10/18/2012 at 2:00pm until 3/07/2017 at 5:00pm and the test set runs from the end of the training set until 10/07/2019 at 8:00pm.

With this diversity of markets, it is possible to test the system developed in different types of markets and to analyze its performance in markets with very different characteristics. For example, the S&P500 is considered to be a low volatile market unlike TGT which is known for its instability.

The main objective of the proposed system is to obtain the best possible return-risk ratio. In order to evaluate the performance of the strategies used in this thesis, four metrics were calculated: Return of Investment (ROI), ROI divided by the number of hours within the market, Maximum Drawdown and Return Over Maximum Drawdown (RoMaD).

A. *Evaluation of the proposed system results*

Some market trend identification algorithms, despite showing good results in financial returns during the tests, end up trading in more unstable periods (high volatility), eventually having the possibility of suffering high drawdowns. As explained above, the main objective of this thesis was to obtain the best possible return-risk ratio and for this reason strategies were created based on volatility and financial market trends. These strategies were combined through a genetic algorithm, thus creating the system presented in this thesis, whose fitness function is equation 1. The objective was to understand if using a market trend forecasting algorithm, together with a volatility strategy, these more volatile periods could be avoided by continuing to have good financial returns.

Before analyzing the performance of the proposed system in the seven financial markets chosen to evaluate the system, it is important to explain that in this case study three strategies were compared despite the figures presenting 5 graphs. The GA volatility strategy and SVM GA volatility charts were only placed to understand how the final strategy is created. The final strategy, is the result of multiplying the binary decision vectors of the two strategies (volatility and trend), or in other words, only when the two strategies agree, the final strategy should invest. In conclusion, the three strategies that are compared in this case study are the Buy&Hold strategy, the final strategy and the final strategy with a 2:1 leverage.

Figures 6, 7 e 8 show the results obtained by the proposed system during the testing period at the S&P500 index.

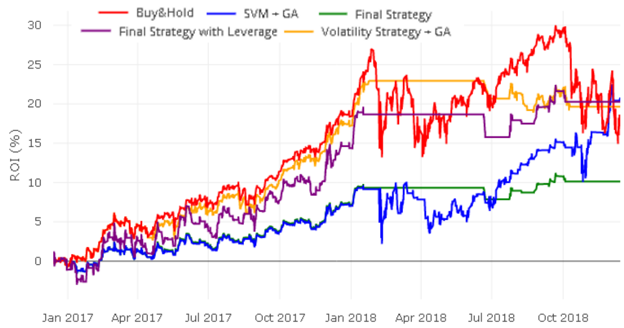

Fig. 6: ROI of the proposed system, the two strategies that were combined to generate the proposed solution and the Buy&Hold strategy on the S&P500 index.

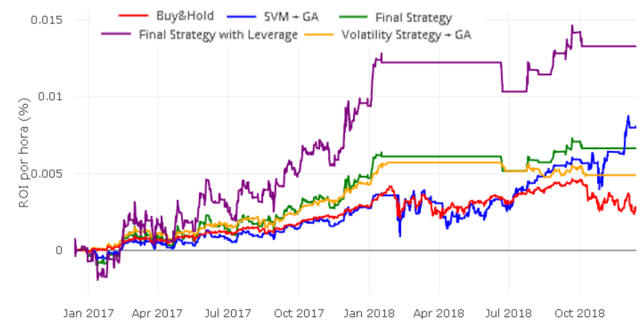

Fig. 7: Proposed system ROI / hour, the two strategies that were combined to generate the proposed solution and the Buy&Hold strategy on the S&P500 index

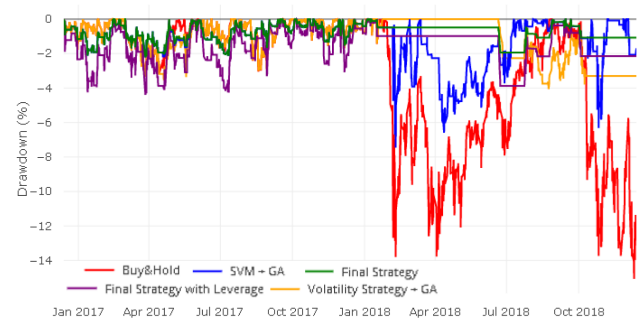

Fig. 8: Drawdown of the proposed system, the two strategies that were combined to generate the proposed solution and the Buy&Hold strategy on the S&P500 index.

With Figures 6, 7 and 8 it can be observed that the final strategy does not surpass the Buy&Hold strategy but the final strategy with Leverage already surpasses the Buy&Hold strategy. However, you need to look at Drawdown to understand if leveraging puts you at greater risk than the Buy&Hold strategy. In Figure 8 you can see that the strategy with the lowest drawdown is the final strategy (of the proposed system), this drawdown being so low that even applying Leverage still has a considerably lower drawdown than the Buy&Hold strategy. As for Figure 7, it can be seen that both the final strategy and the Leverage final strategy have a ROI / hour much higher than the Buy&Hold strategy because the returns obtained were made over a much shorter period. In Figure 6 it is important to highlight the advantages of combining this type of volatility and trend strategies, as during the period between late January 2018 and early July 2018, the trend identification strategy (SVM with GA), could not entirely avoid the big declines of the S&P500 index but the volatility strategy avoided this whole period, thus creating a final strategy with much less drawdown. Between July 2018 and October 2018, the volatility strategy selected some periods that it considered less volatile and eventually lost in some of them, but when the trend identification strategy and the volatility strategy were combined, it is clear that the final strategy avoids most of it.

Figures 9, 10 and 11 show the results obtained by the proposed system during the AMZN stock testing period.

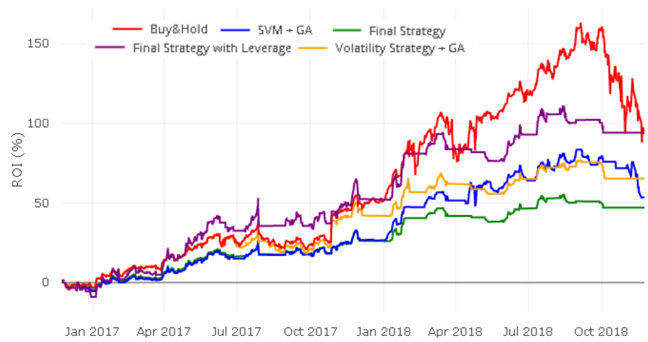

Fig. 9: ROI of the proposed system, the two strategies that were combined to generate the proposed solution and the Buy&Hold strategy in the AMZN stock.

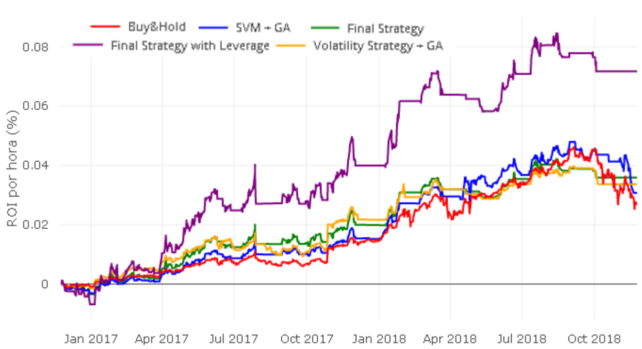

Fig. 10: Proposed system ROI / hour, the two strategies that were combined to generate the proposed solution and the Buy&Hold strategy on the AMZN stock.

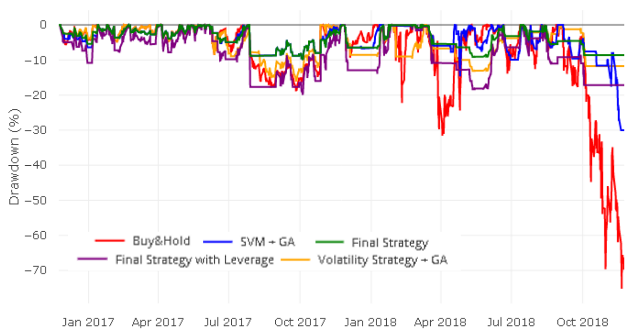

Fig. 11: Drawdown of the proposed system, the two strategies that were combined to generate the proposed solution and the Buy&Hold strategy on the AMZN stock.

Regarding the tests performed in the AMZN stock, represented in Figures 9, 10 and 11, some similar conclusions can be drawn to the conclusions previously mentioned in the S&P500 index test, in this case study. The final strategy can't beat the Buy&Hold strategy but the time it is in the market it is much lower. From the analysis of Figure 9, it is possible to observe that the return obtained by the final strategy, although not higher than the Buy&Hold strategy, is a return very stable over the test period, with little drawdown. Having a strategy that is concerned with minimizing risk allows you to apply Leverage and still gets good drawdown results. This can be seen in more detail in Figure 11, where the final strategy is the least drawdown strategy avoiding the most volatile periods, allowing the final Leverage strategy to still draw significantly less drawdown than the Buy&Hold strategy. The final Leverage strategy, with far fewer hours on the market, observable in Figure 10, and a much lower drawdown, observable in Figure 11, achieves a ROI similar to the Buy&Hold strategy.

Figures 12, 13 and 14 show the results obtained by the proposed system during the TGT stock testing period.

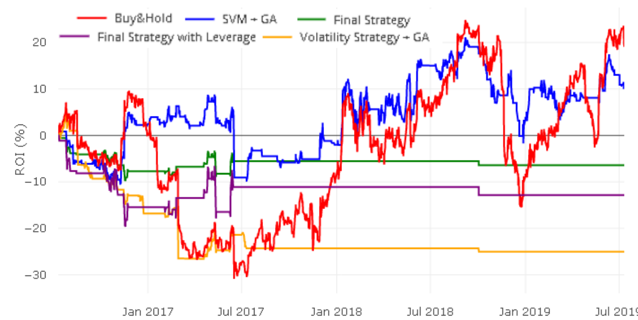

Fig. 12: ROI of the proposed system, the two strategies that were combined to generate the proposed solution and the Buy & Hold strategy on the TGT stock.

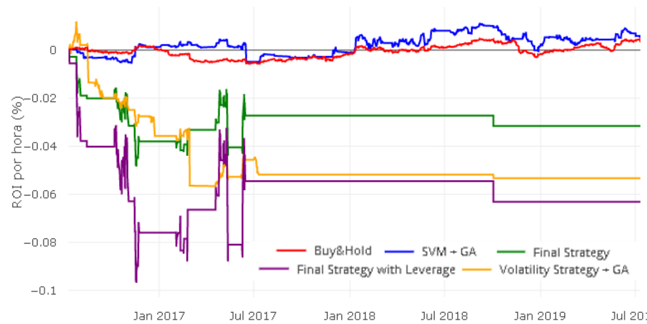

Fig. 13: ROI / hour of the proposed system, the two strategies that were combined to generate the proposed solution and the Buy & Hold strategy on the TGT stock.

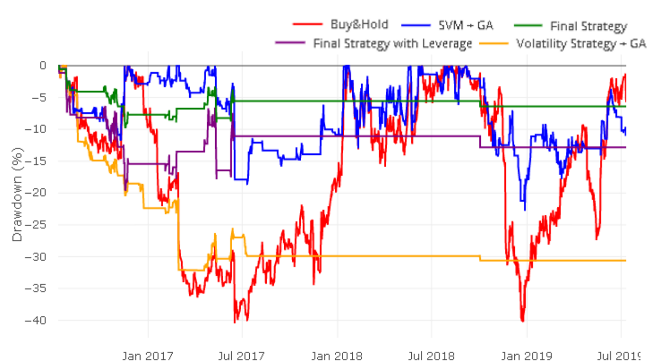

Fig. 14: Drawdown of the proposed system, the two strategies that were combined to generate the proposed solution and the Buy &Hold strategy on the TGT stock.

Of the seven markets tested, the tests performed on the TGT stock showed the worst results. The TGT stock is known for its great volatility and unpredictability, which makes it very difficult for any type of investment strategy to perform well. From the analysis of Figures 12, 13 and 14 it is possible to notice that the final strategy underperformed the Buy&Hold strategy. Figures 12 and 13 show that the volatility strategy realizes that this asset is quite volatile and decides to practically not trade, but when it trades in periods it considered less volatile, it suffered some declines. The trend identification strategy was able to avoid some of the TGT stock declines, achieving a positive ROI, but slightly lower than the Buy&Hold strategy. The final strategy, obtained through the volatility strategy and the trend strategy, practically does not trade, and as you can see in Figure 14, is the strategy that had a lower drawdown, but could not surpass Buy&Hold, obtaining a negative ROI.

Finally, Table 3 presents an example of the best performing chromosome during the second test period performed on the S&P500 index. Here you can get an idea of how SVM trend and momentum indicators are selected and used, the selector variable for each indicator indicates whether this indicator has been selected (if 1) or not (if 0), in addition to the respective periods, which before being used to calculate the indicators, are multiplied by 12, the number of hours the market is open per day.

Table 3: Example of a chromosome obtained during the second test period at the S&P500 index.

Chromosome										
SMA		EMA		WMA		ADX		RSI		
Selector	Period	Selector	Period	Selector	Period	Selector	Period	Selector	Period	
1	14	1	27	1	29	0	14	0	8	
TRIX		MOM		WILLR		CCI		AROON		
Selector	Period	Selector	Period	Selector	Period	Selector	Period	Selector	Period	
1	19	1	7	1	5	1	18	0	30	
ATR		CTC		CSA		CF		Decision	C	γ
Margin	Weight	Margin	Weight	Margin	Weight	Margin	Weight			
0.137	0.381	0.181	0.019	0.128	0.563	0.002	0.035	0.534	0.09	0.009

After a qualitative analysis of the results obtained in this case study on the S&P500 index, the AMZN stock and the TGT stock, it is necessary to quantitatively analyze the results of the strategies used in this case study in all the previously mentioned markets. Table 4 shows the results of the strategies applied in this case study in all markets, where it is possible to analyze all the above mentioned metrics, with the addition of the percentage of hours in the market.

Table 4: Results of final strategy, Leverage final strategy, and Buy&Hold strategy in the seven markets tested

	ROI (%)	**ROI / Hours (%)**	**MDD (%)**	**Hours in market (%)**	**RoMaD**
S&P500					
Final Strategy	10.14	0.0066	2.17	23.58	4.67
With Leverage	20.28	0.0132	4.34	23.58	4.67
Buy&Hold	17.34	0.0027	15.07	100	1.15
AMZN					
Final Strategy	47.08	0.0359	9.99	37.35	4.71
With Leverage	94.16	0.0718	19,98	37.35	4.71
Buy&Hold	93.45	0.0266	75.19	100	1.24
TGT					
Final Strategy	-6.41	-0.0316	9.82	3.85	-0.65
With Leverage	-12.81	-0.0631	19,66	3.85	-0.65
Buy&Hold	19.14	0.0036	40.48	100	0.47
ALGN					
Final Strategy	90.47	0.1121	33.90	22.97	2.67
With Leverage	180.94	0.2242	67.80	22.97	2.67
Buy&Hold	130.61	0.0371	206.47	100	0.63
JNJ					
Final Strategy	9.15	0.0189	6.69	9.18	1.37
With Leverage	18.30	0.0378	13.38	9.18	1.37
Buy&Hold	15.16	0.0028	23.94	100	0.63
MCD					
Final Strategy	22.74	0.0253	7.81	17.04	2.91
With Leverage	45.49	0.0507	15,63	17.04	2.91
Buy&Hold	74.45	0.0141	25.29	100	2.94
WMT					
Final Strategy	10.93	0.0319	7.67	6.50	1.43
With Leverage	21.86	0.0639	15.34	6.50	1.43
Buy&Hold	52.47	0.0099	37.48	100	1.39

Analyzing the table with the results of the proposed solution, Table 4, it can be seen that the results are quite positive. They demonstrated that by combining the volatility strategy with the trend identification strategy a conservative strategy has been created which, while never outperforming the Buy&Hold strategy, achieves a much lower MDD than the Buy&Hold strategy MDD. A strategy with these characteristics allows leverage to be used without entering in very high risk levels. As mentioned above, the leverage used was 2:1, and even with leverage, the proposed strategy MDD is considerably lower than the Buy&Hold strategy MDD in all tested markets. Regarding the ROI obtained in the various stocks, with the final strategy with Leverage, it can be noted that in 4 markets out of the 7 markets tested, the ROI is higher than the ROI obtained with the Buy&Hold strategy. With the RoMaD results we can see the advantage of the final strategy in relation to the return-risk ratio. In the best case, in the S&P500 index, the final strategy RoMaD was 4 times higher than the Buy&Hold strategy RoMaD. Note that the proposed strategy was able to achieve the returns, shown in Table 4.6, with far fewer hours in the market than the Buy&Hold strategy. Once again, the worst results occurred in the tests performed with the TGT stock. The proposed system realized that this was a volatile market and therefore the percentage of hours on the market during testing was only 3.85%. However, during these periods the system did not prevent some market declines, closing with an ROI of -6.41%.

V. Conclusion

This thesis presents a system composed by a genetic algorithm that combines two strategies, a trend identification strategy and a low volatile period identification strategy. The main objective of this system was to create a trading strategy with the best possible return-risk ratio.

In order to evaluate the performance of the proposed system, tests were performed in seven markets with different characteristics, thus allowing to analyze to which type of markets the proposed trading strategy best fits.

In most of the tests performed, the results obtained were very positive and the proposed system demonstrated its great ability to minimize the risk of trading.

Regarding the results obtained by SVM with feature optimization and selection through GA, it was expected to notice more the advantages of using GA over SVM with Grid Search. It is believed that this difference could be increased if the number of indicators used were larger, because the GA would, through a greater diversity of indicators, select the most appropriate ones to improve system performance. The results obtained by the SVM can be considered as positive, and the labeling system presented had a fundamental role to obtain these results.

The volatility strategy, as noted in the last two case studies, played a major role in reducing the risk associated with investing. This strategy, in most tests, has successfully avoided the most volatile periods on the market. However this low volatility strategy is not perfect and in the last case study it was proved that by combining the volatility strategy with the trend identification strategy was possible to get good financial returns with much less risk associated with the investment.

One of the main conclusions that can be drawn from the tests performed is that the developed trading system stably achieves returns over time, with few declines, which makes the system reliable. Due to this confidence and stability, the use of 2:1 leverage has been tested, which not only outperformed the Buy&Hold strategy in most tested markets, but also had much lower Maximum Drawdowns than the Buy&Hold strategy.

References

[1] Murphy, J. J. (1999). *Technical analysis of the financial markets: A comprehensive guide to trading methods and applications*. Penguin.

[2] Malkiel, B. G., & Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, *25*(2), 383-417.

[3] Yoo, P. D., Kim, M. H., & Jan, T. (2005, November). Machine learning techniques and use of event information for stock market prediction: A survey and evaluation. In *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)* (Vol. 2, pp. 835-841). IEEE.

[4] Guresen, E., Kayakutlu, G., & Daim, T. U. (2011). Using artificial neural network models in stock market index prediction. *Expert Systems with Applications*, *38*(8), 10389-10397.

[5] Khaidem, L., Saha, S., & Dey, S. R. (2016). Predicting the direction of stock market prices using random forest.

[6] Cao, L., & Tay, F. E. (2001). Financial forecasting using support vector machines. *Neural Computing & Applications*, *10*(2), 184-192.

[7] Hsu, J., & Li, F. (2013). Low-volatility investing. *The Journal of Index Investing*, *4*(2), 67-72.

[8] Helleiner, E. (2011). Understanding the 2007–2008 global financial crisis: Lessons for scholars of international political economy. *Annual review of political science*, *14*, 67-87.

[9] Ali, M. (2014). Relationship between Financial Leverage and Financial Performance (Evidence of Listed Chemical Companies of Pakistan). *Research Journal of Finance and Accounting*, *5*(23), 46-56.

[10] Tsatsaronis, K., & Yang, J. (2012). Bank stock returns, leverage and the business cycle. *BIS Quarterly Review, March*.

[11] Jasemi, M., Kimiagari, A. M., & Memariani, A. (2011). A modern neural network model to do stock market timing on the basis of the ancient investment technique of Japanese Candlestick. *Expert Systems with Applications*, *38*(4), 3884-3890.

[12] Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, *42*(1), 259-268.

[13] Huang, W., Nakamori, Y., & Wang, S. Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers & operations research*, *32*(10), 2513-2522.

[14] Pinto, J. M., Neves, R. F., & Horta, N. (2015). Boosting Trading Strategies performance using VIX indicator together with a dual-objective Evolutionary Computation optimizer. *Expert Systems with Applications*, *42*(19), 6699-6716.

[15] Yu, H., Chen, R., & Zhang, G. (2014). A SVM stock selection model within PCA. *Procedia computer science*, *31*, 406-412.

[16] Gorgulho, A., Neves, R., & Horta, N. (2011). Applying a GA kernel on optimizing technical analysis rules for stock picking and portfolio composition. *Expert systems with Applications*, *38*(11), 14072-14085.

[17] Huang, C. L., & Wang, C. J. (2006). A GA-based feature selection and parameters optimization for support vector machines. *Expert Systems with applications*, *31*(2), 231-240.

[18] de Almeida, B. J., Neves, R. F., & Horta, N. (2018). Combining Support Vector Machine with Genetic Algorithms to optimize investments in Forex markets with high leverage. *Applied Soft Computing*, *64*, 596-613.

[19] Data Preprocessing for Supervised Leaning. S. B. Kotsiantis, D. Kanellopoulos, and P. E. Pintelas

[20] De Prado, M. L. (2018). *Advances in financial machine learning*. John Wiley & Sons.