Portfolio Selection and Trading Model based on Genetic Algorithms, K-Means Clustering, Fundamental Indicators and Technical Indicators

João Quinas Guterres
joaoqguterres@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

November

Abstract

A model capable of utilizing Evolutionary Algorithms and Fundamental Analysis for stock selection, this alongside a trading system based on Technical Analysis is proposed. Firstly a data scrapping program is created in order to obtain the financial statements of the S&P500 companies for every end of the month in the data range, allowing the calculation of various Fundamental Indicators. A K-means clustering algorithm is then used to classify these indicators among sector peers for every end of the month in the data range. A Genetic Algorithm will then search for the importance that should be given to each indicator using a previous three month stock price variation method as the fitness function. Three different strategies with different fundamental indicators are considered, one focused on growth, another on general performance and another focused on value companies. This stock selection algorithm is then tested under different portfolio compositions and alongside a trading system based on the EMA indicator, with the testing being done over the period of 2017-02-01 up to 2021-10-31. The results show that all strategies have potential, although the value and balanced strategies are the ones that most consistently beat the benchmark, yielding returns up to 119.4% in 2020, displaying their strength when recovering from market downtrends. The EMA trading system proved to be a useful tool as well, mainly when it comes to avoiding heavy losses and reducing volatility

Keywords: Genetic Algorithms; K-means Clustering; Stock Selection; Trading System; Fundamental Analysis; Technical Analysis.

1. Introduction

1.1. Investment Strategies

The stock market encapsulates various venues where shares of publicly held companies can be bought, sold and issued, such as the NYSE, SSE and Euronext.

Investors purchase shares of a company in an attempt to profit via its stock price appreciation and regular dividend payments. The main goal is to maximize profits whilst reducing risk, and, along the years, multiple theories and strategies have been developed in order to thrive in this very volatile and competitive environment. The two most prominent schools of thought that have emerged are Fundamental Analysis and Technical Analysis.

Utilizing Fundamental Analysis, an investor seeks stocks that are undervalued expecting that their market price will eventually rise to its ”fair value”, capitalizing on the aforementioned opportunity. To determine the intrinsic value of a given stock numerous factors are considered, both qualitative, such as the management and business model, and quantitative, namely a company's financial statements and macroeconomic factors. On the other hand, Technical Analysis focuses on finding trading opportunities resorting to price trends, searching for patterns in charts and considering volume in an attempt to forecast price movements.

1.2. Sector Classification

When attempting to analyse the stock market it is important to aggregate companies with similarities into various groups, given that this allows the investor to get a better understanding of how a more specific type of company is performing, which may highlight general underlying problems or advantages. Various industry classification schemes have been developed, such as the Global Industry Classification Standard (GICS), the North American Industry Classification System (NAICS) and the Fama-French (FF) industry classification, with the GICS being considered superior [1]. Developed in 1992 by MSCI and S&P 500 Dow Jones Indices, the GICS is a four-tiered system with the following hierarchy: 11 sectors, 24 industry groups, 69 industries and 158 sub-industries. The sectors defined by this system are: Energy, Materials, Industrials, Consumer Discretionary, Consumer Staples, Health Care, Financials, Information Technology, Communication Services, Utilities and Real Estate.

1.3. Evolutionary Algorithms

In recent years computational intelligence has become more prevalent in the world of investing [2] [3] [4] , mostly due to the amount of information available regarding companies and financial markets. Many studies have been developed "covering techniques for preprocessing and clustering of financial data, for forecasting future market movements, for mining financial text information, among others", [2]. Technical and Fundamental indicators are usually considered, with the former being more commonly used as input for machine learning and deep learning models [3], although some studies advocate for the advantages of combining both indicators [5]. In this work both Technical and Fundamental indicators will be considered.

A particular type of algorithm that has gained traction is the Evolutionary Algorithm. EAs [6] are metaheuristic optimization algorithms inspired by concepts in Darwinian Evolution with mechanisms that might include selection, reproduction, mutation and recombination. The EAs family is composed by the following main algorithms: *genetic algorithm* (GA), *genetic programming* (GP), *differential evolution* (DE), the *evolution strategy* (ES) and *evolutionary programming* (EP).

1.4. Objectives

The main objective of this work is to the develop a trading mechanism capable of selecting the best performing stocks and creating portfolios that can, at a bare minimum, outperform the SP&500 benchmark. In order to achieve this goal various secondary objectives must be met, with these objectives being: (1) Download and use fundamental data in order to calculate various fundamental indicators for SP&500 companies; (2) Create a classifier that ranks a company by classifying its fundamental indicators, this based on contrasting these ratios against sector peers; (3) With the help of a GA, determine weights that indicate the importance of each indicator; (4) Combining both the classifier and weights create a novel stock ranking system; (5) Take advantage of the new ranking system to select stocks for portfolio creation; (6) Test these portfolios not only alone but also when applying a BUY/SELL system based on Technical Indicators.

2. Related Works

Table 1 summarizes the related works analysed.

In [7] a GA was taken advantage of in order to optimize portfolios with the assistance of a momentum strategy and using the CAPM to determine undervalued stocks. The implemented fitness function took into account the Portfolio fund standardization, the Portfolio CAPM and its Sharpe Ratio when evaluating chromosomes. The model was tested in two different markets, SP 500 and KOSPI200, projecting a better performance than both indexes.

In [8] both financial ratios and technical indicators were used, alongside a MOEA (SPEA II). The fitness functions considered were the return and its respective risk, and the models considered up to 10 financial ratios and 5 trading parameters for the chromosomes. Tested for the SP&500 from June 2010 to 2014, the results obtained were not only above the market's average but also paired with low variances.

In [9] a clustering-based portfolio optimization model using a GA and investor information is implemented. Firstly, taking advantage of investor information, various portfolios are generated via clustering analysis, using K-means clustering. Then, in order to optimize the weights of the selected stocks, a GA is used, with the fitness functions being either the Minimum variance weights or the Sharpe ratio weights. The proposed model outperformed previous models when applied to the KOSPI200.

In [10] a MOEA (NSGA-II) portfolio optimization model is paired with TIs. The genetic algorithm aims to minimize risk (Covariance or CVar) and maximize the return function, developed in [16], also considering 4 indicators. Two scenarios are possible, the first one only using the TIs for transactions after the optimal portfolio is chosen at the start of each month and the second will perform the monthly optimization only on the stocks selected by the indicators. The simulation encapsulates 6 years of data from the Brazilian Stock Exchange and the strategy focused on using the optimization first performed better than both the index and the other strategies.

In [11] a SVR method generates replacements for actual stock returns, in order to rank stocks, where the top rated stocks will be used when forming a portfolio. Supporting the SVR, a GA is employed for parameter optimization, and feature selection providing the best input variables to the SVR model. The data selected corresponds to the 200 largest market capitalization stocks in the Taiwan Stock Exchange, with dates ranging from 1996 to 2010. The attributes used in the stock selection model are fundamental ratios related to

Table 1: Related works

Work	Algorithm	Fitness Function	Benchmark	Financial Application	Period	Results
[7]	GA	Sharpe Ratio and Portfolio CAPM	S&P 500; KOSPI200	Portfolio Optimization	2008-2018	Best: 400% cumulative returns
[8]	MOEA (SPEA II)	Returns and variance of returns	S&P 500	Portfolio Composition	2010-2014	Best: 50.24%
[9]	GA	MV weights/Sharpe weights	KOSPI200	Portfolio Optimization	2007-2014	Best: Annual Return of 40.33%
[10]	NSGA-II	Returns and Risk	Brazilian Stock Exchange	Portfolio Optimization	2012-2015	Best: 68.09%
[11]	SVR and GA	Annualized Return of the Portfolio	Taiwan Stock Exchange	Stock Selection	1996-2010	Best: 17.5719 %
[12]	SPEA-II	Returns and Risk	SSE	Portfolio Optimization	2012-2013	Best: 47.07%
[13]	K-means	-	Chinese stock markets	Portfolio Construction and Optimization	2001-2020	Annualized Returns: 30%
[14]	Various clustering algorithms	-	Multiple American Indexes	Stock Selection and Portfolio Optimization	2001-2017	Best: 45.09%
[15]	-	-	Brazilian Stock Market	Trading System	2000-2014	Best average return (day): 0.061%

share price rationality, profitability, leverage, liquidity, efficiency and growth. The models applied significantly outperformed the benchmark, achieving a maximum 17.5719 % mean of annualized model return.

In [12] TIs and FIs are used for the creation of a model, alongside a MOEA that assesses risk and returns for optimization. The study was based on 40 stocks of the SSE A, obtained via the Cathay Pacific database. The results show that the model created is able to outperform the index and, as the number of fundamental indicators increases so does the performance, specially with the addition of cash-flow growth, capital expenditure growth rates and the payout ratio.

In [13] a portfolio construction method is proposed that takes into account the continuous trend characteristics of the market. Firstly K-means clustering is used to cluster stocks, divide the different stock groups and revise the calculation of returns for the Sharpe Ratio, this based on the continuous trend characteristics. Various portfolio theories are then utilized to calculate the required weights. The results obtained show that the proposed method was superior.

In [14] a portfolio selection algorithm is presented that, based on the pattern matching principle, selects the optimal portfolio, this updated periodically. The two steps of the system consist of the sample selection, that utilizes various clustering algorithms including k-means, and the portfolio optimization, where the optimum function and transaction costs are considered. Various data sets were considered for different american indexes and time frames, ranging from 2001 up to 2017. The results indicate that the proposed models outperformed others provided in the literature.

In [15] the performances of the SMA, EMA, MACD and Triple Screen techniques are analyzed in a trading system. 198 stocks traded in the Brazilian stock market were used for various different periods ranging from 2000 up to 2014. Multiple brokerage fees were considered alongside a Stop-Loss mechanism and the benchmark considered was a buy-and-hold strategy. Although most results indicated that the strategies had positive returns only a small percentage was actually capable of overcoming the buy-and-hold strategy.

3. Architecture

The model proposed is composed by five main modules: (1) **Input Strategy**: the investor may choose between three main strategies, a **Growth Strategy**, a **Balanced Strategy** and a **Value Strategy** which will determine the indicators taken into consideration when classifying stocks and training the GA; (2) **Data Acquisition Module**: this module will be responsible for the storage of SP&500's companies fundamental information as well as their historical price variations; (3) **Stock Classifier Module**: based on the strategy chosen, this module will utilize the fundamental indicator information obtained in the **Data Acquisition Module** and classify them for all companies considered at the end of each month via a clustering method; (4) **GA Module**: depending on strategy, this module will attempt to correlate the importance of each indicator with the historical price variations; (5) **Portfolio Module**: this module will generate various portfolios based on the best companies gathered by combining the weights obtained in the **GA Module** and the classifications derived from the **Stock Classifier Module**. The usage of the technical indicator EMA will also be considered to decide IN/OUT positions.

Figure 1 shows the general flow of information of the model.

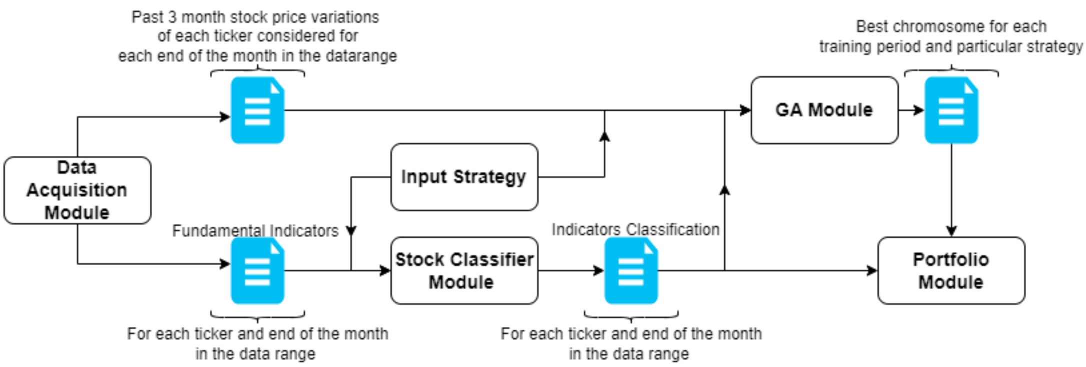

Figure 1: General flow of information

3.1. Input Strategy

The investor may choose various strategies that will define which indicators will be considered in the GA: (1) **Strategy A - Growth**: the indicators taken into consideration will be more focused on growth, overlooking debt to some extent and searching for growing and undervalued stocks that appear to have a good structure/business model; (2) **Strategy B - Balanced**: a balanced set of indicators is considered, taking into account both growth and stability; (2) **Strategy C - Value**: the main objective is to search for companies that have a dominant position in their respective sector, presenting good debt management, sustained growth and solid foundations.

3.2. Data Acquisition Module

The Data Acquisition Module, defined in figure 2, is mainly focused on two tasks: Gathering information regarding the fundamental indicators of each ticker for the data range considered and obtaining the monthly stock prices of tickers, calculating the past three month stock price variations and ranking the companies for each end of the month in the data range based on these variations.

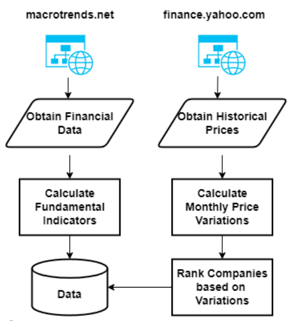

Figure 2: Overview of the Data Acquisition Module

3.3. Stock Classifier Module

The overall structure of the **Stock Classifier Module** is defined in figure 3. The main objective of this module is to classify, for each date in the data range and for each ticker considered, the fundamental indicators of all companies included, with the fundamental indicators selected depending on the strategy chosen. To achieve this objective the tickers were distributed into sectors and their fundamental indicator values stored with sector peers for all dates and fundamental indicators. These values were then clustered utilizing a K-means clustering algorithm and classified from 1-20, providing, for each date and ticker, the various classifications of their fundamental indicators.

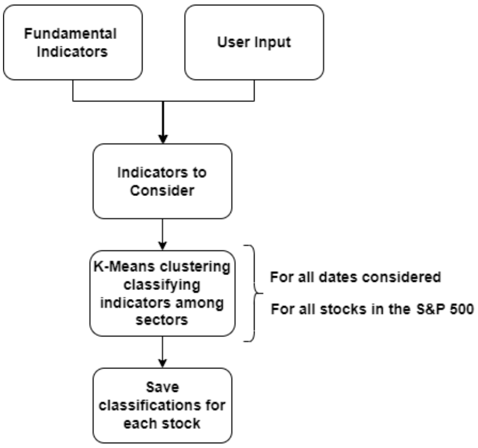

Figure 3: Overview of the Stock Classifier Module

3.4. GA Module

The overall functioning of the GA Module is shown in figure 4. The purpose of this module is to search for fundamental patterns that represent positive price variations in the stock market along various dates. Regarding the fundamental side, for each ticker a classification is done for each end of the month considering every single indicator, following the **Stock Classifier Module**. On the other side, for each end of the month, the past 3 month variation of the stock price of each ticker is calculated and used as the criterion for stock classification of that month. The algorithm will attempt to find the optimal weights that each fundamental indicator should have such that both rankings are the closest to each other for all dates.

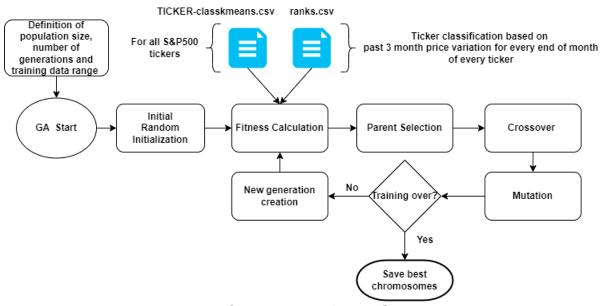

Figure 4: Overview of the GA Module

Chromosome structure: The chromosome structure will depend entirely on the number of fundamental indicators being considered, with each gene representing a ratio. For example, lets assume that N indicators are being considered, then the structure would be:

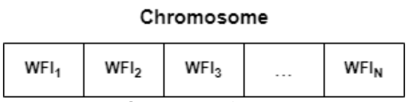

Figure 5: Structure of a chromosome

Each gene will represent the weight given to an indicator and, when initializing, a random value between -5 and 5 will be chosen for every gene of all initial population chromosomes.

Fitness Function: As described before, the goal of this module is to find what weights to give to each classified indicator that better mimic the past three month variations at each end of the month of the training set. This way, a fundamental pattern that reflects good stock price variations among various months is the desired objective. The way this is achieved is done by comparing the ranking achieved with the weights and the ranking achieved due to the 3 month price variations, along all training dates. The equation used to calculate the fitness of a chromosome is:

$$Fitness = \sum_{t=1}^{T} \sum_{n=1}^{N} \frac{1}{|OI_{tn} - I_{tn}| + 1}, \qquad \frac{OI_{tn}}{N} < 0.05 \tag{1}$$

Where T is the total amount of dates being considered, N the number of tickers available, I_{tn} the index of ticker n in the ranking obtained via GA weights for date t and OI_{tn} the index of ticker n for date t in the ranking generated resorting to the past 3 month variations. It is visible from the equation that the distance between both indexes is the value used when calculating the fitness, closer indexes will affect the fitness positively, whilst bigger differences affect it negatively, considering that the goal is to maximize the fitness. The condition $\frac{OI_{tn}}{N} < 0.05$ indicates that only the top 5% of the original tickers are considered, thus aiming at a more specialized evaluation that is only concerned with the top performers.

Training: When training the GA the population size, number of generations and data ranges must be defined. Firstly the population size selected was 100 with 150 generations. Regarding the training, two years of training for a year of testing was the decision, this due to the fact that since companies' financial statements are only updated each quarter a long term training should be beneficial for pattern finding. Figure 6 demonstrates the overall training method, which is the application of a rolling window, thus for each year the GA weights obtained depend on the two years prior. The overall data range considered for training is from **2015** up to **2021**, with testing starting in **2017**

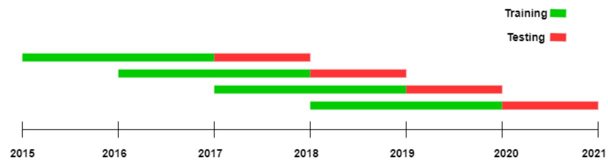

Figure 6: Rolling window training

3.5. Portfolio Module

The Portfolio Module, described in figure 7, is responsible for the selection and testing of various portfolios based on the weights obtained in the GA Module for the range of **01-02-2017** - **31-10-2021**.

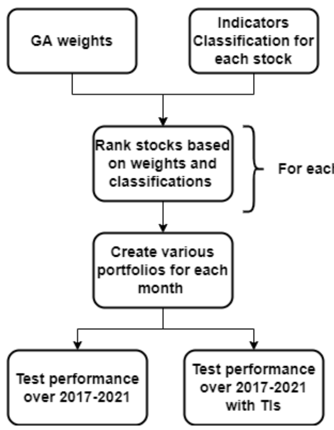

Figure 7: Overall view of the Portfolio Module

Initially, in this module, a python program receives the weights from the **GA Module** and the classifications obtained in the **Stock Classifier Module**. Utilizing these two it is possible to generate a new more nuanced classification that takes into account the importance of each indicator. Taking advantage of this new ranking sets of stocks are selected to generate portfolios. The four types of sets considered are: The best two stocks, the best five stocks, the best seven stocks and finally the best stock for each sector. It is important to reference that, following the way training proceeds these weights obtained will be different every year. For each strategy, **Growth Strategy**, **Balance Strategy** and **Value Strategy**, the algorithm was run four times, totalling twelve sets of chromosomes that will be used to form portfolios, four for each strategy. Each set will have five different weight combinations, one for each year.

Two different methods will be applied, a *laissez-faire* method that merely generates equally distributed portfolios and a more controlling method that, although still applying equal distribution, takes advantage of a simple technical indicator to define entry and exit points for the portfolio stocks each month. The indicator used will be the Exponential Moving Average (EMA) considering a 24-day window, which was the window that delivered the best results. The application of this indicator is simple, if the EMA of the previous day is bellow today's price a BUY signal is sent for that particular stock, else a SELL signal will be registered, as illustrated by 8.

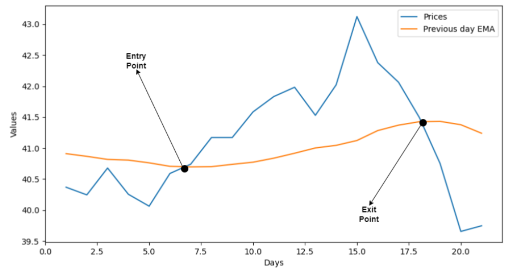

Figure 8: Application of the EMA trading system for one month (Purely exemplary)

4. Results

The GA described above was tested for years 2017 to 2021. For each year the algorithm was run various times in order to get a better understanding of its general performance. It is important to note that **the labeling of test runs is purely for organizational purposes and to get a better understanding of a long term performance, i.e, there is no correlation between Test Run 1 2017 and Test Run 1 2018 and so on**, thus, any random shuffle of Test Runs is valid for a long term (2017-2021) observation, the labeling of Test Runs is merely for simplicity's sake.

As stated before, three different strategies were tested, the **Growth Strategy**, where the algorithm will analyse growth focused fundamental indicators; the **Balanced Strategy**, which will take into consideration a wider range of indicators and the **Value Strategy**, that will attempt to prioritize the most stable companies.

In regards to the portfolio compositions, various different compositions were considered, a 2 stock portfolio (2SP), picking the top two companies for each month, a 5 stock portfolio (5SP), a 7 stock portfolio (7SP), both following the same logic as the 2SP, and a Sector stock portfolio (SSP), where a company from each sector is considered. The reason these are the proposed compositions is to analyse how the GA performs when taking into account variance, since naturally smaller portfolios are considered less stable, allowing for better performances but also greater failures.

These portfolios were all updated monthly according to the combination of classifier + GA. For all test runs two situations were considered, one where, for each month, the portfolio consists of the equal distribution of the stocks selected and another where, although the investment is equally distributed among stocks, a EMA based trading system is implemented to decide entry/exit points from the market.

The initial investment considered was of **10000**, the period considered for the EMA was **24 days**, the commission considered was **2%** of the money invested upon stock purchase and the GA features were as described in table 2.

Table 2: GA features

Number of Generations	150
Population Size	100
Number of Parents Mating	50
Crossover Probability	0.75
Mutation Probability	0.20

In this article only the **2SP results will be shown**.

4.1. Validation Metrics

4.1.1 Return on Investment (ROI)

The ROI is a measure that allows an investor to understand the profitability of a particular investment, specially when comparing to other investments. The basic equation of the ROI is:

$$ROI = \frac{ValueofInvestment - CostofInvestment}{CostofInvestment} \tag{2}$$

4.1.2 Sharpe Ratio (SR)

The SR is used in finance to evaluate the risk/return of an investment. It provides a perception to the investor of how much more risk is being taken for higher returns.

$$SharpeRatio(P) = \frac{R_P - R_F}{\sigma_P} \tag{3}$$

Where R_P is the average rate of return of the portfolio P, R_F is the risk-free rate and σ_P is the portfolio P's standard deviation. In this article a 0 risk-free rate is considered due to currently low interest rates.

4.2. Experiment I - 2 Stock Portfolio

Table 3 shows the yearly as well as total ROIs obtained for all Test Runs and all strategies when considering the 2SP, alongside their Sharpe Ratios.

For the **Growth Strategy** the average compound ROI of all Test Runs, with and without the EMA strategy was 117.6%, which was 34.6 percentage points above the index. It is important to note the instability of the performances, with the gap between best and worst performance being 270.9 percentage points and three Test Runs being bellow the benchmark. The introduction of the EMA trading system improved the returns of all Test Runs, playing a especially important role in **2018** and **2020**. Figure 9 plots the various Test Runs for the **Growth Strategy**.

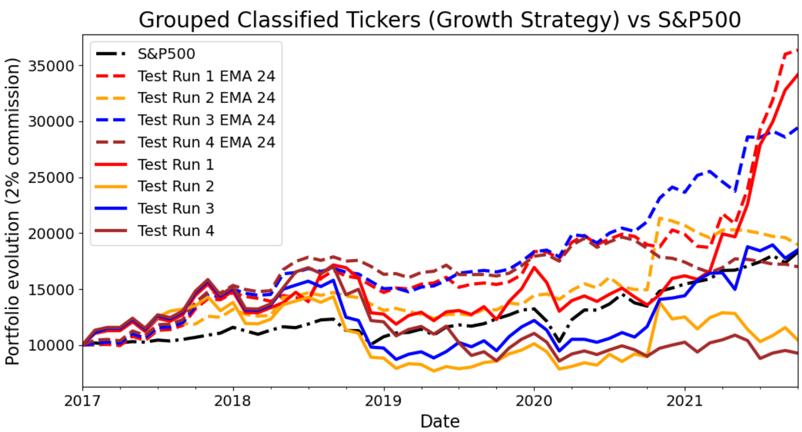

Figure 9: 2 Stock Portfolio evolution (Growth) vs the S&P500

Regarding the **Balanced Strategy** an improvement in results was verified, with the average compound returns being 278.2%, more than 3 times better than the SP&500. Although the gap between best and worst Test Run was of 450.1 percentage points, higher than that of the **Growth Strategy**, all performances were above the index. Barring **Test Run 3**, all Test Runs benefited from the introduc-

Table 3: ROI for all strategies for the 2SP including total compound returns and Sharpe Ratios

Strategy	Test Run	2017	2018	2019	2020	2021	Total	SR
Growth Strategy	1	48.1%	-13.8%	32.7%	-4.4%	111.0%	241.5%	0.69
	1 EMA	46.9%	0.0%	24.9%	8.4%	82.8%	263.6%	0.69
	2	38.1%	-36.1%	14.6%	23.6%	-16.5%	4.5%	0.64
	2 EMA	31.8%	-0.7%	10.2%	43.3%	-8.3%	89.7%	0.64
	3	50.7%	-35.5%	25.3%	18.1%	28.4%	84.9%	0.64
	3 EMA	49.8%	0.3%	22.0%	29.0%	24.4%	194.2%	0.64
	4	52.0%	-20.6%	-8.6%	-7.1%	-7.6%	-7.3%	0.62
	4 EMA	53.3%	6.5%	9.9%	-2.9%	-2.4%	69.9%	0.62
Balanced Strategy	1	60.4%	-6.8%	22.1%	102.7%	59.0%	488.5%	0.64
	1 EMA	44.0%	26.8%	33.2%	119.4%	22.8%	555.2%	0.64
	2	10.8%	-8.9%	-2.6%	75.6%	18.7%	105.1%	0.62
	2 EMA	5.0%	19.5%	7.3%	82.6%	15.6%	184.4%	0.62
	3	52.4%	3.8%	7.5%	57.7%	8.3%	190.5%	0.64
	3 EMA	20.2%	14.0%	5.8%	81.4%	-0.4%	161.9%	0.64
	4	45.8%	14.5%	18.7%	22.0%	17.5%	184.1%	0.63
	4 EMA	13.1%	26.8%	38.8%	94.8%	17.6%	356.1%	0.63
Value Strategy	1	46.7%	34.5%	52.6%	65.3%	13.8%	466.6%	0.72
	1 EMA	13.7%	47.7%	40.0%	62.7%	36.1%	420.5%	0.72
	2	32.1%	10.3%	16.0%	100.2%	17.8%	298.5%	0.77
	2 EMA	17.6%	20.6%	10.3%	89.5%	16.2%	244.6%	0.77
	3	52.0%	6.6%	52.6%	84.0%	49.4%	579.7%	0.78
	3 EMA	28.0%	11.9%	40.0%	74.2%	48.3%	418.0%	0.78
	4	52.0%	-18.6%	21.1%	111.1%	9.1%	245.4%	0.79
	4 EMA	26.7%	-0.8%	12.6%	88.9%	12.4%	200.5%	0.79
S&P 500	-	15.8%	-7.3%	23.2%	16.8%	18.5%	83.0%	N/A

tion of the EMA trading system. Figure 10 plots the various Test Runs for the **Balanced Strategy**.

Finally, for the **Value Strategy** the average compound ROI was of 359.2%, the best of all strategies. Comparing the best and worst Test Runs the difference was of 379.2 percentage points, although the worst Test Run was already 2.4 times better than the market. For this particular strategy the effect of the trading system was diminished, with its introduction being unable to improve any Test Runs, granted that it didn't severely hinder any particular Test Run either. Figure 11 plots the various Test Runs for the **Value Strategy**.

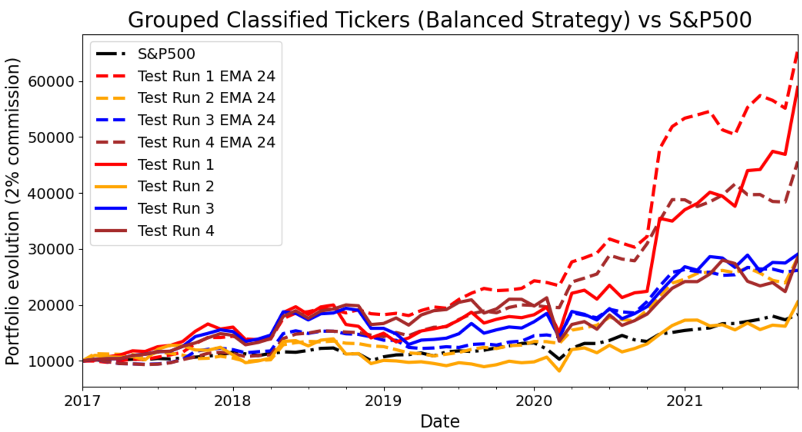

Figure 10: 2 Stock Portfolio evolution (Balanced) vs the S&P500

Overall the best strategy for this portfolio was the **Value Strategy**, followed up by the **Balanced Strategy** and the **Growth Strategy**. Given that the portfolio composition considered was the 2SP the differences between best and worst Test Run can be justified by high volatility, making the results inconsistent, especially when paired with the **Growth Strategy** that already trends towards unstable companies. The EMA trading system proved to be a useful tool, particularly when avoiding losses for more unstable strategies.

As table 3 shows the strategy that had the best SR values for all its Test Runs was the **Value Strategy**, indicating that the best strategy proved to also be the less risky. The **Balanced Strategy** was also considered the riskiest strategy, closely followed by the **Growth Strategy**.

5. Conclusions

In this work a GA based stock selection model paired with a trading system was created in order to take advantage of patterns found in the market. Three main strategies were analysed, a **Growth Strategy** that mainly focuses on the potential of companies, a **Balanced Strategy** that will take into consideration all kinds of companies and a **Value Strategy** that searches for companies that have strong and stable positions in the market. The companies considered were all inserted in the SP&500 and overall the training and testing period ranged from 31-01-2017 up to 31-10-2021. The portfolio considered for testing, applied to each strategy, was the 2SP, with the stocks either being equally distributed and updated every month or equally distributed but traded utilizing an EMA based trading system, although monthly stock updates were still applied. The results were overall extremely positive with most strategies being able to find great results when compared to the benchmark. Given that the portfolios generated are fairly simple these results were mainly due to the GA based stock selection model, proving its capabilities of finding good stocks. It is also important to highlight the impact of the EMA trading system, specially for the **Growth Strategy**, which was able to severely diminish losses during more volatile months.

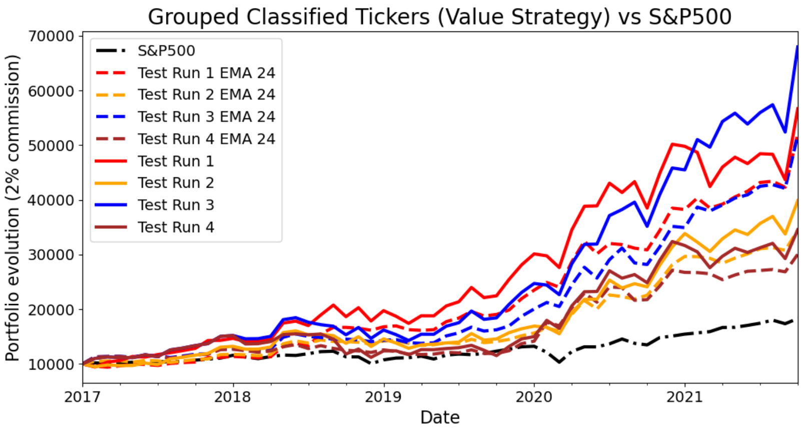

Figure 11: 2 Stock Portfolio (Value) evolution vs the S&P500

Overall this work was able to accomplish several objectives. Firstly data regarding SP&500 companies' financial statements was successfully acquired and utilized to calculate various fundamental indicators for multiple months. Then, utilizing a K-means clustering algorithm the performance of a company's fundamental indicators was classified for every end of the month in the data range, this done by contrasting with sector peers. Following this step a GA was taken advantage of to search for the importance that each indicator has in a company's performance, utilizing a previous three month price variation method as the fitness function. All these steps culminated in the creation of a stock selector that takes into consideration both the performance of fundamental indicators among industry peers and the importance that each fundamental indicator must have. This classifier was updated at the end of every month. Finally the stock selector was tested by forming various different portfolios and analysing the results for various strategies, even introducing a trading system based on the EMA.

Although the model created yielded good returns there is always room for improvement. Regarding future work, these are some suggestions that are believed to further improve the model developed: (1) Test the model with different types of evolutionary algorithms; (2) Explore different trading systems with the assistance of more technical indicators; (3) Apply different types of portfolio techniques like the Modern Portfolio Theory [16] in order to better optimize portfolio weights; (4) Explore the application of the GA Module developed in this thesis to find the worst companies instead of the best companies, opening shorting options; (5) Take into account different types of trading costs in order to get a better grasp of the overall performance; (6)

Explore different markets besides the SP&500; (7) Apply different methods of clustering, perhaps hierarchical, for the classification.

References

[1] K. Hrazdil, K. Trottier, and R. Zhang, "A comparison of industry classification schemes: A large sample study," *Economics Letters*, vol. 118, 2012.

[2] R. C. Cavalcante, R. C. Brasileiro, V. L.F. Souza, J. P. Nobrega, and A. L.I. Oliveira, "Computational intelligence and financial markets: A survey and future directions," *Expert Systems With Applications*, vol. 55, pp. 194–211, 2016.

[3] W. Jiang, "Applications of deep learning in stock market prediction: Recent progress," *Expert Systems With Applications*, vol. 184, 2021.

[4] Y. Hu, K. Liu, X. Zhang, L. Su, E. Ngai, and M. Liu, "Application of evolutionary computation for rule discovery in stock algorithmic trading: A literature review," *Applied Soft Computing*, vol. 36, pp. 534–551, 2015.

[5] E. Beyaz, F. Tekiner, X. Zeng, and J. A. Keane, "Comparing technical and fundamental indicators in stock price forecasting," 2018.

[6] A. Slowik and H. Kwasnicka, "Evolutionary algorithms and their applications to engineering problems," *Neural Computing and Applications*, vol. 32, 2020.

[7] S. Lim, M. Kim, and C. W. Ahn, "A genetic algorithm (ga) approach to the portfolio design based on market movements and asset valuations," *IEEE Access*, vol. 8, 2020.

[8] A. Silva, R. Neves, and N. Horta, "A hybrid approach to portfolio composition based on fundamental and technical indicators," *Expert Systems with Applications*, vol. 42, 2015.

[9] D. Cheong, Y. M. Kim, H. W. Byun, K. J. Oh, and T. Y. Kim, "Using genetic algorithm to support clustering-based portfolio optimization by investor information," *Applied Soft Computing*, vol. 61, 2017.

[10] B. C. Barroso, R. Cardoso, and M. K. Melo, "Performance analysis of the integration between portfolio optimization and technical analysis strategies in the brazilian stock market," *Expert Systems With Applications*, vol. 186, 2021.

[11] C.-F. Huang, "A hybrid stock selection model using genetic algorithms and support vector regression," *Applied Soft Computing*, vol. 12, 2012.

[12] X. Meng and X. Chen, "The investment models based on the fundamental and technical indicators of listed companies," *37th Chinese Control Conference*, 2018.

[13] D. Wu, X. Wang, and S. Wu, "Construction of stock portfolios based on k-means clustering of continuous trend features," *Knowledge-Based Systems*, vol. 252, 2022.

[14] M. Khedmati and P. Azin, "An online portfolio selection algorithm using clustering approaches and considering transaction costs," *Expert Systems with Applications*, vol. 159, 2020.

[15] T. R. C. C. da Costa, R. T. Nazário, G. S. Z. Bergo, V. A. Sobreiro, and H. Kimura, "Trading system based on the use of technical analysis: A computational experiment," *Journal of Behavioral and Experimental Finance*, vol. 6, pp. 42–55, 2015.

[16] H. Markowitz, "Portfolio selection," *The Journal of Finance*, vol. 7, 1952.