STOCK EXCHANGE PREDICTION USING IMPROVED GENETIC ALGORITHMS

Bianca V. L. Pereira
Breno Batista
Luiz E. H. Lopes
Felipe D. Araujo
Marta de O. Barreiros
bianca.pereira@aluno.uema.br
brenobatista@aluno.uema.br
luiz.lopes@aluno.uema.br
Maranhão State University, Paulo VI Campus, University City, Center for Technological Sciences
R. Paulo VI, s/n - São Cristovão, 65055-000, São Luis/Maranhão, Brazil

Abstract. Trading on the stock exchange is a fundamental part of a country's economic development, driving profits directly or indirectly from market transactions. However, the stock market is highly dynamic, stocks grow and fall as a result of changes in various parameters. Predicting the exact future of the stock market is not a simple process as it encompasses several factors needed to demonstrate Emory's current context. Some techniques have already been used to try to predict these changes in the market, but improvements are still needed to ensure greater accuracy in market betting, as most of them are high risk. Predicting the closing value of an asset is a difficult task because the asset's value is constantly changing and has a variety of parameters that influence its increase and decrease. In this work, a genetic algorithm was implemented in MATLAB to estimate the price of assets, trying to predict the closing value of stock market shares. A population of 21 possible solutions was initially designed using as a variable the value of IBM stock closing time during March 2019. The input variables were the closing date, closing value and the variation. The algorithm sought to predict closures in the first half of April 2019. Thus, in this work, the least variable stocks were chosen as the criterion for the suitability of the best-fit chromosomes, thus having a positive result with error rate less than 1 dollar. Highlighting the effects of diversification according to the number of assets and available capital for investment.

Keywords: Genetic Algorithms, Stock Exchange, Prediction, IBM.

1 Introduction

The stock market goes through constant fluctuations that make it difficult to know the right time to perform a buy or sell a stock. This oscillation is due to various factors such as changes in a company, policy, financial market situation, among other factors. According to Nascimento [1], there are commonly used techniques such as fundamental analysis and technical analysis, which seek to predict price changes based on a company's data and information analysis, and on techniques for reading and analyzing graphs.

The ability to predict the closing price on the stock market makes it possible to engage in profitable stock trading. With the development of information technology, new means for price prediction such as Data Mining, Genetic Algorithms and Neural Networks have been tested and improved with the purpose of applying them to the stock market. Recent work involving Genetic Algorithms has said that the most important factor in determining the best means of prediction is from the selection method.In Sable's work, Porwal and Singh [2], a comparison of 6 attributes is performed to determine the weight of the attribute that most contributes to the prediction of quotation through Genetic Algorithms. In Mao, Zhang and Fan's work [3], Genetic Algorithms are also used to determine which factor is most relevant for price prediction using multiple initial populations. While in the work of Xia, Sun, Zhao, Wang, Xiong, Gao, Li, and Yuan [4], Support Vector Machine (SVM) is used to classify stock market elements and determine the most suitable as a selection method to apply to the Genetic Algorithm. K-means along with Genetic Algorithms was also used in the work of Desokey, Badr and Hegazy [5] seeking the closest results to the actual stock price.

In this work, a genetic algorithm was implemented to estimate the price of assets, trying to predict the closing value of stock market shares, using as the selection criterion the closing price of stocks that suffered the smallest variation between previous quotations.

2 Genetic Algorithms

Genetic Algorithms (GAs) are part of the set of Evolutionary Algorithms and were developed based on Evolution and Genetic Heredity [1]. The evolutionary theory of Darwin (1858, 1859) and Wallace (1858) has in natural selection the device that clarifies the principle and heterogeneity of species. The theory of natural selection is differential conception because of fluctuations in the survival potential of the inhabitants of a species in a given environment. This follow-up may lead to a broadening of the proficient hereditary characteristics between a generation and its successor [6].In Table 1, follows the interconnection of biological elements to GAs.

Table 1. Interconnection of biological elements for GAs

Genetic Biology	Genetic Algorithms
Chromosome	Data Structure
Gene	Element occupying a given position of the data structure.
Crossover	Part exchange between data structures.
Mutation	Replacement of one or more genes.
Fitness (Fit)	Value that indicates an individual's quality as a solution to the problem.
Selection	Process that allows the survival and reproduction of individuals.
Genotype	Coding of a candidate solution.
Phenotype	Decoded value of chromosome

Thus, through this technique would evolve, there is a better search and optimization in solving a given problem. The algorithm consists of finding several answers and using the data obtained to get better agents. Having an initial population, through a selection, crossing and mutation criteria, this method aims to evolve the chromosomes trying to get as close as possible to the solution.

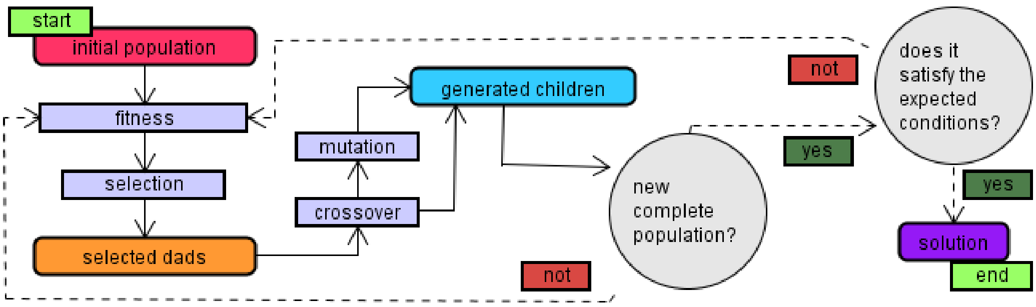


Figure 1. Standard cycle of an AG.

3 Data Set

The action chosen was from the International Business Machines Corporation (IBM). The dataset used for this prediction consists of data from March (day 1 to 29) from 2019. It was extracted from the Investing.com Web site and the attributes (phenotypes) provided were the closing date, the maximum value, the opening value, the minimum value, the variation and the volume. With a population of 21 values, as shown in Table 2, the algorithm sought to predict closures in the first half of April.

Table 2. Input data of the proposed genetic algorithm

Day	Closing Value	Change	Day	Closing Value	Change	Day	Closing Value	Change
01	139.20	0.77	04	138.43	0.55	05	137.88	0.40
06	136.98	0.65	07	135.36	1.18	08	135.09	0.20
11	137.71	1.94	12	138.28	0.41	13	138.56	0.20
14	138.79	0.17	15	139.43	0.46	18	140.21	0.56
19	140.49	0.20	20	139.60	0.63	21	141.44	1.32
22	139.45	1.41	25	139.18	0.19	26	140.22	0.75
27	139.24	0.70	28	139.92	0.49	29	141.10	0.84

All this data was stored in a Comma Separated Values (CSV) file, each line has the peculiarities of each closing value of the respective dates. However, for our algorithm only the closing value and its variation is required.

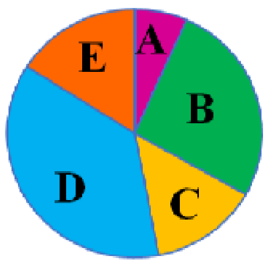
4 Methodology

Predicting the closing value of a stock is a complex activity because the value of the stock is constantly subject to fluctuations and has a number of factors that act on its volatility. In this work, the genetic algorithm was implemented in MATLAB to estimate the price of assets, trying to predict the closing value of stock market shares.

Start and Fitness: Initialization was done by allocating the closing values and the variations of actions in a vector for each element. Then, the adequacy of each chromosome was calculated based on the variation indicator of each one. Also in 2015, the moving average was the most widely used indicator by experts to analyze stock market trends and used in previous work as a criterion for chromosome suitability. However, due to changes in the Brazilian economy, stocks have changed widely over the past few years and the moving average has not been chosen as the most appropriate criterion for chromosomal fitness. Thus, in this work, the actions with the least variation were chosen as criteria for adequacy of the best aptitude chromosomes (Apt_{I_n}). Follows the aptitude formula that gives each individual (I_n) a fit value that is based on the percentage that their variation (Var_{In}) represents about the sum of the variations of all individuals $(\sum Var_{In})$. Since the greatest fit is attributed to the person with the least variation, it is subtracted 1 by the percentage change result.

$$Apt_{I_n} = 1 - \frac{Var_{In}}{\sum Var_{In}}. \tag{1}$$

Selection: Although genetic algorithms have several chromosome selection methods, the method chosen for this work was the roulette method, since it gives a higher probability of choosing that action that did not undergo sudden variations.



Individual	Fitness	Relative Fitness
A	2,27	0,078329883
B	7,44	0,256728778
C	3,87	0,133540373
D	10,4	0,358868185
E	5	0,172532781

Figure 2. Roulette Method - Selection Criteria.

Thus, in the roulette selection function, a random value is generated that if it is equal to or close to the fit of the chromosome, it is the chosen one. Remembering that the technique of selection favors those who have greater aptitude.

Crossover: The crossover to generate a new individual (I_{new}) occurred by the arithmetic mean between two individuals (I_n and I_m) selected.

$$I_{new} = \frac{(I_n + I_m)}{2}. \tag{2}$$

Mutation: It was developed with the goal of looking for the same individuals and applying the media between them (I_{before}) and five individuals (I_1, I_2, I_3, I_4 and I_5) aggregated at a random mutation rate (t_{mut}) that is added to the new individual (I_{after}).

$$I_{after} = t_{mut} + \frac{(I_{before} + I_1 + I_2 + I_3 + I_4 + I_5)}{6}. \tag{3}$$

5 Results

After generating 378 individuals (possible solutions in each execution/prediction), selecting the 11 best children of the genetic algorithm, for the first half of April, we obtained a positive result. We set the error rate limit to 1 dollar, that is, the values obtained that were 1 dollar higher or lower than the actual closing value were considered good results, so with 11 values, a complete prediction occurred within this value limit. We made three predictions with IBM stock, follows the tables and therefore the analyzes. In Table 3, below are the values computed for the 1st prediction.

Table 3. IBM 1st Action Value Prediction

Actual Value	Predicted Value	Difference	Actual Value	Predicted Value	Difference
143.3000	143.1620	0.1380	143.0000	143.1620	0.1620
143.6300	143.1620	0.4680	142.7800	143.6545	0.8745
143.2800	143.6545	0.3745	143.3900	143.6545	0.2645
142.1100	142.2468	0.1368	143.0200	142.2468	0.7732
143.7800	142.0050	1.7750	144.3500	142.0050	2.3450
143.9000	142.0050	1.8950			

In this 1st prediction, we observed that 8 out of 11 individuals had prediction values below 1 dollar. However, of these predicted values, 6 (55%) of them were below the established 0,5 dollar limit. Thus, the algorithm showed 73 % efficiency at the first prediction of the IBM stock exchange.

In Table 4 is the 2nd prediction of IBM stock values.

Table 4. IBM 2nd Action Value Prediction

Actual Value	Predicted Value	Difference	Actual Value	Predicted Value	Difference
143.3000	142.6064	0.6936	143.0000	142.6064	0.3936
143.6300	143.7707	0.1407	142.7800	143.7707	0.9907
143.2800	143.7707	0.4907	143.3900	143.7707	0.3807
142.1100	142.0539	0.0561	143.0200	142.0539	0.9661
143.7800	143.4219	0.3581	144.3500	143.4219	0.9281
143.9000	143.4219	0.4781			

In this 2nd prediction, we observed that all individuals have prediction values below 1 dollar. However, of these predicted values, 7 (64%)of them were below the established 0,5 dollar limit. Thus the algorithm presented 100% closing efficiency determined (1 dollar) for the 2nd IBM stock prediction.

Finally, Table 5 presents the values computed for the 3rd IBM action prediction.

Table 5. IBM 3nd Action Value Prediction

Actual Value	Predicted Value	Difference	Actual Value	Predicted Value	Difference
143.3000	142.8410	0.4590	143.0000	142.8410	0.1590
143.6300	142.8410	0.7890	142.7800	141.9994	0.7806
143.2800	141.9994	1.2806	143.3900	143.4670	0.0770
142.1100	143.4670	1.3570	143.0200	143.4670	0.4470
143.7800	143.4670	0.3130	144.3500	143.4670	0.8830
143.9000	143.4670	0.4330			

In this 3rd prediction, we observed that 9 out of 11 individuals had prediction values below 1 dollar. However, of these predicted values, 6 (55%) of them were below the established 0,5 dollar limit. Thus, the algorithm showed 82% efficiency in IBM's third action prediction.

It is observed that the proposed algorithm was able to fulfill the objective established to approximate the closing values of the IBM stock exchange. Moreover, it is noted that with each interaction of the algorithm, the prediction improved, bringing the values closer to a category below 0,5 dollars difference. Thus, it is understood that genetic algorithms are great alternatives for valuation of stock market closing values. New interactions are still being established to improve predicted values and, as a result, to approximate the exact day's closing value.

6 Final Considerations

Thus, in this work it was observed that using as a criterion the variation to determine the aptitude and the method of selection via roulette we obtained the expected result for the error rate lower than 1 dollar. Highlighting the effects of diversification according to the number of assets and available capital for investment. As also noted, for the 0,5 dollar difference we had about fifty percent of the well-predicted individuals. In the future, if we use other parameters, besides the variation to calculate the fitness, we will undoubtedly have even more effective results.

References

[1] T. P. Nascimento. (2015). Um Serviço Baseado em Algoritmos Genéticos para Predição da Bolsa de Valores. Dissertation (Master Degree on Electrical Engineering) - Universidade Federal do Maranhão. São Luís, p. 8. 2015.

[2] S. Sable, A. Porwal and U. Singh. Stock price prediction using genetic algorithms and evolution strategies. *International Conference of Electronics, Communication and Aerospace Technology*, 2017, Coimbatore, India. IEEE, 2017. Available from: https://ieeexplore.ieee.org/document/8212724. Access in: 15 June 2019.

[3] Y. Mao, Z. Zhang and D. Fan. Hybrid feature selection based on improved genetic algorithm for stock prediction. *6th International Conference on Digital Home (ICDH)*, 2016, Guangzhou, China. IEEE, 2017. Available from: https://ieeexplore.ieee.org/document/8039111. Access in: 15 June 2019.

[4] T. Xia, Q. Sun, A. Zhao, S. Wang, S. Xiong, S. Gao, J. Li and Q. Yuan. Improving the Performance of Stock Trend Prediction by Applying GA to Feature Selection. *IEEE 8th International Symposium on Cloud and Service Computing (SC2)*, 2018, Paris, France. IEEE, 2018. Available from: https://ieeexplore.ieee.org/document/8567383. Access in: 15 June 2019.

[5] E. S. Desokey, A. Badr and A. F. Hegazy. Enhancing Stock Prediction Clustering Using K-Means with Genetic Algorithm. *13th International Computer Engineering Conference (ICENCO)*, 2017, Cairo, Egypt. IEEE, 2018. Available from: https://ieeexplore.ieee.org/document/8289797. Access in: 15 June 2019.

[6] M. T. da Silva. C. M. D. Santos. A historical analysis of natural selection: from Darwin-Wallace to the Evolution extended synthesis. *Revista Amazonica* 2015. Available from: https://periodicos.ufpa.br/index .php/revistaamazonia/article/viewFile/2122/2634. Access in 16 June 2019.