IOSR Journal of Economics and Finance (IOSR-JEF)
e-ISSN: 2321-5933, p-ISSN: 2321-5925.Volume 10, Issue 6 Ser. I (Nov. – Dec 2019), PP 52-55
www.iosrjournals.org

Application of Genetic Algorithms for Stock Market Prediction

Nildson de Castro Pinheiro Mello[1], Danilo José dos Santos Costa[1], Higor Gabriel Rodrigues Lima[1], Lucas Diógenes Pinheiro Barroso[1], Matheus Pereira da Silva[1], Marta de Oliveira Barreiros[1]

[1]*Department of Computer Engineering, State University of Maranhão-UEMA São Luiz/MA – Brazil*

Abstract: *The market for investment in the stock exchange draws attention because of the possibility of very profitable returns, so it is important to have a procedure to estimate possible outcomes. In this scenario, it is interesting to use computation in the prediction of stock market results, since estimating the future price of the stock exchange is a complex process, with the aid of programming it is possible to create an algorithm that helps in the making of decisions. In this context, the present article aims to present an application capable of predicting information for investment in the stock exchange. A python script was created based on genetic algorithms, making use of real data of distinct assets and marketed in the São Paulo city (Brazil) stock exchange (BOVESPA) extracting characteristics and relevant information of the market and a certain period that are used like parameters of entry in the statistic to predict the value of future closure. The result showed an estimate of the future behavior of the asset in question, as closing values. The proposed algorithm serves as an indicator to aid in decision making in the stock market.*

Keywords*: Market, Genetic Algorithm, investment, active.*

Date of Submission: 25-10-2019 Date of acceptance: 09-11-2019

I. Introduction

The stock exchange is a market where shares of publicly traded companies, which may be public or private, and other securities are traded. The stock market is seen as a high potential investment by many shareholders, but even for the most seasoned and knowledgeable of the changes that can happen in the market, the risk of loss exists to the same extent as the possibility of gains.[1]

In everyday life, many technologies are used, such as smartphones, consumer electronics and payment systems. No different on the stock market, to make a higher probability of success, technologies can be employed that help to make a predictive perspective on the end result on the stock market. These technologies are well regarded in the market as they help in organizing and agility in the school of future moves in the stock market.[1]

Many methods have already been used for stock market prediction, however, the use of genetic algorithms has become a feasible possibility, as it is feasible to achieve a result that can be of great help in the decision making of an individual who participates in this extreme market. , where results may vary in a short time.

Thiago Nascimento [2], in its application for stock market prediction based on genetic algorithms using, among others, Petrobras actions for the tests, obtained a percentage of 53.1% of results with differences of up to 50 cents, however, for the fifth day after this percentage drops to 19.6%.

This work proposes the creation of a method based on genetic algorithms capable of predicting stock closing values on the stock market.

II. Material And Methods

Genetic algorithms can be said as one of the paradigms of the so-called evolutionary computation. They are basically an iterative process in which a given population of solutions to a given problem goes through several generations. Depending on the selection methodology, the best individuals survive until they find a satisfactory solution at the end of the generations.[3]

Data for the use of the algorithm and for testing were obtained from the São Paulo Stock Exchange (BOVESPA) through the yahoo finance [4]. Data are from November 2018 through the end of May 2019 for several different actions (such as OI, Petrobras, among others). Being the last seven valid days of May used for testing.

The algorithm was developed in python language version 3.7 in conjunction with the Pycharm IDE. To aid development, the following libraries were used: numpy, random, csv, and matplotlib. The numpy library makes it easy to perform numeric calculations, random lets you create random numbers, csv lets you work with csv spreadsheets, and matplotlib lets you generate graphs. [5]

The algorithm iterates a population of 100 individuals for 400 generations. Each individual has seven positions with each position being the expected closing value for the respective day (position 0 represents the first day of the forecast, position 1 represents the subsequent day, and so on).

The selection method chosen was selection by ranking, in which the 100 individuals are sorted according to fitness and the crossing occurs by this sort and the next generation is all formed by crossings (similar to sorting by selection) applying mutations according to with the probability.

The mutation method selected is from a single randomly chosen position, its value changed by one of the possible values collected from the base.

The crossover method selected was a single point, where the part preceding this point from parent 1 will be joined to the part following that point from parent 2, forming child 1. The remaining parts are joined together forming child 2. It is After this crossover the mutation may occur depending on the probability employed.

To calculate the fitness of individuals, mathematical concepts used in trend indicators are required [1]. Trend indicators are calculations that exemplify the way the market progresses. The simple moving average (Eq. 1), for example, is a trend indicator whose calculation is simply the sum of prices divided by the number of days. However, the mathematical calculation itself carries with it a difference between the predicted value and the actual value. [2]

$$mms_n = \frac{\sum_{i=1}^{n} preço_i}{n} \quad (1)$$

where mmsn is the simple moving average and n represents the number of days.

The prediction average (Eq. 2) is calculated from the 25-day moving averages of each database day, as 100 closing values are used, the moving average of each day is calculated, at the end it is divided this value per 100, thus obtaining the prediction mean.

$$mp_{100} = \frac{\sum_{i=1}^{100} mms_{25}i}{100} \quad (2)$$

where mp100 is the 100 day prediction average and mms25i is the 25 day simple moving average of the day i in question.

The individual's fitness is the sum of the differences in modulus of the predicted values and the prediction averages, so the lower the fitness, the better the individual is. A weighting is added for each individual position to prioritize starting positions.

The algorithm output for each execution is a comparative graph between the predicted values and the current values, and a terminal table displaying the error of each prediction. The reference currency used is the BRL (Brazilian Real).

III. Resultand Discussion

A total of 100 executions were established for evaluation of the proposed algorithm using the OI database [4]. The best results of the proposed algorithm are shown in table 1.

Table 1. Executionswiththe OI base.

Real Values							
DAY	DAY 1	DAY 2	DAY 3	DAY 4	DAY 5	DAY 6	DAY 7
VALUE	1.49	1.49	1.53	1.54	1.48	1.49	1.47
TEST 1							
VALUE	1.51	1.51	1.51	1.51	1.51	1.51	1.51
ERROR	1.32%	1.32%	1.32%	1.99%	1.99%	1.32%	2.65%
TEST 2							
VALUE	1.51	1.51	1.51	1.51	1.51	1.51	1.51
ERROR	1.32%	1.32%	1.32%	1.99%	1.99%	1.32%	2.65%
TEST 3							
VALUE	1.51	1.51	1.51	1.51	1.51	1.51	1.51
ERROR	1.32%	1.32%	1.32%	1.99%	1.99%	1.32%	2.65%
TEST 4							
VALUE	1.51	1.49	1.51	1.51	1.49	1.51	1.49
ERROR	1.32%	0.0%	1.32%	1.99%	0.67%	1.32%	1.34%
TEST 5							
VALUE	1.49	1.52	1.51	1.51	1.51	1.49	1.49
ERROR	0.0%	1.97%	1.32%	1.99%	1.99%	0.0%	1.34%

It is noticeable that the proposed algorithm was able to approximate the actual value of the action, and thus, the percentage error (Eq. 3) for this action was significantly low, even reaching the value of closing the day, in certain executions.

$$\%error = \frac{|approx - exact|}{exact} * 100 \quad (3)$$

For the database used in the period in question, the prediction average was 1.5055 BRL. Analyzing the output graphs shows that the results are very close to the prediction mean, however, the mathematical calculation itself carries an error in relation to the actual value of the stock closing. Thus, during the tests there were no results in which all predicted values were equal to the actual ones.

To evaluate the proposed algorithm in values above 20 reais, we used the Petrobras database, in the same interval of days and with the same number of executions (total of 100), in order to evaluate the best closing values. The top ten results for this database are presented in table 2.

Table 2. Executionswiththe Petrobras base.

Real Values							
DAY	DAY 1	DAY 2	DAY 3	DAY 4	DAY 5	DAY 6	DAY 7
VALUE	25.84	26.14	26.24	26.80	26.50	26.15	25.55
TEST 1							
VALUE	26.00	26.08	26.08	25.88	26.03	26.08	25.93
ERROR	0.62%	0.23%	0.61%	3.55%	1.81%	0.27%	1.47%
TEST 2							
VALUE	26.08	26.03	26.03	26.03	26.08	26.05	25.80
ERROR	0.92%	0.42%	0.81%	2.96%	1.61%	0.38%	0.97%
TEST 3							
VALUE	26.05	26.08	26.08	26.05	26.03	26.08	26.35
ERROR	0.81%	0.23%	0.61%	2.88%	1.81%	0.27%	3.04%
TEST 4							
VALUE	26.03	26.03	26.03	26.03	26.05	25.80	26.03
ERROR	0.73%	0.42%	0.81%	2.96%	1.73%	0.97%	1.84%
TEST 5							
VALUE	26.03	26.05	26.00	26.03	26.08	26.03	26.05
ERROR	0.73%	0.35%	0.92%	2.96%	1.61%	0.46%	1.92%
TEST 6							
VALUE	26.08	26.03	26.00	25.88	26.03	26.05	26.35
ERROR	0.92%	0.42%	0.92%	3.55%	1.81%	0.38%	3.04%
TEST 7							
VALUE	26.00	26.08	26.00	25.88	26.05	25.93	26.05
ERROR	0.62%	0.23%	0.92%	3.55%	1.73%	0.85%	1.92%
TEST 8							
VALUE	26.08	26.03	26.03	26.03	26.08	25.93	26.03
ERROR	0.92%	0.42%	0.81%	2.96%	1.61%	0.85%	1.84%
TEST 9							
VALUE	26.08	26.03	26.00	26.03	26.03	25.93	26.35
ERROR	0.92%	0.42%	0.92%	2.96%	1.81%	0.85%	3.04%
TEST 10							
VALUE	26.00	26.03	26.00	25.88	26.08	25.80	25.80
ERROR	0.62%	0.42%	0.92%	3.55%	1.61%	0.97%	0.97%

The prediction average in this period was 26.1164 BRL and the algorithm sought to minimize the error in relation to this value, but the prediction average itself is very far from the correct one, finding up to 90 cents of difference between the predicted and the real value.

Table 2 shows that for the first 3 days the percentage errors are below 1% and from day 4 the error tends to increase. Through the sample variance formula (Eq. 4) applied to the results of table 2 we arrived at the data presented in table 3.

whereσ^2 is the sample variance, *n*isthenumberof samples (in the case, 10) and$\bar{x}$ representes the arithmetic mean of the predicted values.

Table 3. Sample variance.

Real Values							
DAY	DAY 1	DAY 2	DAY 3	DAY 4	DAY 5	DAY 6	DAY 7
$\bar{X}$	26.043	26.047	26.04	25.972	26.054	25.968	26.074
σ^2	0.001141	0.0005 01	0.000925	0.00567 6	0.0005 04	0.0103 16	0.0404 44

Because it is a genetic algorithm, after each execution we have different results. The days with the highest variance are those with the most distant results. From the table you can see that in the first three days we have greater constancy of results and these days have errors below 1% representing values between 10 and 20 cents.

Relatively low variance was obtained every day (the highest found was 0.040444), which exemplifies the accuracy of the algorithm.

Day 4 is the one that has the greatest difference between the real value and the prediction mean, however, presenting a low variance, this fact shows that the accuracy of the approach is dependent on the oscillation level of the analyzed asset. The results show that the proposed solution behaves better in the first days analyzed, in which there was good precision and accuracy.

It is important to point out that the stock market is surrounded by several variables and that the proposed approach uses closing values and does not foresee sudden changes in the market.

IV. Conclusion

In the present article, programming practices were applied to obtain an estimate of stocks in the stock market, the proposed solution was elaborated as another consultation parameter for individuals active in the stock market.The results show the accuracy and precision of the algorithm for the first days predicted in the test bases. Summarizing the results, it can be concluded that the answer is only an estimate that behaves close to the prediction average. The errors obtained were percentage low during the tests, which allows the use of the script as an aid to guide the user's action plan.

References

[1]. Sanchez, “Bolsa de Valores,” Ecotec, 2011.
[2]. T. P. Nascimento. Um ServiçoBaseadoemAlgoritmosGenéticos para Predição de Bolsa de Valores. Dissertação de Pós-GraduaçãoemEngenharia de Eletricidade. Universidade Federal do Maranhão. São Luis. 2015.
[3]. E. M. Gallego Rendón, Ramón A.; Escobar Zuluaga, Antonio H.; Toro Ocampo, “Algoritmogenético,” in Técnicasmetaheurísticas de optimización, 2008.
[4]. Yahoo. Yahoo Finance - Business Finance, Stock Market, Quotes, News. Disponívelem: <https://finance.yahoo.com/>. Acessoem: 2 jun. 2019
[5]. Python Software Foundation Python 3.7.4rc1 documentation. Disponívelem<https://www.python.org/doc/>. Acessoem: 28 mai. 2019.

Nildson de Castro Pinheiro Mello " Application of Genetic Algorithms for Stock Market Prediction. " IOSR Journal of Economics and Finance (IOSR-JEF) , vol. 10, no. 6, 2019, pp. 52-55.