Genetic Algorithm Portfolio Optimization with Stock Return Forecast Comparison

Anonymous Author(s)*

ABSTRACT

This paper presents a comparison study between two portfolio optimization approaches based on Genetic Algorithm. Both approaches rely on the GA as the portfolio optimizer. On the other hand, each approach uses as input for the optimization stock return forecasts based on different models: a naïve predictor and Facebook Prophet. Each model was used in other to forecast one week ahead stock return. Both approached relied on 10 years of weekly stock returns in training and 2 years of weekly updates in a test period.

CCS CONCEPTS

• **Computing methodologies** → *Heuristic function construction*; *Machine learning approaches*.

KEYWORDS

Genetic Algorithms, Facebook Prophet, Stock Return Forecasting, Portfolio Optimization

ACM Reference Format:

Anonymous Author(s). 2022. Genetic Algorithm Portfolio Optimization with Stock Return Forecast Comparison. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Selecting the optimal allocation for an investment portfolio is a widely faced problem with innumerous approaches, one of them being a mixed strategy that combines stock return forecast with genetic algorithms optimization [5].

The first task would be to select how many, and witch assets would compose the portfolio in other to maximize its return while also maintain its risk accounted for. It is well know that a diverse portfolio, with weakly correlated assets is an efficient strategy to minimize its risks, usually expressed in terms of the portfolio volatility **[aqui cabe uma citação sobre diversificação de portifolio]**.

This paper has arbitrarily chosen 20 assets that are traded on the São Paulo Stock Exchange (BOVESPA) being one of them a risk-free asset called SELIC, to represent the investor's possibility to apply on fixed income. A Genetic Algorithm ([4],[6]) was implemented to find the optimum weights for each of the 20 assets, based on its expected returns. Prior to that, two time series forecast models were implemented in order obtain the expected return one week ahead for each asset. It is stablished that predicting stock return time series in the short term is a difficult task and, because of that, two models were implemented. First, as a benchmark model, it was designed a Naïve Predictor based on the expected return of a Random Walk Model. Secondly, the Facebook Prophet [7] was used for the same task. The reason to implement two predictors is to compare its results.

All models were trained during a 10-year period, from 2010 until 2019 and tested on a 2-year period, from 2020 until the end of 2021. As a benchmark to be beaten, the BOVESPA Index was chosen, since all the assets select are traded on it. No fees or transaction costs are assumed, and this paper does not aim to be used as a tool for individual investors, given its academic nature.

The next section presents how both the Genetic Algorithm and the stock return models were implemented and how the results were computed. Furthermore, the results of the two approached are presented and compared. Lastly, there is a section designed for conclusions and possible next steps to improve the methodology and results from this work.

2 METHODOLOGY

In order to give the optimizer a diverse range of assets, from different markets, the following 20 were manually selected:

- Banks: BBDC4, BBAS3, ITUB4, SANB11
- Insurance: SULA11
- Energy: ELET3, CPLE6, ENBR3
- Gas: PETR4
- Non Cyclical Consumption: ABEV3, JBSS3, WEGE3, CSAN3
- Mining and Steel: USIM5, VALE3, CSNA3, GGBR4
- Communications: VIVT3, TIMS3
- Risk-Free: SELIC

With each asset price in hands, the return series were computed as follows:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$$

where P_t is the asset adjusted close price and ln is the natural logarithm.

The Naïve Predictor is based on a Random Walk Process, which can be expressed as:

$$r_t = \phi r_{t-1} + \eta_t$$

where η_t is a random shock, usually is comes from a $Normal(0, \sigma^2)$ distribution, and $|\phi| = 1$. From this model, it is possible to obtain the expected return given the past of the series as:

$$E[r_t] = E[r_{t-1} + \eta_t] = E[r_{t-1}] + E[\eta_t] = E[r_{t-1}] = r_{t-1}$$

That means that next week expected return is exactly the observed price at the current week.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Since it is a challenging task to implement relevant stock return series models, the Naïve Predictor is used as a reference and as a first approach to give the Genetic Algorithm the expected returns in order to find the optimal weights. Furthermore, is common practice to include the Random Walk Model in forecast comparisons, as done in [2].

The second approach implemented used the Facebook Prophet Model for time series forecasting [7]. Even tough this model was not designed for stock return series, research has been done in this field. See [3] and [1].

An individual model was trained for each asset during a 10-year period with the following hyper-parameters:

- Flat Growth (no trend)
- 5 change point
- Daily, Weekly and Year Seasonality
- Brazilian Holidays Calendar

Following the training period, the model was used to forecast each asset return during a 2-year period, from January 2020 until December 2021. It was used a rolling window approach with length of 12 weeks and always forecasting only one week ahead.

Once all the expected returns were obtained from each model, those forecasted series were passed as input to the Genetic Algorithm, so that it could be used to determine the optimal weight for each asset every week. The chromosomes are composed of 20 genes, which represents the weights to be optimized.

The objective function to be maximized is the portfolio return, which can be expressed as:

$$max_w \left\{ \sum_{i=1}^{20} r_{t,i} w_{t,i} \right\}$$

The next step when implementing a Genetic Algorithm is to define its restrictions. Only three restrictions were given:

- No weight can be negative (notice that the GA can choose not to include a given asset)
- Weight sum must be 1
- Portifolio volatility must be less or equal to 15%

It is important to notice that the only time any risk measure is considered is in this restriction.

After stablished the basics elements of the Genetic Algorithm, the training period was used to experiment different parameters, so that the set of parameters that generates the highest cumulative return over the end of the training period would be chosen. The parameters used in those experiments were:

- Population: 50,100
- Crossover Probability: 0.3, 0.5
- Mutation Probability: 0.05, 0.1
- Number of Generations: 20,50

And the fixed evolutionary parameters are:

- Window Size: 12 weeks
- Gaussian Mutation with $\mu = 0$, $\sigma^2 = 1$ and probability = 0.05
- Two Point Crossover
- Selection Tournament of size 3

Once selected the best set of parameters during the training period, the GA was used in the test period in order to compute the portfolio cumulative return over two years. Once again, the algorithm was re-estimated every week in order to compute the optimum weights for that week. The results obtained for using both the Naïve Predictor and the Prophet Model are presented in the next section.

Table 1: Cummulative Returns at the end of test period

BOVESPA	GA+Naive	GA+Prophet
40%	39%	146%

3 RESULTS

Since the main goal of this paper is to compare the results obtained by the portfolio optimization, no result from the individual asset return series will be presented. It is important to mention that this paper was not aiming to produce a return series forecasting model that completely captures the return's internal structure, since this is much challenging. This paper aims to check if, even if the individual stock return predictions are not perfect, it is possible to give the optimization algorithm enough information about the future of each asset so that it can find the optimum weight distributions over time.

With that being said, table 1 gives the cumulative return from the BOVESPA index, used as benchmark, from the portfolio that used the Naïve Predictor and from the portfolio that used the Prophet Model. Furthermore, figures 1 and 2 present the cumulative return from each portfolio, using the Naïve Predictor and the Prophet Model, respectively. Note that the dashed line is fixed in one, while the whole line represents the portfolio cumulative return divided by the BOVESPA index cumulative return. Those figures were chosen as a direct way to infer how many times the optimum portfolio's return was higher than the benchmark.

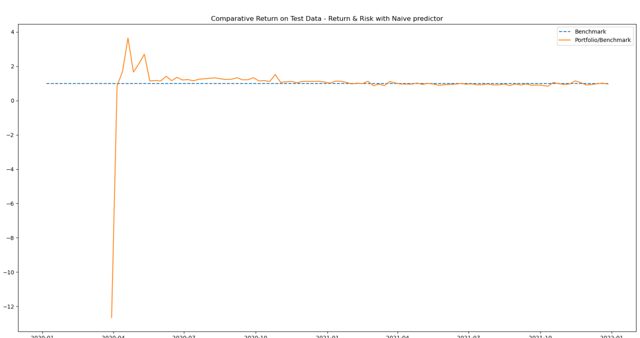

Figure 1: GA+Naive Cummulative Return against BOVESPA Cummulative Return as benchmark

4 CONCLUSION

As the results shows, there is a significant difference between the two approaches implemented. It is important to clarify, once more, that the only difference between those approaches is the method used to compute the individual assets expected returns. The portfolio optimization was done in the exact same way.

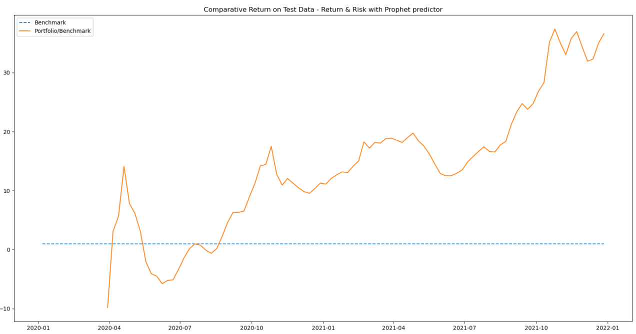

Figure 2: GA+Prophet Cummulative Return against BOVESPA Cummulative Return as benchmark

With that being said, using the Naïve predictor to compute the expected returns seems to result in no difference when compared to the benchmark, which means that the optimum portfolio is no better than the index itself, since the cumulative return of both of them at the end of the test period is the same. On the other hand, when the genetic algorithm received as inputs the expected returns from the Prophet Model, it was able to generate an optimum portfolio that far surpass the index cumulative return, giving a final value almost 30 times bigger.

This result does not mean that the approach developed in this paper is to be implemented in real investment scenarios. However, it seems to indicate that combining returns predictions with portfolio optimization is possibly a powerful strategy when implementing investing mechanisms.

As last, it is important to mention that there are several aspects of this paper that should be further explored, such as: compute a measurement of the portfolio risk over time, develop better stock return models to pass as input and implement different and more complex objective functions.

REFERENCES

[1] Wen-Xiang Fang, Po-Chao Lan, Wan-Rung Lin, Hsiao-Chen Chang, Hai-Yen Chang, and Yi-Hsien Wang. 2019. Combine Facebook Prophet and LSTM with BPNN Forecasting financial markets: the Morgan Taiwan Index. In *2019 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*. 1–2. https://doi.org/10.1109/ISPACS48206.2019.8986377

[2] Márcio GP Garcia, Marcelo C Medeiros, and Gabriel FR Vasconcelos. 2017. Real-time inflation forecasting with high-dimensional models: The case of Brazil. *International Journal of Forecasting* 33, 3 (2017), 679–693.

[3] Anusha Garlapati, Doredla Radha Krishna, Kavya Garlapati, Nandigama mani Srikara Yaswanth, Udayagiri Rahul, and Gayathri Narayanan. 2021. Stock Price Prediction Using Facebook Prophet and Arima Models. In *2021 6th International Conference for Convergence in Technology (I2CT)*. 1–7. https://doi.org/10.1109/I2CT51068.2021.9418057

[4] D. E. Goldberg. 1989. *Genetic Algorithms in Search, Optimisation, and Machine Learning*. Addison-Wesley.

[5] Juan G Lazo Lazo, Marco Aurélio C Pacheco, and Marley Maria R Vellasco. 2002. Portfolio selection and management using a hybrid intelligent and statistical system. In *Genetic Algorithms and Genetic Programming in Computational Finance*. Springer, 221–238.

[6] Zbigniew Michalewicz. 1996. GAs: Why Do They Work? In *Genetic Algorithms+ Data Structures= Evolution Programs*. Springer, 45–55.

[7] Sean J Taylor and Benjamin Letham. 2018. Forecasting at scale. *The American Statistician* 72, 1 (2018), 37–45.