



CIÊNCIA DA COMPUTAÇÃO

**ALGORITMOS GENÉTICOS PARA OTIMIZAÇÃO DE
PORTFÓLIO DE INVESTIMENTO**

GABRIEL GARCIA PAVAN

**Rio Verde - GO
2021**

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
GOIANO - CAMPUS RIO VERDE
CIÊNCIA DA COMPUTAÇÃO

**ALGORITMOS GENÉTICOS PARA OTIMIZAÇÃO DE PORTFÓLIO
DE INVESTIMENTO**

GABRIEL GARCIA PAVAN

Trabalho de Curso apresentado ao Instituto Federal
Goiano – Campus Rio Verde, como requisito parcial
para a obtenção do Grau de Bacharel em Ciência da
Computação.

Orientador: Prof. Dr. Márcio Antônio Ferreira Belo
Filho

Rio Verde - GO
2021

Sistema desenvolvido pelo ICMC/USP
Dados Internacionais de Catalogação na Publicação (CIP)
Sistema Integrado de Bibliotecas - Instituto Federal Goiano

Pva Pavan, Gabriel
 Algoritmos Genéticos para Otimização de Portfólio
 de Investimento / Gabriel Pavan; orientador Márcio
 Antônio Ferreira Belo Filho. -- Rio Verde, 2021.
 49 p.

 TCC (Graduação em Bacharelado em Ciências da
 Computação) -- Instituto Federal Goiano, Campus Rio
 Verde, 2021.

 1. Teoria Moderna do Portfólio. 2. Programação
 Linear Inteira Mista. 3. Algoritmo Genético. I.
 Antônio Ferreira Belo Filho, Márcio , orient. II.
 Título.



SERVIÇO PÚBLICO FEDERAL
MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA GOIANO



INSTITUTO FEDERAL
Goiano

Repositório Institucional do IF Goiano - RIIF Goiano
Sistema Integrado de Bibliotecas

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR PRODUÇÕES TÉCNICO-CIENTÍFICAS NO REPOSITÓRIO INSTITUCIONAL DO IF GOIANO

Com base no disposto na Lei Federal nº 9.610/98, AUTORIZO o Instituto Federal de Educação, Ciência e Tecnologia Goiano, a disponibilizar gratuitamente o documento no Repositório Institucional do IF Goiano (RIIF Goiano), sem ressarcimento de direitos autorais, conforme permissão assinada abaixo, em formato digital para fins de leitura, download e impressão, a título de divulgação da produção técnico-científica no IF Goiano.

Identificação da Produção Técnico-Científica

- | | |
|--|---|
| <input type="checkbox"/> Tese | <input type="checkbox"/> Artigo Científico |
| <input type="checkbox"/> Dissertação | <input type="checkbox"/> Capítulo de Livro |
| <input type="checkbox"/> Monografia – Especialização | <input type="checkbox"/> Livro |
| <input checked="" type="checkbox"/> TCC - Graduação | <input type="checkbox"/> Trabalho Apresentado em Evento |
| <input type="checkbox"/> Produto Técnico e Educacional - Tipo: _____ | |

Nome Completo do Autor: Gabriel Garcia Pavan

Matrícula: 2017102201910340

Título do Trabalho: Algoritmos Genéticos para Otimização de Portfólio de Investimento

Restrições de Acesso ao Documento

Documento confidencial: ☒ Não ☐ Sim, justifique: _____

Informe a data que poderá ser disponibilizado no RIIF Goiano: 14/08/2021

O documento está sujeito a registro de patente? ☐ Sim ☒ Não

O documento pode vir a ser publicado como livro? ☐ Sim ☒ Não

DECLARAÇÃO DE DISTRIBUIÇÃO NÃO-EXCLUSIVA

O/A referido/a autor/a declara que:

1. o documento é seu trabalho original, detém os direitos autorais da produção técnico-científica e não infringe os direitos de qualquer outra pessoa ou entidade;
2. obteve autorização de quaisquer materiais inclusos no documento do qual não detém os direitos de autor/a, para conceder ao Instituto Federal de Educação, Ciência e Tecnologia Goiano os direitos requeridos e que este material cujos direitos autorais são de terceiros, estão claramente identificados e reconhecidos no texto ou conteúdo do documento entregue;
3. cumpriu quaisquer obrigações exigidas por contrato ou acordo, caso o documento entregue seja baseado em trabalho financiado ou apoiado por outra instituição que não o Instituto Federal de Educação, Ciência e Tecnologia Goiano.

Rio Verde, 12/08/2021.

Gabriel Garcia Pavan
Assinado Eletronicamente

Ciente e de acordo:

Márcio Antônio Ferreira Belo Filho
Assinado Eletronicamente

Documento assinado eletronicamente por:

- **Gabriel Garcia Pavan, 2017102201910340 - Discente**, em 12/08/2021 23:05:21.
- **Marcio Antonio Ferreira Belo Filho, PROFESSOR ENS BASICO TECN TECNOLOGICO**, em 12/08/2021 23:04:04.

Este documento foi emitido pelo SUAP em 12/08/2021. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifgoiano.edu.br/autenticar-documento/> e forneça os dados abaixo:

Código Verificador: 299144

Código de Autenticação: a7b008b268



INSTITUTO FEDERAL GOIANO
Campus Rio Verde
Rodovia Sul Goiana, Km 01, Zona Rural, None, RIO VERDE / GO, CEP 75901-970
(64) 3620-5600



SERVIÇO PÚBLICO FEDERAL
MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA GOIANO

Ata nº 24/2021 - GEPTNM-RV/DE-RV/CMPRV/IFGOIANO

ATA DE DEFESA DE TRABALHO DE CURSO

Ao(s) 29 dia(s) do mês de julho de 2021, às 16 horas e 15 minutos, reuniu-se a banca examinadora de forma virtual pela plataforma Google Meet composta pelos docentes: Márcio Antônio Ferreira Belo Filho (orientador), André da Cunha Ribeiro (membro), Douglas Cedrim Oliveira (membro), para examinar o Trabalho de Curso intitulado "Modelos de Otimização de Portfólio de Investimento" do(a) estudante Gabriel Garcia Pavan, Matrícula nº 2017102201910340 do Curso de Bacharelado em Ciência da Computação do IF Goiano – Campus Rio Verde. A palavra foi concedida ao(a) estudante para a apresentação oral do TC, houve arguição do(a) candidato pelos membros da banca examinadora. Após tal etapa, a banca examinadora decidiu pela APROVAÇÃO do(a) estudante. A banca alterou o título do trabalho de conclusão para "Algoritmos Genéticos para Otimização de Portfólio de Investimento". Ao final da sessão pública de defesa foi lavrada a presente ata que segue assinada pelos membros da Banca Examinadora.

(Assinado Eletronicamente)

Márcio Antônio Ferreira Belo Filho

Orientador(a)

(Assinado Eletronicamente)

André da Cunha Ribeiro

Membro

(Assinado Eletronicamente)

Douglas Cedrim Oliveira

Membro

Observação:

() O(a) estudante não compareceu à defesa do TC.

Documento assinado eletronicamente por:

- Andre da Cunha Ribeiro, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 29/07/2021 18:48:31.
- Douglas Cedrim Oliveira, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 29/07/2021 18:48:26.
- Marcio Antonio Ferreira Belo Filho, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 29/07/2021 18:46:26.

Este documento foi emitido pelo SUAP em 29/07/2021. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifgoiano.edu.br/autenticar-documento/> e forneça os dados abaixo:

Código Verificador: 294770

Código de Autenticação: 53e424d4af



INSTITUTO FEDERAL GOIANO
Campus Rio Verde
Rodovia Sul Goiana, Km 01, Zona Rural, None, RIO VERDE / GO, CEP 75901-970
(64) 3620-5600

Aos meus pais e a minha irmã.

AGRADECIMENTOS

Agradeço primeiro a Deus por ter me mantido na trilha certa durante este projeto de pesquisa com saúde e forças para chegar até o final.

Sou grato à minha família pelo apoio que sempre me deram durante toda a minha vida.

Deixo um agradecimento especial ao meu orientador pela oportunidade, incentivo e dedicação do seu tempo ao meu projeto de pesquisa.

Também quero agradecer ao Instituto Federal de Educação, Ciência e Tecnologia Goiano - Campus Rio Verde e a todos os professores do meu curso pela elevada qualidade do ensino oferecido.

*“Investir em conhecimento rende sempre os
melhores juros.” (Benjamin Franklin)*

RESUMO

Nem sempre possuir um patrimônio em dinheiro é sinônimo de proteção financeira a longo prazo. Determinadas moedas podem se desvalorizar e com isso investí-las torna-se uma boa opção. Além de proteger, o investidor pode obter lucro, ou seja, aumentar a quantidade de capital inicialmente investido. Com o avanço tecnológico, o número de oportunidades de investimentos aumentou drasticamente, fazendo com que o investidor encare duas perguntas: “Em quais ativos devo investir? E quanto devo investir nos mesmos?”. Diante disso, diversos investidores e pesquisadores visam desenvolver modelos com o uso de ferramentas matemáticas/computacionais para solucionar tais perguntas. Markowitz (1952) cita que a primeira pergunta se relaciona mais com o comportamento e crença do investidor, enquanto diz que para a segunda pode-se utilizar mais facilmente de conceitos estatísticos para solução. O presente trabalho visa responder às perguntas, trazendo uma solução para o investidor, inserindo características reais em modelos teóricos. Além disso, propomos uma meta-heurística Algoritmo Genético para solucionar os mesmos modelos com características reais embarcadas, visto que muitos possuem uma complexidade computacional elevada. O algoritmo proposto apresentou bons resultados em tempo computacional reduzido.

Palavras-chave: Teoria Moderna do Portfólio, Programação Linear Inteira Mista, Algoritmo Genético.

ABSTRACT

PAVAN, Gabriel Garcia. **Genetic Algorithms for Investment Portfolio Optimization**. 2021. 32 f. Trabalho do Curso – Bacharelado em Ciência da Computação, Instituto Federal de Educação, Ciência e Tecnologia Goiano - Campus Rio Verde. Rio Verde - GO, 2021.

Cash equity does not mean long-term financial protection. Some currencies can depreciate and thus investing them becomes a good option. In addition to protecting, the investor can make a profit, that is, increase the amount of capital invested. With technological progress, the number of investment opportunities has increased dramatically, making investors face two questions: “In which assets should I invest? And how much should I invest in them? ”. Therefore, investors and researchers aim to develop models using mathematical/computational tools to solve such questions. Markowitz (1952) says that the first question is more related to the investor’s behavior, while for the latter statistical concepts can be used for solution. The present work aims to answer the questions, bringing a solution to the investor, embedding real features in theoretical models. Furthermore, we propose a Genetic Algorithm meta-heuristic to solve the same models with real embedded features, as long as they have a high computational complexity. The proposed algorithm presents good results in reduced computational time.

Keywords: Modern Portfolio Theory, Mixed Integer Linear Programming, Genetic Algorithm.

LISTA DE FIGURAS

| | |
|---|----|
| Figura 1 – Representação gráfica de um problema linear. A região factível está representada em cinza. Fonte: Cormen et al. (2009) | 11 |
| Figura 2 – Distribuições de pesos para portfólio de exemplo. Fonte: O autor. . . . | 13 |
| Figura 3 – Distribuições de pesos para portfólio de exemplo. Fonte: O autor. . . . | 13 |
| Figura 4 – Distribuições de pesos para portfólio de exemplo. Fonte: O autor. . . . | 22 |
| Figura 5 – Distribuições de pesos para portfólio de exemplo. Fonte: O autor. . . . | 23 |
| Figura 6 – Média de risco por instâncias testadas. Fonte: O Autor. | 26 |
| Figura 7 – Quantidade de soluções factíveis geradas por instância testada. Fonte: O Autor. | 26 |
| Figura 8 – Comparação entre riscos até diferença absoluta de 20%. Fonte: O Autor. | 27 |
| Figura 9 – Média de tempo computacionais por classe de instância. Fonte: O Autor. | 28 |
| Figura 10 – Convergência média das instâncias de soluções factíveis. Fonte: O Autor. | 28 |

LISTA DE TABELAS

| | |
|---|---|
| Tabela 1 – Principais artigos e suas metodologias abordadas, no presente trabalho | 5 |
|---|---|

LISTA DE ABREVIATURAS E SIGLAS

| | |
|----------|---------------------------------------|
| GA | Algoritmo Genético |
| CSV | <i>Comma-Separated Values</i> |
| MAD | <i>Mean Absolute Deviation</i> |
| MG | <i>Mean-Gini</i> |
| MV | Modelo Média-Variância |
| OPL | <i>Optimizer Programming Language</i> |
| Semi-MAD | <i>Semi-Mean Absolute Deviation</i> |

LISTA DE ALGORITMOS

| | |
|---|----|
| Algoritmo 1 – Crossover | 18 |
| Algoritmo 2 – Operação de Mutação 1 | 19 |
| Algoritmo 3 – Operação de Mutação 2 | 19 |
| Algoritmo 4 – Heurística Construtiva. | 21 |

SUMÁRIO

| | |
|--|-----------|
| 1 – INTRODUÇÃO | 1 |
| 2 – REVISÃO BIBLIOGRÁFICA | 3 |
| 2.1 Modelagem de Otimização | 7 |
| 2.2 Modelo de Média-Variância (MV) | 7 |
| 2.3 Modelo de Desvio Absoluto Médio (MAD) | 8 |
| 2.4 Características Reais | 8 |
| 2.4.1 Limitante de Investimento | 9 |
| 2.4.2 Cardinalidade | 9 |
| 3 – PROBLEMA ESTUDADO | 10 |
| 3.1 Programação Linear Inteira Mista | 10 |
| 3.2 Exemplo de Problema Estudado | 12 |
| 4 – Solução Meta-Heurística | 15 |
| 4.1 Algoritmo Genético | 15 |
| 4.2 Algoritmo Proposto | 16 |
| 4.2.1 Operação de Seleção de Pais | 17 |
| 4.2.2 Operação de <i>Crossover</i> | 17 |
| 4.2.3 Operação de Mutação | 18 |
| 4.2.4 Método Principal | 20 |
| 4.2.5 Exemplo | 22 |
| 5 – RESULTADOS COMPUTACIONAIS | 24 |
| 5.1 Conjunto de Instâncias | 24 |
| 5.2 Resultados Computacionais | 25 |
| 6 – CONCLUSÃO E PERSPECTIVA | 29 |
| Referências | 30 |
| ANEXO A – Código fonte Algoritmo Genético | 32 |

1 INTRODUÇÃO

Qualquer objeto que possua valor e que possa ser negociado no mercado financeiro é denominado como ativo. Exemplos de ativos são: ações, moedas, títulos públicos etc. Um conjunto de ativos a serem investidos é denominado portfólio ou carteira de investimento. Ativos podem ser analisados separadamente com base no seu histórico ou em seu comportamento futuro. Sendo assim, ativos com um bom histórico recente tendem a ser mais atrativos e desejados.

Mansini et al. (2015) citam:

Para avaliação da performance de um ativo olhamos o preço de cotação ou também preço de mercado. Este corresponde ao valor econômico o qual cada ativo é comprado ou vendido. A estimação de cotações futuras é uma tarefa difícil, pois a cotação é o resultado de diversos componentes e fatores tais como a força e estabilidade de uma companhia, a lucratividade e projeção futura do setor que a mesma opera. [...] Apesar disso, há alguns fatores que possam facilitar essa previsão, por exemplo: se os bancos entrarem em crise, os ativos do setor bancário tendem a perder valor; já se a uma empresa automobilística criar um novo modelo de sucesso, provavelmente seus ativos tendem a se valorizar. [...] Há também outras variáveis macroeconômicas que são relacionadas a economia nacional e internacional, ou a políticas fiscais e monetárias que podem ser observadas pois afetam também os valores de um ativo de uma companhia, por exemplo: PIB (Produto Interno Bruto). [...] Outra variável que podemos citar é a inflação, esta impactando no crescimento esperado da economia, e as taxas de câmbio, que não podem ser ignoradas caso se deseja operar no mercado global.

Cotações de ativos de mesma classe/setor tendem a mover-se no mesmo sentido. Sendo assim, pode-se obter vantagens em investir em diferentes setores ao mesmo tempo. A diversificação permite a redução da volatilidade de um portfólio (que pode ser considerada um dos fatores que compõem o risco). Fundos de investimento e *Private equity* são exemplos de portfólios que praticam a diversificação.

Mansini et al. (2015) enumeram o processo de construção de um portfólio por meio de uma abordagem científica em 5 fases:

1. Identificação dos ativos disponíveis;
2. Coleta de informações, e métodos de estimação e hipóteses da performance dos ativos disponíveis;
3. Seleção de modelos de otimização, o qual com base nas estimativas obtidas anteriormente, geram um portfólio;
4. Reavaliação dos modelos adotados por meio de *feedback* às fases anteriores;
5. Implementação do portfólio.

O problema de otimização de portfólio consiste em determinar quais ativos de um conjunto de ativos deve-se investir e quanto investir em cada um destes ativos, com

o objetivo de maximizar o retorno do investimento e minimizar o risco envolvido. Este portfólio deve incluir características impostas pelo mercado financeiro e particularidades do investidor. Atualmente, mesmo um investidor iniciante possui uma quantidade elevada de possibilidades de investimento.

Markowitz (1952) apresentou a Teoria Moderna do Portfólio, o que lhe rendeu um prêmio Nobel de Economia por seu trabalho pioneiro. Além disso, esta teoria incentivou outros trabalhos como Yitzhaki (1982), Konno e Yamazaki (1991), em que visam utilizar de diferentes medidas estatísticas para se calcular o risco de um portfólio (o quanto pode-se perder financeiramente), afim de reduzir a complexidade computacional de seus modelos comparados ao elaborado por Henry Markowitz.

Determinados modelos são trazidos para uma realidade mais próxima do investidor ao adicionar características reais, tais como custo de gerência, percentual de investimento por setor, quantidade de ativos, percentuais de investimentos. Entretanto, alguns aspectos tornam os modelos matemáticos mais complexos e computacionalmente mais difíceis de se resolver. Assim, investidores e pesquisadores estão sempre a buscar abordagens mais eficientes de solução. O uso de meta-heurísticas possibilita encontrar boas soluções em um tempo computacional satisfatório. Trabalhos como Woodside-Oriakhi, Lucas e Beasley (2011) utilizam meta-heurísticas para achar boas soluções para o problema.

Diante disto, o presente trabalho irá abordar o problema de otimização de portfólio de investimento, levando em consideração características reais do problema como retorno mínimo, quantidade de ativos e percentual de capital investido por ativo. Um Algoritmo Genético para o problema de otimização de portfólio de investimento é proposto a investidores, que possuam um conhecimento básico de linha de comando e programação, para que possam utilizar para se obter uma solução robusta e flexível em tempo computacional satisfatório. O algoritmo proposto será comparado com a resolução do modelo matemático por um *solver* comercial com tempo de execução limitado.

O presente trabalho está organizado da seguinte forma: o Capítulo 2 apresenta uma revisão de literatura de modo a deixar o leitor a par dos trabalhos mais relevantes para este trabalho; o Capítulo 3 aborda detalhadamente o problema de pesquisa; já o Capítulo 4 explica a metodologia adotada para resolução do problema proposto; no Capítulo 5 são comparados e discutidos os resultados computacionais dos métodos abordados; por fim, o Capítulo 6 realiza a conclusão do presente estudo e aponta perspectivas futuras de pesquisa.

2 REVISÃO BIBLIOGRÁFICA

A relação do homem com o dinheiro é algo muito antigo. Atualmente, o poder financeiro possibilita que o homem possua segurança e bem-estar. O capital de uma pessoa está distribuído em vários ativos: propriedades, moedas, joias, títulos de empresas e governos, dentre outros. Vários destes ativos podem ser negociados por meio de bolsas de valores.

Segundo Pinheiro (2014):

A palavra bolsa, no sentido comercial e financeiro tem origem no fim do século 13 do nome da família de nobres belgas, os Van der Buerse, cujo brasão de armas era três bolsas de pele, simbolizando honradez e méritos por sua atuação na área mercantil. [...] A primeira Bolsa oficial surgiu em 1531 em Antuérpia na Bélgica, designada por Nieuwe Beurse. Nesta bolsa, realizavam-se negócios especulativos influenciados por boatos que afetavam a evolução dos preços. [...] A partir do século XIX, as bolsas restringiram sua atuação aos mercados de capitais (títulos e valores mobiliários) e, à medida que surgiam os mercados de títulos representativos de mercadorias (commodities), foram criados locais específicos para sua negociação (bolsas de mercadorias). Segundo a Federação Mundial de Bolsas, durante as últimas décadas as bolsas de valores mobiliários passaram a desempenhar um novo papel de grande importância no sistema financeiro internacional, qualitativamente diferente de qualquer outro visto desde a Segunda Guerra Mundial. Em termos quantitativos, os mercados que operam por meio de bolsas regulamentadas cresceram em uma escala nunca antes imaginada, o que lhes atribuiu papel ativo e grande responsabilidade no centro da economia mundial.

Ativos possuem valores variáveis de acordo com sua importância histórica e futura. Ao fazer um investimento um investidor busca ativos cujo valor aumente com o tempo obtendo rendimentos e/ou lucro na venda deste ativo. Com o passar do tempo diversos investidores e pesquisadores começaram a estudar e desenvolver novas teorias para criação e gestão de seus investimentos. Dentre diversas teorias podemos destacar a Teoria Moderna do Portfólio, introduzida em Markowitz (1952). Esta teoria assume que o investidor deseja minimizar o risco para um dado retorno mínimo. Esta relação de retorno e risco pode ser representada graficamente por uma curva conhecida como Fronteira Eficiente, no qual cada coordenada indicada representa um portfólio eficiente bem diversificado.

Markowitz (1952) cita:

O processo de seleção de um portfólio pode ser dividida em dois estágios. O primeiro começa com a observação e experiência e finaliza com a crença sobre o desempenho futuro de títulos disponíveis. Já o segundo começa a crença no desempenho futuro e finaliza com a escolha do portfólio.

Markowitz (1952) formaliza matematicamente os conceitos de retorno esperado e risco, define diversificação e propõe um modelo de otimização no qual utiliza como medida de risco a variância (*Mean-Variance* ou MV).

Yitzhaki (1982) propõem uma nova abordagem denominada *Mean-Gini* (MG) a qual se utiliza de Diferenças de Média de Gini como medida de risco. Esta medida consiste na variabilidade das diferenças dos retornos em diferentes cenários. A utilização da *Mean-Gini* permite a redução da complexidade computacional encontrada no modelo Markowitz, pois trata-se de um modelo matemático linear enquanto de Markowitz é quadrático.

Jaaman e Lam (2012) realiza uma comparação entre os modelos MV e MG. Para realização dos testes foram selecionados 29 ativos presentes no Mercado de Ações da Malásia. Como resultado, o portfólio MV apresentou uma composição de 8 ativos, enquanto o portfólio MG apresenta 7 ativos e uma performance melhor.

Em Konno e Yamazaki (1991), os autores propõem um novo modelo utilizando outra medida de risco, o Desvio Absoluto Médio (*Mean Absolute Deviation* - MAD), e o aplicam na Bolsa de Valores de Tokyo. Alegam que o modelo de Markowitz não tem sido usado extensivamente na sua forma original para construção de portfólio em larga escala e uma das razões mais significativas é a dificuldade computacional com a resolução de problemas quadráticos em larga escala com uma matriz densa de covariância.

Em Feinstein e Thapa (1993), os autores reformularam o modelo apresentado em Konno e Yamazaki (1991), resultando em uma medida de risco denominada Semi-MAD. Ao invés de usar o MAD, esta medida de risco ignora os desvios das taxas acima do retorno esperado e contabiliza apenas os desvios de performance inferior, visto que a preocupação do investidor se concentra em portfólios com performance inferior.

Speranza (1996) propõe um modelo matemático o qual também utiliza o Desvio Absoluto Médio como medida de risco, além de considerar características reais tais como custos de transação e lotes de transação. A autora utiliza ativos do mercado de ações italiano e mostra que seu modelo gera soluções em tempo satisfatório para instâncias até 20 ativos. Para instâncias maiores, a autora propõe uma heurística que utiliza soluções do modelo relaxado e as factibiliza.

Apesar disso, determinadas medidas de risco quando implementadas utilizando-se de grandes conjuntos de ativos podem fazer com que o modelo possua um alto custo computacional para se resolver. Com isso, diversos pesquisadores vem a utilizar de heurísticas e meta-heurísticas como forma de obter uma boa solução em tempo satisfatório.

Em ChiangLin (2006) o autor reformula o modelo de Markowitz de forma que se possa definir uma taxa de aversão ao risco. Além disso, apresenta três variações do modelo utilizando características reais tais como: lote de transação mínima e custos de transação. Como forma de se buscar a solução utiliza-se da meta-heurística Algoritmo Genético. Apresenta como resultado uma eficiência, quando não igual, maior por parte do modelo de Markowitz com taxa de aversão ao risco sem características reais.

Chang, Yang e Chang (2009) abordam o problema de otimização de portfólio com diferentes medidas de risco utilizando-se de Algoritmo Genético como meta-heurística e compara com o modelo de Markowitz com restrição de cardinalidade. Como medidas de risco

são escolhidas três baseadas na média-variância de Markowitz, sendo elas: Semi-Variância, Desvio Absoluto Médio e Variância Assimétrica. Como resultado pode-se observar que para as três medidas é aconselhável que não se invista na carteira mais que um terço da quantidade do conjunto de ativos desejados, caso contrário o portfólio é dominado por outros portfólios com quantidade menores de ativos.

Já em Woodside-Oriakhi, Lucas e Beasley (2011), os autores abordam três meta-heurísticas para o problema de otimização de portfólio com cardinalidade restrita além de realizar mesclagens entre elas, sendo as meta-heurísticas: Algoritmo Genético, Busca Tabu e *Simulated Annealing*. Como modelo base é escolhido o modelo de Markowitz. Como resultado observa-se que *Simulated Annealing* não é competitivo como o Algoritmo Genético e a Busca Tabu, que tiveram performances parecidas. Quanto ao tempo computacional para todas as meta-heurísticas o tempo não foi excessivo, sendo o maior deles de aproximadamente 15 minutos.

Kamali (2014), em seu artigo, realiza a comparação entre a meta-heurística de Enxame de Partículas e Algoritmo Genético. A autora utiliza como medida de risco a variância e mostra que a metodologia de Enxame de Partículas gera portfólios com menores riscos e também tempos de execuções menores.

Sefiane e Benbouziane (2012) em seu trabalho explicam sobre o uso de Algoritmos Genéticos em aplicações de finanças, especialmente no problema de otimização de portfólio. Como experimento para comprovar a eficiência e confiabilidade da meta-heurística, propõem três algoritmos com diferentes operadores de *crossover* (um ponto, dois pontos e média aritmética), sem operador de mutação e característica real de limitação de investimento. Com foco nos operadores de cruzamento, concluem que, dentre os três, a média aritmética apresenta um portfólio com maior retorno e menor risco.

Abaixo apresentamos uma tabela que contém os principais artigos a respeito de otimização de portfólio de investimento. Na mesma também é explicado brevemente as características de cada trabalho, bem como o método de solução usados.

Tabela 1 – Principais artigos e suas metodologias abordadas, no presente trabalho

| Autor(es) | Características | Método de Solução |
|-------------------------|--|----------------------|
| Markowitz (1952) | Variância como medida de risco, modelo quadrático | <i>Não informado</i> |
| Konno e Yamazaki (1991) | Desvio Absoluto Médio como medida de risco, modelo pode ser quadrático ou linear | Modelo Matemático |
| Yitzhaki (1982) | Médias De Gini com medida de risco | Modelo Estatístico |

| | | |
|--|--|--------------------|
| Jaaman e Lam (2012) | Comparação entre Médias de Gini e Variância, ambos em sua forma quadrática | Modelo Matemáticos |
| Feinstein e Thapa (1993) | Reformulam o modelo MAD, dando origem a medida de risco Semi-MAD | Modelo Matemático |
| Speranza (1996) | Utiliza Desvio Absoluto Médio como medida de risco e custos de transação e lotes de transação como características reais | Modelo Matemático |
| ChiangLin (2006) | Utiliza de Algoritmo Genético para resolver modelo Média-Variância com características reais (lote de transação e custos de transação) | Heurístico |
| Chang, Yang e Chang (2009) | Compara as medidas de risco Variância, Semi-Variância e Desvio Absoluto Médio, fazendo uso de Algoritmo Genético para resolver o modelo | Heurístico |
| Woodside-Oriakhi, Lucas e Beasley (2011) | Utiliza de 3 meta-heurísticas (Algoritmo Genético, Busca Tabu e <i>Simulated Annealing</i> para resolução do modelo Média-Variância com restrição de cardinalidade | Heurístico |
| Kamali (2014) | Comparação de meta-heurísticas Exame de Partículas e Algoritmo Genético | Heurístico |
| Sefiane e Benbouziane (2012) | Utiliza de Algoritmos Genéticos com diferentes operadores de <i>crossover</i> | Heurístico |

2.1 Modelagem de Otimização

Seja N um conjunto de ativos e um determinado capital a ser investido neste conjunto de ativos. Devemos tomar a decisão de quais ativos investir e quais as porcentagens do capital a ser investido nestes ativos. As decisões visam minimizar o risco envolvido e maximizar a expectativa de retorno. Além disso devem atender a restrições de investimento, relacionadas ao investidor, aos ativos e ao mercado, por exemplo.

Seja o vetor x um portfólio, em que cada posição $x_i, i \in N$ denota o peso a ser investido no ativo i , $0 \leq x_i \leq 1$. Seja $\mu(x)$ a taxa de retorno médio do portfólio x , a qual o investidor pretende maximizar. Outro objetivo do investidor é minimizar o risco, o qual denotamos por $\varrho(x)$ para o portfólio x . Assim, trata-se de um problema multiobjetivo, com dois objetivos interrelacionados:

$$\min\{ [-\mu(x), \varrho(x)] : x \text{ é um portfólio factível.} \} \quad (1)$$

Apesar de teoricamente a formulação de otimização de portfólio ser multiobjetiva, modelos na prática, como Markowitz (1952), retiram o objetivo secundário da função objetivo e o passam como restrição. Pode-se observar este comportamento nos modelos apresentados nas subseções abaixo.

As seções seguintes abordam modelos matemáticos propostos para resolução do problema de otimização de portfólios. Os modelos a seguir buscam minimizar apenas o risco, sujeitos a um retorno mínimo denotado por μ_0 .

2.2 Modelo de Média-Variância (MV)

Markowitz (1952) foi o idealizador da Teoria Moderna do Portfólio, propondo medidas para o retorno e o risco. O modelo proposto por Markowitz é dado por:

$$\text{minimizar : } \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} x_i x_j \quad (2a)$$

$$\text{s.a. : } \sum_{j=1}^n x_j \mu_j \geq \mu_0 \quad (2b)$$

$$\sum_{j=1}^n x_j = 1 \quad (2c)$$

$$x_j \geq 0, \quad j \in N \quad (2d)$$

A equação (2a) referência a função objetivo, a qual visa minimizar o risco, ou seja, a covariância entre os retornos dos ativos. A variável σ_{ij} corresponde a covariância entre as taxas de retornos dos ativos i e j . A restrição (2b) impõe que o retorno de portfólio deve ser maior ou igual que um retorno mínimo μ_0 . Quanto a (2c), esta garante que 100% do capital seja investido. Já a restrição (2d) define o domínio das variáveis x_j .

Pode-se observar que o modelo matemático de Markowitz é um problema de programação quadrática, com uma função objetivo quadrática, duas restrições lineares e uma variável contínua não-negativa.

2.3 Modelo de Desvio Absoluto Médio (MAD)

Konno e Yamazaki (1991) utilizam o Desvio Absoluto Médio como medida de risco e propõem um modelo matemático linear para o problema de otimização de portfólio. Desta forma, o problema é computacionalmente mais tratável que a forma quadrática de Markowitz. Além disso, os autores mostraram que as soluções encontradas pelas duas formulações são similares.

A modelagem utilizada pelos autores para o MAD é dada por:

$$\text{minimizar : } \frac{1}{T} \sum_{t=1}^T \left| \sum_{j=1}^n a_{jt} x_j \right| \quad (3a)$$

$$\text{s.a. : } \sum_{j=1}^n x_j \mu_j \geq \mu_0 \quad (3b)$$

$$\sum_{j=1}^n x_j = 1 \quad (3c)$$

$$x_j \geq 0, \quad j \in N \quad (3d)$$

A equação (3a) consiste na função objetivo a qual visa minimizar o risco, ou seja, o desvio absoluto médio. A variável a_{jt} consiste na diferença entre o retorno esperado do ativo j com o retorno do ativo j no tempo t . Sendo T o número total de períodos. As restrições (3b), (3c) e (3d) correspondem às restrições (2b), (2c) e (2d) do modelo de Markowitz.

2.4 Características Reais

Para o problema de otimização de portfólio outras características podem ser consideradas, no intuito de promover restrições de mercado como taxas e impostos, ou preferências do investidor como quantidades mínimas e máximas de ativos a serem investidos ou mesmo valores de investimentos mínimos e máximos.

Queiroz et al. (2019) em seu trabalho utiliza do Modelo de Desvio Absoluto Médio e Modelo Média-Gini junto de características reais e os comparam a um indicador de retorno médio (*benchmark*). Mostram que para uso destas restrições é necessário uma variável binária z_j que denota a existência ou não de investimento no ativo j . Nas subseções abaixo listamos as características reais definidas no modelo utilizado em nosso trabalho.

2.4.1 Limitante de Investimento

Os limitantes de investimento definem porcentagens mínima e máxima do capital que pode ser investido por ativo. O limitante inferior do ativo j é dado por l_j e o limitante superior é dado por u_j . A restrição é apresentada conforme abaixo:

$$l_j z_j \leq x_j \leq u_j z_j, \quad j \in N \quad (4)$$

2.4.2 Cardinalidade

As restrições de cardinalidade definem as quantidades mínimas K_{min} e máximas K_{max} de ativos do portfólio. Tal medida pode ser utilizada no intuito de ajustar a diversificação do portfólio.

$$K_{min} \leq \sum_{j=1}^n z_j \leq K_{max} \quad (5)$$

3 PROBLEMA ESTUDADO

Dados os modelos matemáticos discutidos e as características reais abordadas no capítulo anterior, vamos considerar neste trabalho o Desvio Absoluto Médio (MAD) como medida de risco, por possuir menor complexidade computacional. O modelo matemático a ser utilizado é definido abaixo:

$$\text{minimizar : } \frac{1}{T} \sum_{t=1}^T \left| \sum_{i=1}^n a_{it} x_i \right| \quad (6a)$$

$$\text{s.a. : } \sum_{j=1}^n x_j \mu_j \geq \mu_0 \quad (6b)$$

$$\sum_{j=1}^n x_j = 1 \quad (6c)$$

$$l_j z_j \leq x_j \leq u_j z_j, \quad j \in N \quad (6d)$$

$$K_{min} \leq \sum_{j=1}^n z_j \leq K_{max} \quad (6e)$$

$$x_j \geq 0, \quad j \in N \quad (6f)$$

$$z_j \in \{0,1\}, \quad j \in N \quad (6g)$$

$$(6h)$$

A função objetivo (6a) minimiza o risco, conforme o MAD. O portfólio x está sujeito às restrições de retorno mínimo (6b), investimento de 100% do capital (6c), limitantes de investimento (6d), restrições de cardinalidade (6e) e do domínio das variáveis x_j e z_j (6f) e (6g).

3.1 Programação Linear Inteira Mista

Os problemas de Programação Linear Inteira Mista são problemas matemáticos cuja função objetivo e suas restrições são lineares, podendo suas variáveis serem inteiras ou reais. A função objetivo pode ser de minimização ou de maximização. Há diversos algoritmos que podem ser utilizados para resolução destes problemas, tais como o método *Simplex* e *Branch-And-Bound*.

Um dos métodos que podemos utilizar consiste no método *Simplex*, proposto em 1947 por George Dantzig e cuja a história é relatada pelo próprio autor em Dantzig (1990). Este algoritmo se inicia em um vértice da região factível do problema. Em seguida, verifica se este vértice é uma solução ótima. Caso falhar, um conjunto de vértices vizinhos são testados e um deles é escolhido, a depender principalmente da melhoria da função objetivo.

Este procedimento se repete até que no vértice corrente não haja vizinho melhor. Abaixo, pode-se visualizar um exemplo de problema linear e sua representação gráfica:

$$\text{minimizar : } x_1 + x_2 \quad (7a)$$

$$\text{s.a. : } 4x_1 - x_2 \leq 8 \quad (7b)$$

$$2x_1 + x_2 \leq 10 \quad (7c)$$

$$5x_1 - 2x_2 \geq -2 \quad (7d)$$

$$x_1, x_2 \geq 0 \quad (7e)$$

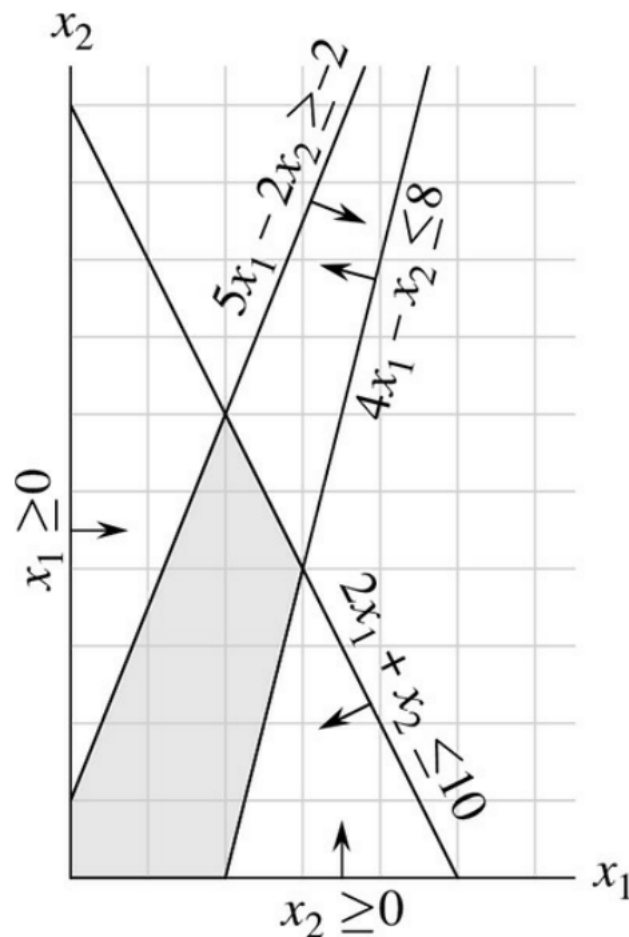


Figura 1 – Representação gráfica de um problema linear. A região factível está representada em cinza. Fonte: Cormen et al. (2009)

Já o algoritmo *Branch-And-Bound* utiliza da técnica dividir para conquistar. O algoritmo enumera todas as possíveis soluções do problema a ser resolvido, armazenando soluções parciais denominadas de subproblemas em uma estrutura de árvore. Nós não explorados de uma subárvore gerado pelo particionamento do espaço de solução em regiões menores podem ser resolvidos recursivamente e são utilizadas regras para reduzir as regiões

do espaço de busca que são consideradas subótimas. Uma vez que toda a árvore é explorada, é retornada a melhor solução encontrada pela busca (MORRISON et al., 2016).

No contexto de pesquisa operacional/otimização, *Solver* consiste em um software para resolução de problemas de programação matemática. Tais problemas podem ser a minimização ou maximização de uma dada função objetivo sob um conjunto de restrições. Há diversos *solvers* no mercado, tais como a função *Solver* do Microsoft Excel, Gurobi, LINDO e o IBM ILOG CPLEX.

O *solver* IBM ILOG CPLEX atua na resolução de problemas de programação matemática: problemas de programação linear inteira mista, problemas de programação quadráticos convexos e não-convexos e problemas de restrições quadráticos convexos. Alguns dos algoritmos encontrados dentro da caixa de ferramenta para resolução destes problemas são: *Branch And Bound*, *Simplex* e *Dual-Simplex*. A linguagem de programação utilizada é a *Optimizer Programming Language* (OPL). O software também conta com uma camada de modelagem a qual provê interfaces para linguagens C, C++, C#, Java e Python. Além disso, também é possível conectar com Microsoft Excel e MATLAB.

3.2 Exemplo de Problema Estudado

Como exemplo podemos citar um modelo com duas instâncias de otimização de portfólio de investimento de 10 ativos relacionados ao setor de tecnologia. Os 10 ativos em questão escolhidos foram: Amazon (AMZN), Apple (APPL), Facebook (FB), Google (GOOG), Microsoft (MSFT), Netflix (NFLX), Tesla (TSLA), Litecoin (LTC-USD), Ethereum (ETH-USD) e Bitcoin (BTC-USD). Pertencentes a um período entre 1 de Janeiro de 2017 e 1 de Janeiro de 2021.

A medida de risco a ser minimizada é o Desvio Absoluto Médio (MAD) em sua forma linear. Para a primeira instância, foram escolhidas as seguintes restrições: restrição de cardinalidade, limitando inferiormente em 1 ativo e superiormente em 10 ativos; e, limitante de investimento, com um percentual mínimo de capital de 0% e máximo de 100%.

Para estas configurações, tidas como relaxadas, podemos obter um portfólio com a seguinte distribuição de pesos:

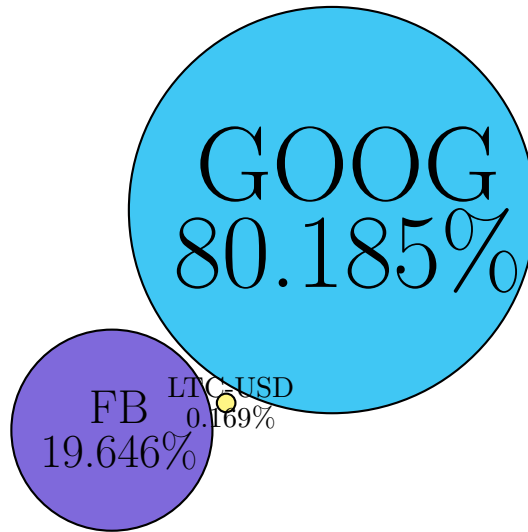


Figura 2 – Distribuições de pesos para portfólio de exemplo. Fonte: O autor.

Determinados ativos não foram representados no gráfico acima devido seu peso ser nulo. Além disso, o mesmo portfólio conta com um retorno de 50,194% e um risco 21,69%.

Já para a instância, foram escolhidas as mesmas restrições, porém a restrição de cardinalidade limita inferiormente em 4 ativos e superiormente em 8 ativos enquanto o limitante de investimento conta com um percentual mínimo de capital de 12.5% e máximo de 25%.

Logo, para estas configurações, um possível portfólio resultante pode possuir a seguinte distribuição de pesos:

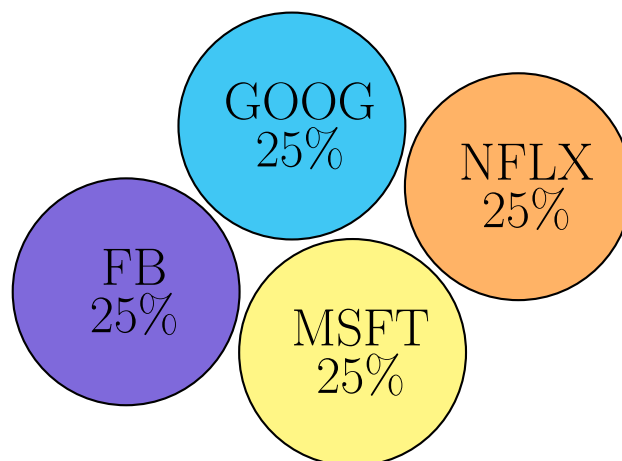


Figura 3 – Distribuições de pesos para portfólio de exemplo. Fonte: O autor.

Ressalta-se que determinados ativos não foram representados no gráfico acima devido seu peso ser nulo. Além disso, o mesmo portfólio conta com um retorno de 84,342% e um risco 39,855%.

Diante dos dois exemplos, pode-se perceber que o o investidor ficaria diante de um *trade-off*. Sendo assim, o primeiro portfólio, apesar de possuir um menor risco, estaria sujeito a um maior impacto diante de uma crise no Google, por exemplo. Com isso, a

pensar no longo prazo, seria uma escolha mais racional o investidor preferir um portfólio mais diversificado e assim percebemos o quão restrições de características reais trazem tais modelos de otimização para uma realidade mais próxima do investidor.

4 Solução Meta-Heurística

O presente trabalho visa como método de solução o uso de meta-heurística. Glover e Kochenberger (2006) define meta-heurística como:

Meta-heurísticas, em sua definição original, são métodos de solução que orquestram uma iteração entre procedimentos de melhoria local e estratégias de alto nível para criar processos capazes de escapar de ótimos locais e promover uma busca robusta pelo espaço de solução.

Conforme o próprio capítulo de revisão bibliográfica aponta, vários autores da literatura utilizaram de meta-heurísticas para resolver problemas de otimização de portfólio com grande número de ativos. Uma das classes de algoritmos que se destaca é a classe do Algoritmo Genético, a qual escolhemos para desenvolver neste trabalho.

4.1 Algoritmo Genético

Um Algoritmo Genético, ou simplesmente AG, fundamentado pelo americano Holland et al. (1992), é uma classe de algoritmos evolutivos a qual utiliza de técnicas inspiradas na evolução das espécies, primeiro observado por Charles Darwin e publicado em seu livro “A Origem das Espécies”, em 1859.

Três pontos principais devem ser definidos para execução de um AG:

- Representatividade de Indivíduo;
- Estratégia de Seleção;
- Operadores de Busca (*Crossover* e *Mutação*).

Uma população é o conjunto de indivíduos que por sua vez são portadores de um código genético. Este código pode ser representado de diversas maneiras, de acordo com o problema abordado. Para problemas de otimização, em geral, a função objetivo, as restrições e as variáveis de decisão são definidas neste código genético.

A estratégia de seleção determina quais indivíduos são escolhidos para reprodução e há diversos meios, os mais conhecidos, abaixo explicados, são: Roleta e Torneio. Para problemas de otimização, a seleção de indivíduos ocorre em termos da função objetivo dos indivíduos, entretanto outros critérios podem ser utilizados, por exemplo, a factibilidade.

A operação de busca *Crossover* imita o processo biológico na reprodução sexuada. Dois indivíduos recebem em seu código genético parte do código do primeiro pai e parte do código do segundo pai. Esta recombinação pode proporcionar que indivíduos troquem entre si as informações que os levam a serem mais aptos a sobreviver, isto é, atingir um melhor valor de função objetivo.

A operação de Mutação propõe a cada indivíduo uma probabilidade de modificação da solução. Esta operação pode ser feita para diversificar os indivíduos ou intensificar a busca por uma melhor solução vizinha.

Uma observação importante é que algumas populações, com o passar das gerações, podem vir a convergir para soluções iguais. Isto ocorre quando os indivíduos com bom valor de função objetivo geram filhos com poucas diversidades.

Outro fator a ser observado no Algoritmo Genético é o critério de parada. Podendo ser estático ou dinâmico. Este deve ser bem definido, sendo possível selecionar um dos seguintes:

- Número máximo de gerações;
- Convergência na função objetivo, ou seja, quando não ocorre melhoria significativa da função;
- Número de gerações sem melhoria na função objetivo;
- Tempo.

4.2 Algoritmo Proposto

O algoritmo abordado no presente trabalho foi construído sobre a linguagem de programação Python, escolhida por seu caráter científico que possibilita o uso, caso necessário, de bibliotecas/dependências, nativas ou não, para manipulações de dados. Sendo assim, necessitando um conhecimento prévio de linha de comando e programação por parte do investidor o qual o deseja utilizar.

O Algoritmo Genético utiliza-se dos módulos: Pandas, Numpy, dentre outros módulos. A primeira é uma ferramenta *open-source* de análise de dados bastante flexível. Já a segunda é uma ferramenta que possibilita uma manipulação mais eficiente de estrutura de dados como *arrays* e matrizes.

Quanto ao paradigma de programação, se é utilizado a Orientação a Objeto. Com isso, o algoritmo pode ser separado em 3 classes, sendo uma destinada para arquivos de entrada de dados, representação de possíveis soluções e lógica de otimização.

A classe de arquivos de entradas de dados consiste na leitura de um arquivo no formato CSV (*comma-separated values*). Cada linha presente no arquivo consiste em uma configuração para o algoritmo. Para cada configuração se faz necessário ajustar 8 parâmetros, sendo eles:

- Arquivo de retorno de ativos de uma dada instância X;
- Arquivo de desvio de retornos dos ativos de uma dada instância X;

- Retorno desejado do portfólio;
- Quantidade mínima de ativos por portfólio;
- Quantidade máxima de ativos por portfólio;
- Investimento mínimo dos ativos do portfólio;
- Investimento máximo dos ativos do portfólio;
- Quantidade de indivíduos pertencentes a população;
- Quantidade de gerações a serem executadas.

A classe de representação de possíveis soluções armazena informações como os ativos pertencentes àquela solução e seus respectivos pesos, qual o retorno e risco daquele portfólio, se o mesmo é uma solução factível. A classe conta também com métodos auxiliares das operações de *crossover* e mutação, encontradas na classe de lógica de otimização (meta-heurística).

Já classe de lógica de otimização consiste no algoritmo genético em si. A operação de seleção de pais, *crossover* e mutação, além de cada método auxiliar utilizado e o método principal presente serão explicados abaixo de forma sucinta.

4.2.1 Operação de Seleção de Pais

Como operação de seleção, foi proposto que todos os indivíduos da população serão escolhidos para a operação de *crossover* por pares aleatórios. Tal método é executado $\frac{p}{2}$ vezes, sendo p o tamanho da população.

4.2.2 Operação de *Crossover*

Como operação de *crossover*, ou cruzamento, foi proposto o seguinte método, cujos parâmetros são os dois pais escolhidos pela seleção. São obtidas duas listas de ativos de pesos não nulos do primeiro e segunda pai. Em seguida, as listas são embaralhadas e se é calculada conforme a equação abaixo a quantidade de trocas entre ativos que haverá:

$$nTroca = 1 + \lfloor \text{Aleatório}(0, \log_2(1 + \min(nAtivosPai1, nAtivosPai2))) \rfloor \quad (8)$$

Após isso inicia-se um laço de repetição, em que $nTroca$ ativos são selecionados aleatoriamente da lista de ativos de pesos não nulos do primeiro e segundo pais. Em sequência, é feita a operação de *crossover*, na qual irá modificar o peso dos ativos selecionados, de modo a continuar obedecendo as restrições de integralidade do capital investido, domínio das variáveis dos pesos, quantidade mínima e máxima de ativos e limitantes de

investimento. Essa modificação é feita por uma função denominada *move_all*(*dados*, *A*, *B*), que tenta mover todo o investimento do ativo A para o ativo B. Esta operação possui complexidade $O(|N|T)$.

Algoritmo 1: Crossover

Parâmetros: *pai_1*, *pai_2*, *dados*

Resultados : Lista contendo os dois filhos resultantes da operação

filho_1, *filho_2* = *Copia*(*pai_1*), *Copia*(*pai_2*)

ativos_filho_1 = *ObterListaAtivos*(*filho_1*)

ativos_filho_2 = *ObterListaAtivos*(*filho_2*)

Embaralhar(*ativos_filho_1*)

Embaralhar(*ativos_filho_2*)

nTroca = $1 + \lfloor \text{Aleatorio}(0, \log_2(1 + \min(nAtivosPai1, nAtivosPai2))) \rfloor$

mudancas_1 = *Lista*()

mudancas_2 = *Lista*()

para *i* ∈ *Sequencia*(*nTroca*) **faça**

se *ativos_filho_1*[*i*] == *ativos_filho_2*[*i*] **então**

 | **continue**

mudancas_1.add(*filho_1.move_all*(*dados*, *ativos_filho_1*[*i*], *ativos_filho_2*[*i*]))

mudancas_2.add(*filho_2.move_all*(*dados*, *ativos_filho_2*[*i*], *ativos_filho_1*[*i*]))

filho_1.recalcula_factibilidade(*dados*, *mudancas_1*)

filho_2.recalcula_factibilidade(*dados*, *mudancas_2*)

retornar *Lista*(*filho_1*, *filho_2*)

4.2.3 Operação de Mutação

Como operação de mutação, foram projetados dois métodos cujo parâmetro é o indivíduo e ocorre para uma quantidade limitada de indivíduos sorteados ao acaso e após o término de todas as operações de *crossover* a cada geração. O indivíduo sorteado sofrerá uma das mutações, sorteada uniformemente.

A mutação 1 gera uma cópia da solução passada como parâmetro, em seguida sorteia um ativo A que possua peso e outro ativo B que não possua peso. Após isso, utiliza-se do método *move_all*(*dados*, *A*, *B*) para mover todo o investimento de A para B e então calcula a factibilidade desta nova solução. Caso a nova solução for melhor que a original tal é retornada, se não o processo de mutação é ignorado. Esta operação possui complexidade $O(|N|T)$ em seu pior caso. O pseudo-código desta operação pode ser visualizado abaixo:

Algoritmo 2: Operação de Mutação 1**Parâmetros:** *solucao, dados**solucao_antiga* = *Copia(solucao)**ativos_solucao* = *ObterListaAtivos(solucao)**ativo_novo* = $\lfloor \text{Aleatorio}(0, \text{dados.numero_ativos} - 1) \rfloor$ **enquanto** *ativo_novo* \in *ativos_solucao* **faça**| *ativo_novo* = $\lfloor \text{Aleatorio}(0, \text{dados.numero_ativos} - 1) \rfloor$ *ativo_velho* = *EscolhderAleatorio(ativos_solucao)**mudancas* = *solucao.move_all(dados, ativo_novo, ativo_velho)**solucao.recalcula_factibilidade(dados, mudancas)***se** *not solucao.is_better(solucao_antiga)* **então**| *mudancas* = *solucao.move_all(dados, ativo_novo, ativo_velho)*| *solucao.recalcula_factibilidade(dados, mudancas)*

Já a mutação 2 possui a mesma estrutura da mutação 1, porém é realizada uma modificação no método *move_all(dados, A, B)* em que irá tentar mover parte do investimento do ativo A para o Ativo B, mantendo o investimento mínimo no ativo A. Esta nova função é denominada *move_all_less_min_inv(dados, A, B)*. Esta operação de mutação possui complexidade computacional $O(|N|T)$. O pseudo-código desta operação de mutação é apresentado abaixo:

Algoritmo 3: Operação de Mutação 2**Parâmetros:** *solucao, dados**solucao_antiga* = *Copia(solucao)**ativos_solucao* = *ObterListaAtivos(solucao)**ativo_novo* = $\lfloor \text{Aleatorio}(0, \text{dados.numero_ativos} - 1) \rfloor$ **enquanto** *ativo_novo* \in *ativos_solucao* **faça**| *ativo_novo* = $\lfloor \text{Aleatorio}(0, \text{dados.numero_ativos} - 1) \rfloor$ *ativo_velho* = *EscolhderAleatorio(ativos_solucao)**solucao.move_all_less_min_inv(dados, ativo_novo, ativo_velho)**solucao.factibilidade(dados)***se** *not solucao.is_better(solucao_antiga)* **então**| *mudancas* = *solucao.move_all(dados, ativo_novo, ativo_velho)*| *solucao.factibilidade(dados)*

4.2.4 Método Principal

Este método é responsável pela execução da meta-heurística Algoritmo Genético. E utiliza as operações de seleção, *crossover* e mutação propostas anteriormente.

Para se gerar uma população inicial, um laço de repetição instância uma nova solução a cada iteração. Um método de sorteio de pesos lista alguns ativos e define pesos aleatoriamente, buscando obedecer as restrições de integralidade do capital investido, domínio das variáveis dos pesos, quantidade mínima e máxima de ativos e limitantes de investimento. Abaixo pode-se visualizar o pseudo-código da instanciação de cada solução.

Algoritmo 4: Heurística Construtiva.

```

Parâmetros: sol, dados
ativos_sorteados = Sample(ativos, max_k)
para ativo ∈ ativos_sorteados[0:min_k] faça sol.mapa_pesos[ativo] = min_inv;
sol.num_ativos = min_k
peso_restante =  $1.0 - \text{min\_inv} * \text{min\_k}$ 
enquanto peso_restante > TOL faça
    para ativo ∈ ativos_sorteados faça
        se ativo ∈ sol.mapa_pesos então
            se sol.mapa_pesos[ativo] + peso_restante < max_inv então
                | sol.mapa_pesos[ativo] += peso_restante
            se não, se Aleatorio(0, 1) < 0.5 então
                | peso_restante -= (max_inv - sol.mapa_pesos[ativo])
                | sol.mapa_pesos[ativo] = max_inv
            se não
                | peso_sorteado = Uniform(0, max_inv - sol.mapa_pesos[ativo])
                | peso_restante -= peso_sorteado
                | sol.mapa_pesos[ativo] += peso_sorteado
        se não, se peso_restante < min_inv então continue ;
        se não, se peso_restante ≤ max_inv então
            | sol.mapa_pesos[ativo] = peso_restante
            | peso_restante = 0
            | sol.num_ativos += 1
        se não, se Aleatorio(0, 1) < 0.5 então
            | peso_restante -= max_inv
            | sol.mapa_pesos[ativo] = max_inv
            | sol.num_ativos += 1
        se não
            | peso_sorteado = Uniform(min_inv, max_inv)
            | peso_restante -= peso_sorteado
            | sol.mapa_pesos[ativo] = peso_sorteado
            | sol.num_ativos += 1

```

Após isso inicia-se um laço de repetição o qual irá iterar até a quantidade de gerações desejada. A cada iteração se é feito o processo de seleção de pais, em seguida aplicada a operação de *crossover* entre os mesmos. Os dois filhos resultantes de cada operação anterior são enfileirados na população atual. A população é ordenada com base na factibilidade do indivíduo, menor risco e maior retorno. Então, mantém-se na lista somente os p primeiros indivíduos (elitismo). Após isso, sorteia-se parte da população que irá sofrer mutação. Em seguida, a população é novamente ordenada e uma nova iteração é feita.

4.2.5 Exemplo

Os dois exemplos citados na Seção 3.2 serão utilizados nesta subseção através do Algoritmo Genético.

Para ambos os exemplos, o Algoritmo Genético foi parametrizado da seguinte maneira:

- População: 100 indivíduos;
- Iterações: 1000;
- Parte da população que sofrerá mutação: 20%.

O primeiro exemplo de portfólio tinha os seguintes dados: quantidade inferior e superior de ativos, respectivamente, os valores de 1 ativo e 10 ativos, enquanto que para o limitante de investimento inferior e superior foram os valores de, respectivamente, 0% e 100% do capital total a se investir.

Embora a heurística construtiva não garanta factibilidade para a restrição de retorno mínimo, todos os portfólios da população inicial gerados foram factíveis. A melhor solução da população inicial teve 0,001% de capital para o nono ativo (Ethereum-USD), em seguida foi atribuído ao terceiro ativo (Facebook) 99,999% do capital, o portfólio então possui um retorno de 46,62% e risco de 22,15%. Nesta parametrização houve-se uma convergência nesta melhor solução da população inicial.



Figura 4 – Distribuições de pesos para portfólio de exemplo. Fonte: O autor.

O segundo exemplo de portfólio foi parametrizado da seguinte maneira: para a quantidade inferior e superior de ativos, respectivamente, os valores de 4 ativo e 8 ativos, enquanto que para o limitante de investimento inferior e superior foram os valores de, respectivamente, 12,5% e 25% do capital total a se investir.

Quanto a geração da população inicial, todos os portfólios iniciais gerados foram factíveis. Para a primeira solução da população inicial, foi gerado um portfólio com 6 ativos,

sendo distribuído da seguinte forma o capital: o primeiro ativo escolhido foi o Facebook, com 25% do capital; o segundo ativo foi a Microsoft, já com 12,5% de capital, o terceiro consiste em 12,5% para a Netflix, já o quarto, quinto e sexto, foram, respectivamente, 19,637% para o Google, 15,076% para Apple e 15,287% para Amazon. Esta solução foi o portfólio com maior performance gerado pela população inicial, possuindo um retorno de 85,435% e risco de 42,418%. Com isso, a melhor solução ao término do processamento consiste em um portfólio com quatro ativos (Facebook, Google, Microsoft e Netflix), em que possuem todos o 25% de capital alocado, ou seja, o limitante superior de investimento. Este portfólio possui um retorno de 84,342% e risco de 39,855%, atingindo a solução ótima obtida pelo *solver* IBM ILOG CPLEX.

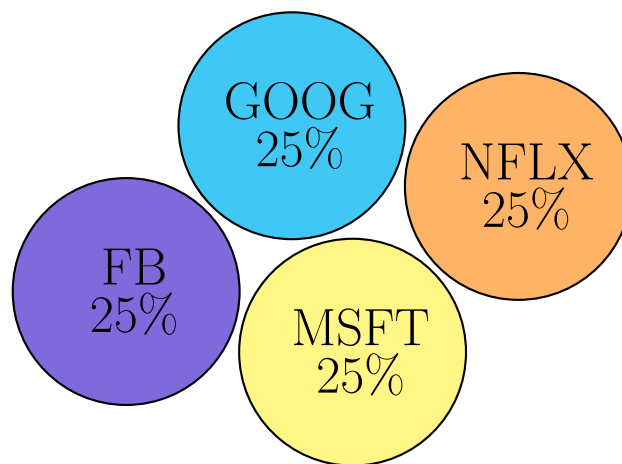


Figura 5 – Distribuições de pesos para portfólio de exemplo. Fonte: O autor.

5 RESULTADOS COMPUTACIONAIS

Esta seção destina-se a comparação de resultados de testes realizados entre o algoritmo genético proposto na seção anterior com o software de otimização comercial IBM ILOG CPLEX. São comparados a quantidade de soluções factíveis obtidas, o tempo computacional e o risco apresentado por cada conjunto de instância.

O computador utilizado para a realização dos testes consiste em um Acer Aspire 5 com 8GB de memória RAM e Intel Core i7-7500U 2.7GHz. Já o sistema operacional consiste no Windows 10 Home Single Language.

5.1 Conjunto de Instâncias

Para a análise de performance foram sorteadas 1386 ações dentre todas listadas na NASDAQ, o mercado de ações americano de tecnologia. A partir das 1386 ações foram selecionadas 13 classes de instâncias de acordo com o número de ativos definidos: 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200 e 1300 ativos. Tendo como objetivo verificar a eficiência e robustez do Algoritmo Genético (GA), foram elaborados um total de 2496 instâncias de teste, variando alguns dos parâmetros listados abaixo:

- Ativos da instância X;
- Retorno desejado do portfólio: 0% e 20%;
- Quantidade mínima de ativos por portfólio: 1 ativo, 10% e 20% do total de ativos;
- Quantidade máxima de ativos por portfólio: 100%, 20% e 10% do total de ativos;
- Investimento mínimo dos ativos do portfólio: 0.01%, 1%, 5% e 10% do capital disponível;
- Investimento máximo dos ativos do portfólio: 100%, 20% e 10% do capital disponível;

Para o Algoritmo Genético, são utilizados os seguintes parâmetros:

- Porcentagem da população a sofrer mutação: 20%;
- Quantidade de indivíduos pertencentes a população: 100;
- Quantidade de gerações: 100 e 1000.

Os valores passados aos parâmetros a ambos meios de resolução foram escolhidos levando em consideração diferentes cenários de investimentos. Cenários, por exemplo, em que o investidor há de preferir poucos ativos sob sua gestão e maior concentração de capital

nestes ativos. Quanto aos parâmetros passados ao Algoritmo Genético, foram realizados um conjunto de testes a fim de validar qual configuração retornaria um melhor resultado para a função objetivo.

Podemos perceber, uma vez que gerado todas as combinações dos parâmetros acima, que determinadas soluções serão infactíveis por natureza. Por exemplo, não há como satisfazermos quando a quantidade mínima de ativos é igual a 80 ativos e 10% de investimento mínimo. Algumas classes de instâncias naturalmente infactíveis foram dispensadas.

O IBM ILOG CPLEX é configurado para um tempo limite de 30 segundos de processamento e o resto da configuração padrão. Tal medida é tomada pois o software busca sempre a solução ótima, com isso determinadas instâncias terão um tempo computacional muito elevado, fazendo com que determinados testes ultrapassem até um dia de execução. Porém, pode-se notar que a média de valores para determinadas instâncias superam tal valor, pois o software contabiliza o tempo que leva para carregar todos os dados da instância. Além disso, para tornar mais justa a comparação com o algoritmo genético, foi determinado este tempo limite, pois o IBM ILOG CPLEX é também paralelizado, chegando a utilizar até 4 *threads* em seu processo, enquanto que o nosso algoritmo proposto utiliza somente de 1 *thread*.

5.2 Resultados Computacionais

Para resolução do modelo matemático por parte do IBM ILOG CPLEX, o *solver* utiliza dos algoritmos: *Branch And Bound* e *Simplex* de modo paralelizado.

Iremos denotar por GA 100 a execução de testes cuja quantidade de gerações é igual a 100 e por GA 1000 a execução de testes cuja quantidade de gerações é igual a 1000.

A primeira vista, analisando os resultados da execução dos testes, o algoritmo genético obteve um total de 1632 soluções factíveis, uma média de 3,73 segundos/teste para a execução GA 100, e 1632 soluções factíveis e uma média de 16,9 segundos/teste para execução GA 1000. Enquanto isso, o IBM ILOG CPLEX obteve um total de 1555 soluções factíveis, e a seguinte média para tempo de execução: 46,25 segundos.

Nos próximos parágrafos iremos analisar mais a fundo os dois softwares em relação às 13 classes de instâncias de acordo com número de ativos utilizadas nos testes.

O gráfico abaixo mostra a média de risco para cada classe de instância (em centenas). Podemos notar que na maioria das classes o IBM ILOG CPLEX obteve um risco abaixo de ambas as versões do Algoritmo Genético. Houve uma boa melhoria nas soluções encontradas pelo GA 1000, em comparação às soluções do GA 100.

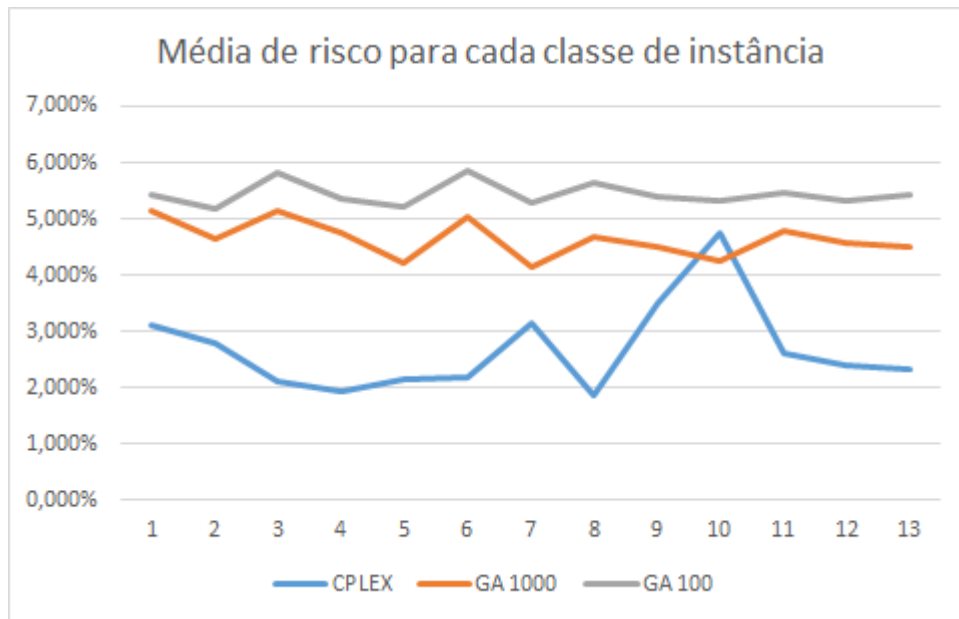


Figura 6 – Média de risco por instâncias testadas. Fonte: O Autor.

Complementando o gráfico acima, apresentamos outro gráfico que apresenta, ao longo das classes de instâncias, o número de soluções factíveis encontradas. O Algoritmo Genético encontrou soluções para as mesmas instâncias encontradas pelo IBM ILOG CPLEX e para outras instâncias, principalmente na classe de 1000 ativos.

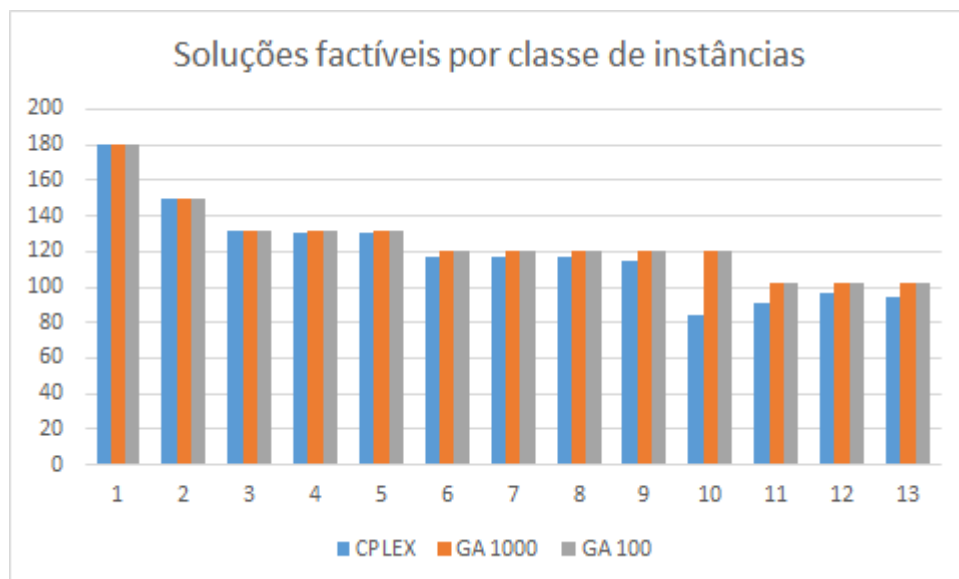


Figura 7 – Quantidade de soluções factíveis geradas por instância testada. Fonte: O Autor.

Vale ressaltar que o algoritmo utilizado pelo IBM ILOG CPLEX não é um método exato, pois não possui tempo suficiente para encerrar a busca no espaço de soluções factíveis.

Conforme proposto por Dolan e Moré (2002), iremos analisar abaixo os perfis de performance dos três procedimentos (conjunto S). A performance de um procedimento

s em uma instância p , dada por $t_{p,s}$, é definida pelo risco da melhor solução factível encontrada. A análise realizada define uma taxa de performance $r_{p,s}$ dada pela Equação(9), em que é feita a diferença entre a performance do procedimento s pela melhor performance encontrada. Em seguida, irá se calcular a porcentagem de soluções que possuem taxa de performance inferior a um dado valor τ (Equação (10)).

$$r_{p,s} = t_{p,s} - \min\{t_{p,s} : s \in S\} \quad (9)$$

$$\rho_s(\tau) = \frac{|p \in P : r_{p,s} \leq \tau|}{|P|} \quad (10)$$

No gráfico abaixo podemos visualizar a porcentagem de soluções factíveis encontradas até uma diferença absoluta de $t\%$ de risco entre os métodos de execução e a melhor solução encontrada.

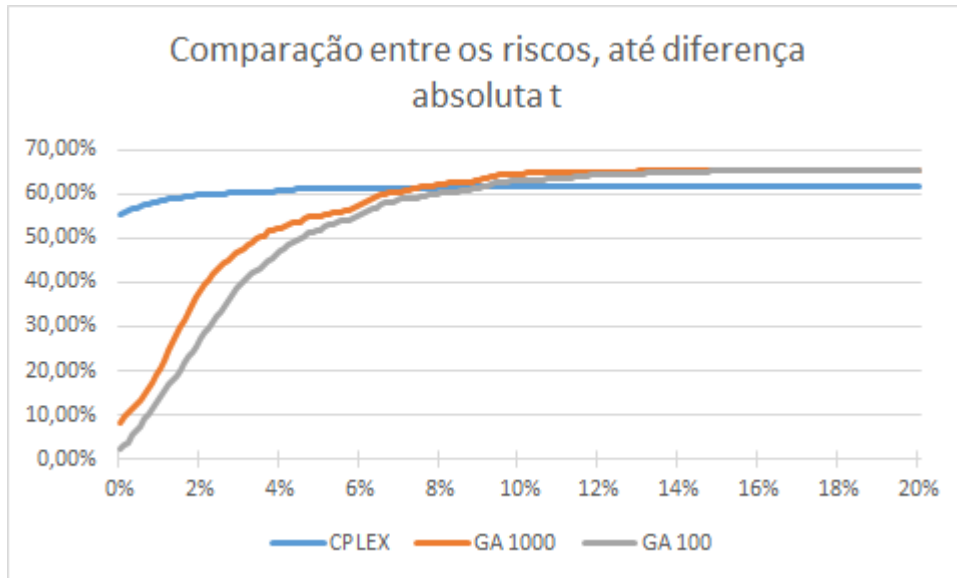


Figura 8 – Comparação entre riscos até diferença absoluta de 20%. Fonte: O Autor.

O IBM ILOG CPLEX não encontrou soluções factíveis em 37,7% dos casos. Tanto o GA 1000 como GA 100 não encontraram soluções em 34,62% dos casos.

O IBM ILOG CPLEX obteve a melhor solução em 55,2% dos casos. Enquanto isso, o GA 1000 e GA 100 obtiveram a melhor solução em 8,33% e 2,32% dos casos, respectivamente. Entretanto, os algoritmos genéticos propostos rapidamente encontram a linha de performance do IBM ILOG CPLEX, e por volta de 8% de diferença absoluta de risco, as três execuções começaram a apresentar performances semelhantes.

O gráfico abaixo apresenta as médias dos tempos computacionais por classe de instâncias (em milissegundos). O GA 100 possuiu pouca variação de tempo ficando, em média, abaixo de 10 segundos. Já o GA 1000 e o IBM ILOG CPLEX tiveram comportamentos semelhantes, com o GA 1000 mais rápido em geral, ficando com a média abaixo de 30 segundos para todas as classes de instâncias.

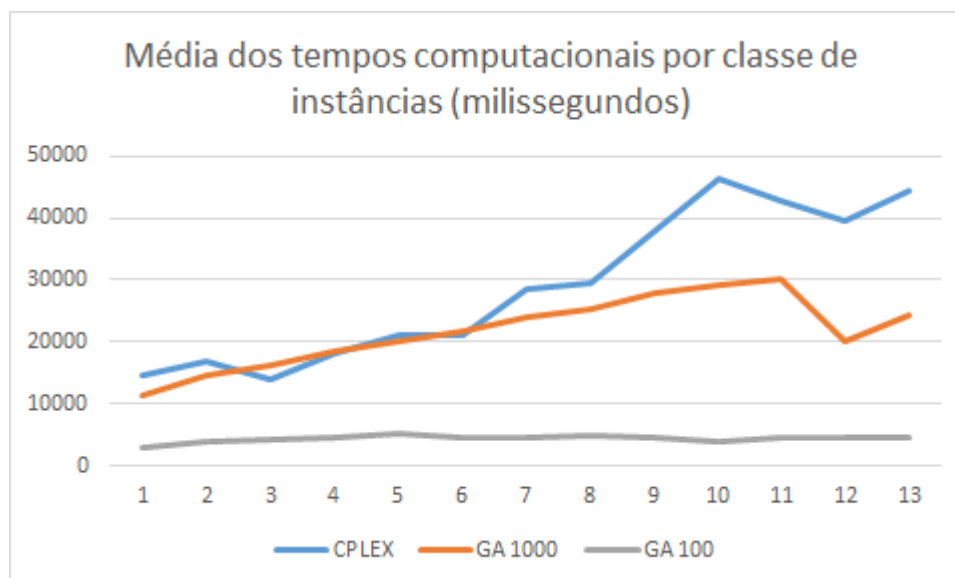


Figura 9 – Média de tempo computacionais por classe de instância. Fonte: O Autor.

Já no gráfico abaixo podemos visualizar a convergência média das instâncias com soluções factíveis.

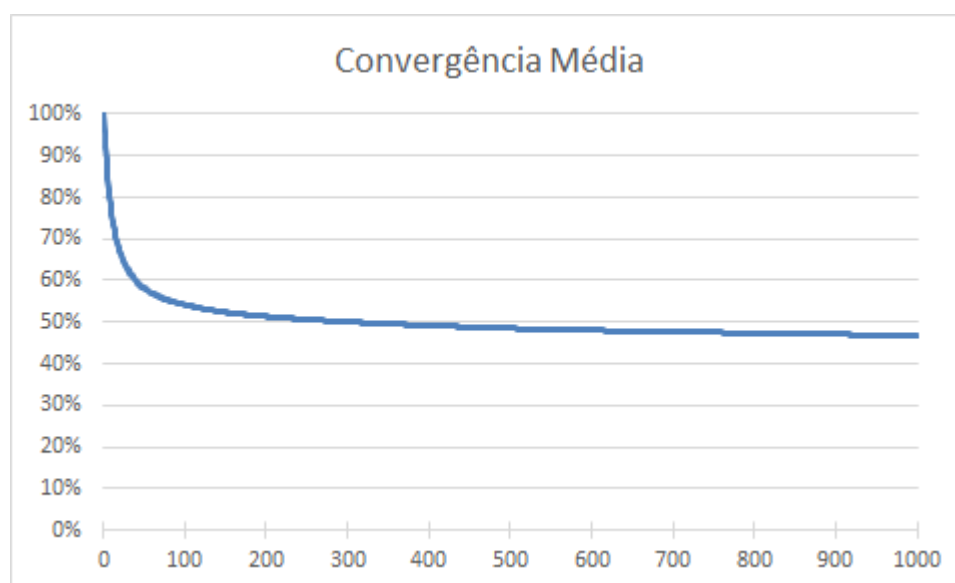


Figura 10 – Convergência média das instâncias de soluções factíveis. Fonte: O Autor.

Pode-se observar que para as gerações iniciais há uma grande redução do valor do risco. Em seguida, a redução do risco torna-se menos significativa conforme a quantidade aumenta. Ao final, o risco na milésima geração é cerca de 46,83% da melhor solução factível encontrada na população inicial. Enquanto isso, na centésima geração o risco é cerca de 54,16% da melhor solução factível encontrada na população inicial. Com isso, podemos concluir que o aumento de gerações nos permite reduzir ainda mais o risco das soluções factíveis.

6 CONCLUSÃO E PERSPECTIVA

O presente trabalho é responsável pela resolução do problema de geração de portfólios de investimento com características reais. A resolução deste problema permite com que investidores e praticantes busquem carteiras de investimento com risco reduzido. Como metodologia de resolução foi proposto o uso da meta-heurística Algoritmo Genético e comparado à resolução do modelo matemático de programação inteira mista no software de otimização IBM ILOG CPLEX.

Os dois métodos de solução foram submetidos a um conjunto de instâncias com diferentes parâmetros de retorno mínimo, quantidades de ativos mínimos e máximos da carteira de investimento e limitantes inferiores e superiores de investimento.

O algoritmo genético obteve mais soluções factíveis que o IBM ILOG CPLEX quanto maior o número de ativos disponíveis. O IBM ILOG CPLEX obteve em geral soluções com menor risco, com uma média 2,68% ante 4,64% do GA 1000 e 5,44% do GA 100.

Nosso algoritmo se apresenta como uma solução melhor para investidores iniciantes e intermediários, pois trata-se de uma solução mais viável financeiramente e principalmente flexível, pois permite com que o usuário realize modificações em seu código fonte de acordo com a estratégia que queira adotar. Por exemplo, caso deseje adicionar restrições para priorizar a presença de ativos do setor de tecnologia, utilizando uma base de dados como a IBOVESPA.

O Algoritmo Genético proposto foi projetado para ter um tempo computacional reduzido. O tempo de execução médio do GA 100 corresponde a 8% do tempo computacional do IBM ILOG CPLEX, sendo que este foi limitado. O tempo reduzido permite que o investidor faça diferentes testes a partir de mudanças nos parâmetros, podendo assim obter os parâmetros que melhor satisfazem suas preferências.

Trabalhos futuros podem ser propostos a fim de diminuir o espaço existente entre os riscos obtidos pelo IBM ILOG CPLEX e o algoritmo aqui proposto. Tais como o desenvolvimento de novos operadores de *crossover* e mutação, outras estratégias meta-heurísticas, e o uso de *matheuristics* (uma meta-heurística híbrida com procedimentos de solução exata). Outra proposta, pode ser a inserção de outras características reais tais como classificar os ativos por setor e limitar o número mínimo/máximo de setores e a quantidade de ativos investidos em cada um deles, como também limitar inferiormente e superiormente o capital investido por setor, e até mesmo a inserção de restrições quanto a taxas e impostos no contexto brasileiro. Outra proposta consiste na análise de múltiplos cenários de geração de portfólios, utilizando variáveis estocásticas.

Referências

- CHANG, T.-J.; YANG, S.-C.; CHANG, K.-J. Portfolio optimization problems in different risk measures using genetic algorithm. **Expert Systems with applications**, Elsevier, v. 36, n. 7, p. 10529–10537, 2009. Citado 2 vezes nas páginas 4 e 6.
- CHIANGLIN, C.-Y. Applications of genetic algorithm to portfolio optimization with practical transaction constraints. In: ATLANTIS PRESS. **9th Joint International Conference on Information Sciences (JCIS-06)**. [S.l.], 2006. Citado 2 vezes nas páginas 4 e 6.
- CORMEN, T. H. et al. **Introduction to algorithms**. [S.l.]: MIT press, 2009. Citado 2 vezes nas páginas 4 e 11.
- DANTZIG, G. B. Origins of the simplex method. In: **A history of scientific computing**. [S.l.: s.n.], 1990. p. 141–151. Citado na página 10.
- DOLAN, E. D.; MORÉ, J. J. Benchmarking optimization software with performance profiles. **Mathematical programming**, Springer, v. 91, n. 2, p. 201–213, 2002. Citado na página 26.
- FEINSTEIN, C. D.; THAPA, M. N. A reformulation of a mean-absolute deviation portfolio optimization model. **Management Science**, v. 39, n. 12, 1993. Citado 2 vezes nas páginas 4 e 6.
- GLOVER, F. W.; KOCHENBERGER, G. A. **Handbook of metaheuristics**. [S.l.]: Springer Science & Business Media, 2006. v. 57. Citado na página 15.
- HOLLAND, J. H. et al. **Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence**. [S.l.]: MIT press, 1992. Citado na página 15.
- JAAMAN, S.; LAM, W. Mean-variance and mean-gini analyses to portfolio optimization in malaysian stock market. **Economic and Financial Review**, v. 2, n. 2, p. 60–64, 2012. Citado 2 vezes nas páginas 4 e 6.
- KAMALI, S. Portfolio optimization using particle swarm optimization and genetic algorithm. **Journal of mathematics and computer science**, v. 10, n. 2, p. 85–90, 2014. Citado 2 vezes nas páginas 5 e 6.
- KONNO, H.; YAMAZAKI, H. Mean-absolute deviation portfolio optimization model and its applications to tokyo stock market. **Management science**, INFORMS, v. 37, n. 5, p. 519–531, 1991. Citado 4 vezes nas páginas 2, 4, 5 e 8.
- MANSINI, R. et al. **Linear and mixed integer programming for portfolio optimization**. [S.l.]: Springer, 2015. Citado na página 1.
- MARKOWITZ, H. Portfolio selection. **The Journal of Finance**, [American Finance Association, Wiley], v. 7, n. 1, p. 77–91, 1952. ISSN 00221082, 15406261. Disponível em: <<http://www.jstor.org/stable/2975974>>. Citado 5 vezes nas páginas , 2, 3, 5 e 7.

- MORRISON, D. R. et al. Branch-and-bound algorithms: A survey of recent advances in searching, branching, and pruning. **Discrete Optimization**, Elsevier, v. 19, p. 79–102, 2016. Citado na página 12.
- PINHEIRO, J. L. **História da Bolsa de Valores**. 2014. Acesso em: 12/08/2021. Disponível em: <https://www.mercadocomum.com/historia_da_bolsa_de_valores/>. Citado na página 3.
- QUEIROZ, T. A. de et al. Linear models for portfolio selection with real features. In: **Advances in Optimization and Decision Science for Society, Services and Enterprises**. [S.l.]: Springer, 2019. p. 35–45. Citado na página 8.
- SEFIANE, S.; BENBOUZIANE, M. Portfolio Selection Using Genetic Algorithm. **Journal of Applied Finance Banking**, v. 2, n. 4, p. 143–154, 2012. Citado 2 vezes nas páginas 5 e 6.
- SPERANZA, M. G. A heuristic algorithm for a portfolio optimization model applied to the milan stock market. **Computers & Operations Research**, Elsevier, v. 23, n. 5, p. 433–441, 1996. Citado 2 vezes nas páginas 4 e 6.
- WOODSIDE-ORIAKHI, M.; LUCAS, C.; BEASLEY, J. E. Heuristic algorithms for the cardinality constrained efficient frontier. **European Journal of Operational Research**, Elsevier, v. 213, n. 3, p. 538–550, 2011. Citado 3 vezes nas páginas 2, 5 e 6.
- YITZHAKI, S. Stochastic dominance, mean variance, and gini's mean difference. **The American Economic Review**, JSTOR, v. 72, n. 1, p. 178–185, 1982. Citado 3 vezes nas páginas 2, 4 e 5.

ANEXO A – Código fonte Algoritmo Genético

Abaixo é possível acessar o repositório com os arquivos utilizados para executar o algoritmo proposto no presente trabalho.

<<https://github.com/gabrielgarciapavan2/tcc-algoritmo-genetico>>