

Universidade Federal do Paraná
Setor de Ciências Exatas
Departamento de Estatística
Programa de Especialização em *Data Science* e *Big Data*

José Matias de Oliveira Junior

Otimização da Alocação de Investimentos em Mídia Online usando Otimização Linear

**Curitiba
2020**

José Matias de Oliveira Junior

Otimização da Alocação de Investimentos em Mídia Online usando Otimização Linear

Monografia apresentada ao Programa de Especialização em Data Science e Big Data da Universidade Federal do Paraná como requisito parcial para a obtenção do grau de especialista.

Orientador: Prof. Abel Soares Siqueira

Curitiba
2020

Otimização da Alocação de Investimentos em Mídia Online usando Otimização Linear

José Matias de Oliveira Junior¹
Prof. Abel Soares Siqueira²

Resumo

A alocação de investimento em plataformas de mídia online é um problema onde os profissionais de Mídia Online e Marketing Digital precisam avaliar a capacidade de gerar os resultados pretendidos para o projeto em diversas plataformas em períodos diferentes, para então, baseados em determinados critérios, indicar qual o investimento ideal em cada uma destas plataformas. No entanto, o processo de muitos profissionais ainda é baseado em médias simples, dados históricos parciais ou mesmo feito de forma intuitiva. Por isto este projeto prevê o desenvolvimento de um programa capaz de receber os dados do profissional de mídia e, usando programação linear, alocar os investimentos de forma a maximizar os resultados pretendidos pelo usuário.

Palavras-chave: Programação Linear, Alocação de investimentos em mídia, Julia

Abstract

The budget allocation in online media platforms is a problem where online media and Digital Marketing professionals need to assess the capacity to generate the intended results for the project on different platforms in different periods, then, based on certain criteria, indicate which is the ideal investment in each of these platforms. However, this process for many professionals still based on simple averages, partial historical data or even done intuitively. This project foresees to develop a program capable of receiving data from the media professional and, using linear programming, allocate investments in order to maximize the results intended by the user.

Keywords: Linear Programming, online media budget allocation, Julia

1 Introdução

O investimento em mídia realizado por agências de propaganda é uma prática existente há décadas. Com o surgimento da internet e a capacidade de se trabalhar com verbas de menor porte e em formato *self-service*, o próprio anunciante pode planejar, configurar e executar suas campanhas em plataformas como o Facebook ou

mesmo o Google, exibindo seus anúncios nas buscas feitas por milhões de usuários.

Estas plataformas, diferente de mídias tradicionais, como Rádio e TV, permitem a coleta de vários dados de desempenho das campanhas, anúncios e segmentações. Este cenário fez com que os profissionais de comunicação passassem a trabalhar de forma orientada a dados.

Isso ainda gera diversas dificuldades para as agências, e entre estes desafios está a necessidade de planejar o investimento do cliente entre os diferentes canais disponíveis, em posse de dados de desempenho de cada plataforma.

Em geral, são feitas regras simples considerando o Alcance (quantos usuários cada plataforma pode atingir, ou seja, exibir um anúncio) e o custo médio de Cliques, conhecido como CPC (Custo Por Clique), desta plataforma ao longo do tempo, para assim estimar o número de visitantes a um site ou *app*.

Em posse disto, é avaliada em seguida a taxa de conversão destes usuários, ou seja, a relação entre o total de conversões (resultados pretendidos pelo cliente, como uma venda ou um cadastro) e o total de visitantes, de forma ao multiplicarmos estes dados, chegarmos ao total de conversões estimado daquela plataforma.

No entanto, em muitos casos não é considerado o Alcance, ou mesmo variações como sazonalidades, onde em determinados meses, uma plataforma pode impactar mais usuários, ou mesmo ter custos menores.

Com isso em mente, foi proposto um modelo matemático onde estas variáveis, além daquelas já trabalhadas por profissionais de mídia, possam ser avaliadas frente à variações mensais na capacidade de alcance de cada plataforma.

Em suma, este é um problema de otimização onde o objetivo é a maximização do Total de Conversões, sejam quais forem, baseado em restrições de público e orçamento, e em vista das taxas de conversão e custo por clique de cada plataforma em N meses.

Para o projeto foi usada a Linguagem Julia[1] e o pacote JuMP[2], uma linguagem de modelagem e pacotes de suporte para otimização matemática em Julia., com o *solver* GLPK[3].

2 Programação Linear

A programação linear é uma área da pesquisa operacional muito utilizada em processos para apoio à tomada de decisão. O termo “programação”, não tem a ver diretamente com programação de computadores, ou linguagem de programação. Este termo pode ser visto como um sinônimo de planejamento [4].

O termo “programação” da programação linear está relacionado ao planejamento de recursos visando atender determinadas condições. Estas condições, ou restrições, são representadas por equações e funções lineares.

"Assim, podemos definir a programação linear como sendo o planejamento de atividades para obter um resultado ótimo, isto é, um resultado que atenda, da melhor forma possível, a um determinado objetivo"[4]. Por isso caracterizamos a programação linear como uma técnica de otimização.

No problema de otimização em uma indústria, por exemplo, buscamos determinar a alocação ótima dos recursos de produção, de forma a atender as limitações de capacidades produtivas de cada unidade e maximizar o lucro resultante.

Tanto a função de maximizar o lucro quanto as restrições de capacidade de cada planta são representados por funções lineares. Com isso, apesar do tomador de decisão pode escolher entre diversas possibilidades de alocação de seus recursos, apenas uma combinação é a mais lucrativa. Este é o resultado ótimo que maximiza o lucro, uma função, ou seja um problema de programação linear.

O problema de programação linear foi inventado pelos matemáticos L. Kantorovich e T. Koopmans em 1939 [5]. O algoritmo mais popular para resolver problemas de programação linear é o Simplex e suas variações, criado por George Dantzig em 1.947 [6].

2.1 Aplicações da Programação Linear

Muitas áreas utilizam de forma recorrente a programação linear no processo de apoio a decisão. Entre elas estão: planejamento logístico, planejamento da produção, setor financeiro, indústrias em geral, entre outros.

Em basicamente qualquer cenário onde precisa se reduzir os custos, prazos ou maximizar o lucro e resultados de um processo, pode ser formulado um modelo matemático do problema, com suas restrições, e então aplicar a programação linear na busca de um resultado ótimo.

Porém, tais ferramentas são em grande parte desconhecidas do universo dos profissionais de comunicação e marketing.

2.2 Principais implementações

O algoritmo mais popular para resolver problemas de programação linear é o Algoritmo Simplex, desenvolvido na década de 1940[6]. O Simplex possui diver-

sas implementações tecnicamente válidas, como por exemplo LINDO, CPLEX, Gurobi e Solver (no Excel).

O Simplex é usado para valores ideais em situações em que diversos aspectos precisam ser respeitados. Diante de um problema, são estabelecidas inequações que representam restrições para as variáveis. A partir daí, testam-se diversas possibilidades com o objetivo de otimizar o resultado da forma mais rápida possível.

As restrições são apresentadas como inequações, como por exemplo, o fato de uma empresa só conseguir produzir uma determinada quantidade dos produtos de cada tipo. Assim, entre as possibilidades de valores para as variáveis que atendam às restrições, o algoritmo deve encontrar aqueles que dão à função objetivo o maior resultado possível.

Este objetivo é representado pela função objetivo, onde o objetivo de maximizar ou minimizar também é expressado através de uma equação.

Para o projeto foi usado o GLPK (GNU Linear Programming Kit), um pacote de software destinado a resolver a programação linear em grande escala, programação inteira mista (MIP) e outros problemas relacionados. É um conjunto de rotinas escritas em ANSI C, organizadas na forma de uma biblioteca que pode ser usada em diversas linguagens, inclusive o Julia[3]. O pacote é parte do Projeto GNU e é lançado sob a Licença Pública Geral GNU.

Por este motivo o GLPK foi selecionado para ser usado no projeto, uma vez que o problema apresentado é razoavelmente simples, não existe necessidade de algoritmos mais complexos, e ainda, com a vantagem de ser software-livre.

3 Desenvolvimento do projeto

Com o objetivo de disponibilizar o uso do modelo desenvolvido foi criado um site onde o usuário pode informar os dados como investimento, CPC, etc. e gerar um arquivo Excel com a alocação do investimento por plataforma ao longo dos meses.

Para o desenvolvimento deste site foi usado o Genie Framework [7]. O Genie é um framework web MVC full-stack que promove um fluxo de trabalho simplificado e eficiente para o desenvolvimento de aplicativos web modernos em Julia.

O modelo MVC significa Model – View – Controller (MVC) é um padrão de arquitetura de software para implementar sistemas utilizado de forma recorrente em projetos de software para web. Ele divide um software em três partes interconectadas, de modo a separar as funções internas das funções com as informações que são apresentadas para o usuário [8].

Com o Genie foi possível criar páginas (views) onde o usuário pode consultar informações sobre o projeto, bem como gerar o arquivo XLS do seu plano de mídia.

Para a geração do arquivo XLS foi usado o pacote XLSX.jl, desenvolvido por Felipe Noris[9].

Com este pacote, os dados gerados pelo resultado do GLPK foram colocados em um arquivo Excel de forma

Plataforma	Maio	Junho	Julho	Agosto	Total
Google Search	R\$ 3.250,00	R\$ 6.500,00	R\$ 6.500,00	R\$ 3.250,00	R\$ 19.500,00
Google Display	R\$ 1.500,00	R\$ 3.000,00	R\$ 3.000,00	R\$ 1.500,00	R\$ 9.000,00
Facebook	R\$ 2.500,00	R\$ 5.000,00	R\$ 5.000,00	R\$ 2.500,00	R\$ 15.000,00
Instagram	R\$ 750,00	R\$ 1.500,00	R\$ 1.500,00	R\$ 750,00	R\$ 4.500,00
					R\$ 48.000,00

Tabela 1: Tabela de Investimento Alocado pelo Analista do Primeiro Plano Avaliado

simples, permitindo que os usuários possam trabalhar nestes documentos como acharem necessário.

No arquivo XLS gerado existem duas abas, na primeira são informados o resultado da solução do GLPK, investimento previsto, alocado e saldo do investimento, se houver 1.

O status da solução pode ser *INFEASIBLE*, onde não foi encontrada uma solução ótima que atenda todas as restrições, ou *OPTIMAL* onde foi encontrada a solução que maximiza os resultados de conversões.

Uma vez que o sistema nem sempre aplica todo o investimento disponível, já foi colocado o investimento informado pelo usuário (previsto) e o alocado, bem como a diferença entre estes valores (saldo).

por plataforma em N meses informados pelo usuário 4.

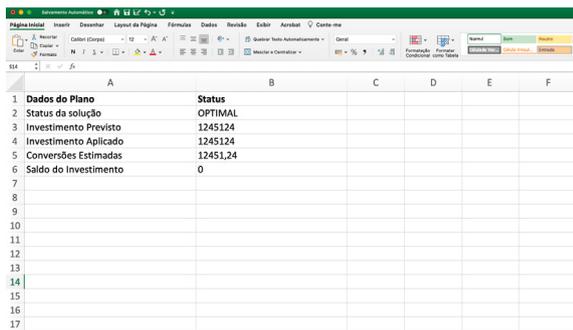


Figura 1: Aba inicial do Excel gerado pelo sistema.

Na segunda aba é apresentada a alocação de investimento gerada pelo GLPK, agrupada por plataforma em N meses informados pelo usuário 2.

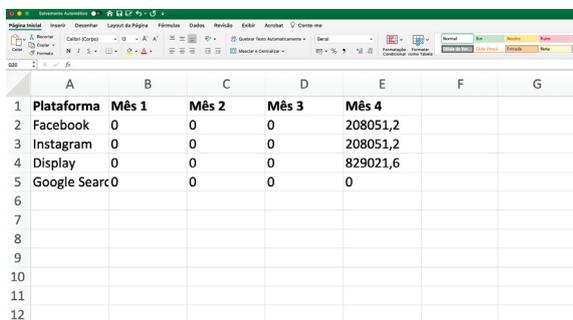


Figura 2: Aba da alocação de investimento.

Caso o sistema não encontre uma solução ótima, será impresso o status *INFEASIBLE* no arquivo gerado, ainda que seja feita uma alocação dos recursos 3.

E da mesma forma, na segunda aba é apresentada a alocação de investimento gerada pelo GLPK, agrupada

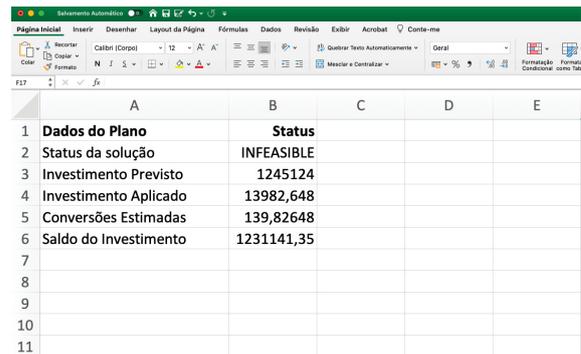


Figura 3: Aba inicial do Excel gerado pelo sistema.

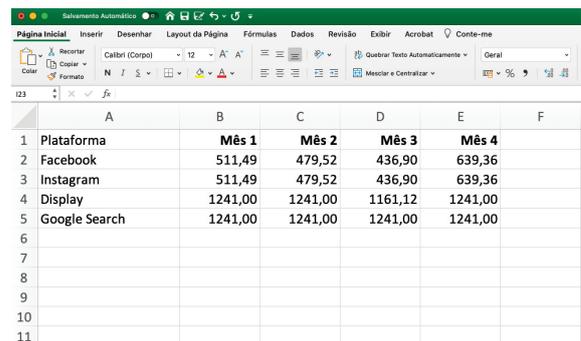


Figura 4: Aba da alocação de investimento.

Para a configuração do servidor da Aplicação foi usado o ubuntu server.

O site está disponível em <https://budget.app.br/> 5 e o código-fonte do projeto está disponível no Github sob a licença GPL v3.0[10].

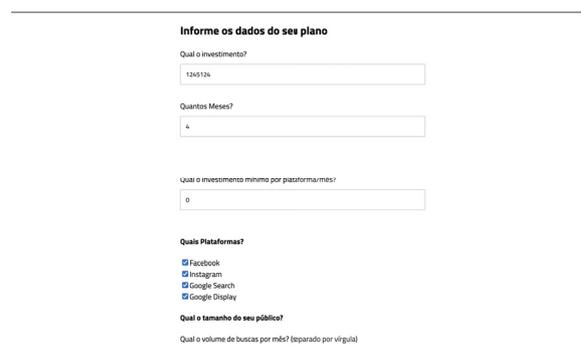


Figura 5: Tela inicial do site budget.app.br. Formulário de Entrada das informações do usuário.

O fluxo do projeto prevê as seguintes interações: O usuário informa os dados de investimento, total de meses e métricas de cada plataforma; o GLPK processa estes dados e é gerado então o XLS com o resultado da alocação 6.



Figura 6: Ilustração do fluxo de interações com o sistema.

3.1 O Modelo Matemático

O principal desafio na construção do modelo matemático do problema é o fato de que não existe um método institucionalizado entre os profissionais de mídia online para a alocação de investimento.

De forma geral existem basicamente dois tipos de campanhas: Campanhas de Resposta Direta ou Conversão, e Campanhas de *Awareness*.

O primeiro tipo, é onde existe um resultado claro e quantificável, como vendas, cadastros, inscrições ou matrículas. Já o segundo é geralmente usado para criar lembrança de marca, onde mensagens publicitárias são repetidas com frequência, usando de *slogans* das empresas, como por exemplo: "Amo muito tudo isso".

Este projeto aborda a alocação de investimentos para o primeiro tipo de campanha: de resposta direta, onde existem conversões quantificáveis.

Neste contexto, são usados dados de desempenho de campanhas anteriores ou *benchmarks* para definir a melhor alocação dos investimentos. Onde os profissionais podem usar equações simples como o Investimento disponível e o CPC e taxa de conversão média para estimar resultados de conversão, ou ainda, em um segundo modelo usar o CPM (custo por mil impressões) multiplicado pelo CTR (*Click Through Rate*), definido pelo total de cliques a cada 100 impressões de um anúncio, para em seguida multiplicar pela taxa de conversão, assim estimando os resultados de uma determinada campanha.

No entanto, estes modelos não costumam trabalhar com restrições em seus planejamento, como o número de usuários em uma plataforma ou o volume de buscas feitas mensalmente no Google.

Estas restrições foram representadas como sendo o público potencial geral em todos os meses, ou chamada também de impressões na rede de display. Já para Google Search, é considerado o volume de buscas mensal, referentes as buscas feitas mensalmente na plataforma.

Por isso foi selecionado o primeiro modelo apresentado, onde já temos o CPC (uma vez que ele resume a relação de CPM e CTR) e colocadas as restrições do tamanho do público-alvo e o volume de buscas mensais, dado informado pelas ferramentas, foi considerado como uma restrição necessária. Do contrário o algoritmo iria alocar

tudo o investimento na plataforma com o menor Custo por Clique e maior Taxa de Conversão.

Além disso foi colocada uma restrição onde o usuário pode informar o investimento mínimo em determinada plataforma. Já que em muitos casos, mesmo que esta plataforma não tenha as maior capacidade de gerar resultados e conversões, pode ser demandando pelo cliente do projeto a presença de mídia.

Além destas restrições, existe o investimento disponível, onde este não pode ser ultrapassado. Porém eventualmente o resultado do algoritmo pode alocar menos investimento do que o disponível, o que inclusive é uma rotina pouco vista entre os profissionais.

Nos processos atuais os planos costumam alocar todo investimento de forma muito similar entre as plataformas e meses, usando o total da verba na maioria dos casos, pela falta de informações sobre outras restrições nestes canais.

3.2 Parâmetros do Modelo Matemático

Os parâmetros que usamos no modelo são

- ▶ A : Investimento disponível para todo o período e todas as plataformas (em R\$).
- ▶ ℓ : Mínimo a ser alocado para todas as plataforma em todos períodos (em R\$).
- ▶ c_{pt} : Custo por clique da plataforma p no período t (em R\$).
- ▶ q_{pt} : Taxa de conversão da plataforma p no período t (em %).
- ▶ b_{pt} : Público potencial da plataforma p .

Todos os parâmetros são não negativos.

Nossa variável de decisão é a alocação de verba em cada plataforma p em cada período t , descrita por x_{pt} .

Agora descrevemos o modelo.

Primeiro, a alocação x_{pt} deve satisfazer a alocação mínima para as plataformas. Isso se traduz na restrição $x_{pt} \geq \ell$ para cada $p = 1, \dots, n$, e $t = 1, \dots, T$.

Segundo, o total de investimento feito deve ser limitado pelo disponível. Isso se traduz em $\sum_p \sum_t x_{pt} \leq A$.

O investimento em cada plataforma em cada mês deve ser menor ou igual ao público potencial multiplicado pelo custo por clique, que representa quantos cliques cada plataforma é capaz de gerar em cada mês dado pela restrição: $x_{pt} \leq c_{pt}b_p$ para cada p .

Com isso chegamos a função objetivo de maximizar as conversões (resultados quaisquer de uma campanha). O custo de conversão do serviço p no período t pode ser calculado por q_{pt}/c_{pt} .

Multiplicando por x_{pt} , e somando para cada p e t chegamos a conversão total.

O modelo completo se encontra abaixo:

$$\max \sum_p \sum_t \frac{x_{pt}}{c_{pt}} q_{pt}$$

$$\text{suj. a } \sum_p \sum_t x_{pt} \leq A$$

$$x_{pt} \leq c_{pt}b_p \quad \forall p = 1, \dots, n, \quad t = 1, \dots, T$$

$$x_{pt} \geq \ell \quad \forall p = 1, \dots, n, \quad t = 1, \dots, T$$

Plataforma	Mês 1	Mês 2	Mês 3	Mês 4	Total
Google Search	R\$ 9.814,00	R\$ 9.520,00	R\$ 9.891,00	R\$ 10.689,00	R\$ 39.914,00
Google Display	R\$-	R\$ -	R\$-	R\$-	R\$ -
Facebook	R\$-	R\$ -	R\$-	R\$-	R\$ -
Instagram	R\$ 8.086,00	R\$ -	R\$-	R\$-	R\$ 8.086,00
					R\$ 48.000,00

Tabela 2: Tabela de Investimento Otimizado Alocado pelo Sistema do Primeiro Plano Avaliado

3.3 Comparações dos resultados obtidos no programa com os planos reais

Para avaliação da qualidade dos dados gerados pelo GLPK na alocação dos investimentos, foram comparados planos desenvolvidos anteriormente, em que a alocação da verba foi feita de forma intuitiva, usando parâmetros similares.

Uma vez que se tratavam de planos que já tiveram suas campanhas executadas, é possível aferir com os dados fornecidos pelas plataformas o volume de conversões registradas e comparar com os resultados apresentados pelo sistema.

Porém, o desafio nestes casos é que as estimativas de público de uma campanha não possuem dados históricos. Além disso, alguns dados estão disponíveis apenas por 12 meses.

Para este plano o resultado atingido foi de 1.598 conversões. A tabela 1 mostra a alocação dos investimentos feita pelo analista de mídia. Já no plano estimado pelo sistema com o GLPK foi de 4.252 conversões, um total 166% superior. Na tabela 2 é possível ver a alocação otimizada.

Na segunda comparação foi avaliado um plano de investimento de R\$23.673,00 no período de Outubro a Dezembro de 2019, onde foram atingidas 2.735 conversões. Já no plano estimado pelo sistema, o resultado foi de 5.727 conversões, total 109% maior.

Um ponto interessante na alocação feita pelo sistema é que este colocou todo o investimento em uma única plataforma, a Google Search.

A Google Search é a rede de anúncios nas pesquisas do Google, onde temos uma precisão maior já que selecionamos palavras-chaves, termos digitados pelos usuários, para exibir os anúncios.

Por exemplo, um usuário que buscar no Google "comprar sapatos" vai receber anúncios relacionados a esta busca. Uma grande vantagem dessa plataforma é seu poder de conversão, que costuma ser superior a de outras redes, porém tem limites no número de usuários que podem ser atingidos, medido pelo volume de buscas em cada mês para cada palavra-chave.

Este dado de volume de buscas é fornecido pelas ferramentas de gerenciamento de mídia do Google, com um histórico de até 12 meses dos dados.

No plano executado este volume de buscas não foi considerado, além disso, foi alocada verba em todas as plataformas em vista de demandas do projeto.

Nas comparações não foram colocados valores de investimentos mínimos por plataforma, o que permitiu ao

sistema alocar todo o recurso apenas no Google Search.

Este tipo de alocação em apenas uma única plataforma não é comum entre os profissionais de mídia por conta das demandas dos clientes e a falta de uso de dados de restrição, porém com a maior estimativa de conversões oferecida pela sistema, existiria a oportunidade de apresentar tal solução para o cliente.

Outra vantagem da alocação em uma única plataforma, seria o menor esforço de manutenção das campanhas por parte das agências. Isso torna ainda mais interessante os resultados apresentados com o GLPK.

4 Conclusão

As agências e empresas têm buscado, cada vez mais, decisões orientadas a dados, onde termos como *data-oriented* ou ainda como *data-driven* despontam nas publicações da área e são jargões repetidos por diversos profissionais. Entretanto, o processo de alocação de investimentos em mídia ainda é feito de forma incipiente, sem relação com os modelos matemáticos disponíveis.

Apesar da popularidade de algoritmos como o Simplex em áreas da indústria e previsão de demanda, no universo de agências de comunicação e Marketing Digital existe uma distância e desconhecimento destas ferramentas.

Com o programa espera-se que os profissionais possam ter melhores resultados em suas campanhas, e também aproximar o universo da estatística e da matemática de uma área ainda muito movida por *big ideas, guts* (grandes ideias ou intuição), permitindo o acesso de qualquer usuário a plataforma.

Além disso, com o código aberto da aplicação outros desenvolvedores podem contribuir com o projeto, que para o futuro pode incorporar novas funcionalidades, como a capacidade de conectar através de APIs (Application Programming Interfaces) as ferramentas de anúncios do Facebook e Google, para extrair dados de forma que os usuários não precisem preencher todas as informações manualmente.

Para o futuro é possível armazenar os dados dos planos gerados para feedback posterior dos profissionais, para gerar análises preditivas sobre a probabilidade de sucesso de cada plano gerado. Além da possibilidade de trabalhar com outras plataformas, ou mesmo formatos variados que estas plataformas apresentam, por exemplo, no Facebook é possível usar diferentes tipos de anúncios e campanhas, que tendem a apresentar resultados diferentes, por isso são incluídas no planejamento

de mídia.

Por hora o projeto prevê apenas quatro plataformas, de maneira agrupada, sem a granularidade dos formatos.

Agradecimentos

Este projeto não seria possível sem a ajuda de diversas pessoas, entre elas quero agradecer imensamente ao Professor Orientador Abel Siqueira e também a Adrian Salceanu, criador do Genie Framework, que em diversas vezes prestou ajuda nas dúvidas sobre como desenvolver a aplicação.

Não posso esquecer também Jean Diogo, pelas incontáveis horas de ajuda, desde o início do projeto nos primeiros esboços em R e no suporte do desenvolvimento do site.

Além de toda comunidade Julia, que de forma solícita respondeu as inúmeras dúvidas sobre a linguagem e suas possibilidades.

Referências

- [1] Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B Shah. Julia: A fresh approach to numerical computing. *SIAM Review*, 59(1):65–98, 2017.
- [2] Iain Dunning, Joey Huchette, and Miles Lubin. Jump: A modeling language for mathematical optimization. *SIAM Review*, 59(2):295–320, 2017.
- [3] GLPK. Glpk. <https://www.gnu.org/software/glpk/>, Acessado em 13 Setembro 2020.
- [4] Mauricio Pereira dos Santos. *Programação Linear*. Departamento de Matemática Aplicada - Instituto de matemática e Estatística - Universidade do Estado do Rio de Janeiro-RJ, 2000.
- [5] Wikipedia. Programação linear. https://en.wikipedia.org/wiki/Linear_programming, Acessado em 13 Setembro 2020.
- [6] Wikipedia. George dantzig. https://en.wikipedia.org/wiki/George_Dantzig, Acessado em 13 Setembro 2020.
- [7] Adrian Salceanu. Genie framework. <https://github.com/GenieFramework/Genie.jl>, Acessado em 13 Setembro 2020.
- [8] Wikipedia. Mvc. <https://pt.wikipedia.org/wiki/MVC>, Acessado em 13 Setembro 2020.
- [9] Felipe Noris. Xlsx.jl. <https://github.com/felipenoris/XLSX.jl>, Acessado em 13 Setembro 2020.
- [10] José Matias. Código-fonte buget app no github. <https://github.com/zehmatias/budget.app>, Acessado em 13 Setembro 2020.