

Superpagg

Uma Ferramenta Crowdsourcing para Comparação de Preços em Supermercados

Paulo Henrique Gouveia Arthur
26138744@sempreunama.com.br
Universidade da Amazônia
Belém, Pará

Bruno Carvalho Santos
26143098@sempreunama.com.br
Universidade da Amazônia
Belém, Pará

Leandro Martha Pontes
26146196@sempreunama.com.br
Universidade da Amazônia
Belém, Pará

Maycon Andrade Duarte
26007345@sempreunama.com.br
Universidade da Amazônia
Belém, Pará

Adonias Pinheiro Pires
adoniaspp@gmail.com
Universidade da Amazônia
Belém, Pará

Alan Marcel Fernandes Souza
alanmarcel2@gmail.com
Universidade da Amazônia
Belém, Pará

ABSTRACT

People often try to spend as little as possible when shopping at a supermarket. Many people can not reduce their costs when buying, either because of lack of time to research or because of the wide variety of products and prices. This article presents the Superpagg software, whose main objective is to help people to spend less on purchases in supermarkets by comparing the prices of different products in different supermarkets, so that the user can see where to pay less for the same products, allowing them to make their purchases with more economy. The results obtained were that the information presented by the software, using a limited sampling, is that the consumer saves from 2.01% to 17.64% monthly per product purchased.

KEYWORDS

supermarket, comparator, prices

1 INTRODUÇÃO

Para muitos consumidores, decidir em qual supermercado fazer compras não é somente uma questão de preço, eles buscam uma experiência completa. Além disso, os grandes *players* do mercado não concorrem somente entre si, mas também com novos modelos de compra online, e novos pontos de venda focados em nichos, bairros, personas e variedade de produtos.

As empresas enfrentam um mercado consumidor cada vez mais exigente e concorrido, com produtos e serviços de melhor qualidade e menores custos. Desta forma, as organizações enfrentam uma verdadeira maratona para conquistar o mercado consumidor e obter maior participação no mercado. Com a globalização e a integração do mercado, surgem centenas de ofertas inovadoras assim como informações para cada tipo ou modelo de produto e serviço.

Em todo o país, as famílias brasileiras têm sempre um grande desafio todo o mês, fazer compras que, por necessidade, não ultrapassem o orçamento dos gastos, o que acaba se tornando uma tarefa das mais cansativas - todos os meses uma busca contínua por bons preços. Segundo levantamento realizado pela [4] na capital paraense, foi concluído que a diferença nos preços de um único

produto pode variar até 100%. Ainda segundo essa pesquisa, utilizar a estratégia de comprar o produto no supermercado mais barato, a economia pode ser de até R\$82,90 por mês e de R\$ 1.059,35 por ano, assim então reduzindo em até R\$ 994,87 nos custos. No Pará, segundo pesquisa do DIEESE [5], o preço final da cesta básica em julho de 2018 chegou a R\$361,11 reais, e apesar de ter registrado queda de 5,38% no período, ainda assim é uma das mais caras do país. A despesa com o orçamento doméstico compromete bastante e depende de uma atenção eficaz para superar esse problema.

Um dos problemas enfrentados pelos consumidores no dia a dia é a diferença de preços entre supermercados, atacados e varejos. Um levantamento feito pela PROTESTE [3] com diversos supermercados por todo o país, mostra que é possível fazer economia de até 29%, ao optar pelo estabelecimento com menores preços. Assim, o preço de R\$361,11 de uma cesta básica pode variar para mais ou para menos em diferentes supermercados.

Como o deslocamento e o tempo tornam inviável uma pesquisa presencial de preços, a utilização de recursos tecnológicos pode auxiliar o consumidor nessa pesquisa, possibilitando aos clientes de supermercados agilidade, facilidade e maior comodidade na hora de comparar os preços dos produtos sem ser preciso visitar os estabelecimentos. Para atingir esse objetivo, foi desenvolvido um software - Superpagg, o qual reunirá os produtos e mostrará o preço dos mesmos em diferentes franquias de supermercado, permitindo ao consumidor escolher em qual loja comprar determinado produto por um menor preço. Esse software apresenta ao consumidor uma listagem ordenada de preços por supermercado e gráficos de variação de preço em diferentes períodos de tempo, o que pode proporcionar ao consumidor uma economia mensal de, por exemplo, 2,01% até 17,64% em um único produto.

Este trabalho apresenta na seção 2 a metodologia utilizada para reunir e processar os dados brutos oriundos de notas fiscais, fazendo a conversão em informação útil. A seção 3 demonstra os resultados obtidos em forma de estatísticas e gráficos e como estes serão apresentados ao consumidor. As considerações finais se apresentam na seção 4.

2 METODOLOGIA

O Superpagg funciona através de *Crowdsourcing* (ver subseção 2.1), um processo colaborativo que utiliza da inteligência coletiva para obter resultados inovadores, refinados, inteligentes e mais precisos [11]. Aplicamos esse conceito utilizando informações fornecidas de

In: XVIII Workshop de Ferramentas e Aplicações (WFA 2019), Rio de Janeiro, Brasil. Anais Estendidos do Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia). Porto Alegre: Sociedade Brasileira de Computação, 2019.
ISSN 2596-1683

forma opcional e voluntária por nossos usuários, de forma prática e simplificada, para gerar novas informações disponíveis para todos os usuários, as quais os ajudarão a economizar em suas compras de supermercado.

2.1 Crowdsourcing

Segundo Howe [8], jornalista criador do termo *crowdsourcing* em um artigo da revista *WIRED*, define que esta técnica é um modo para utilizar o "poder de processamento" de milhões de seres humanos. O conceito consiste em utilizar a contribuição de um número significativo de pessoas para resolver desafios.

2.2 Tecnologias utilizadas

O Superpagg utiliza a técnica de *Web scraping* para interação com o Portal Nota Fiscal Cidadã (SEFA/PA), aplicando soluções em *Node.js 11.15.0*, *XML 1.0*, *PHP 7.2.19* e *MySQL 5.7.26*.

Web scraping é uma técnica para extração de dados disponibilizados em sites de forma automatizada, dados esses que serão posteriormente analisados e combinados para gerar informações úteis.

Node.js ([6]) é um interpretador para códigos desenvolvidos utilizando a linguagem de programação *JavaScript*, focado em lidar com aplicações de alta escalabilidade suportando dezenas de milhares de conexões simultaneamente. Essa tecnologia é utilizada para interagir com o Portal Nota Fiscal Cidadã (SEFA/PA) automaticamente.

XML ([12]) (eXtensible Markup Language ou *Linguagem Extensível de Marcação*) é uma linguagem desenvolvida para descrever dados de forma padronizada, proporcionando sua integridade e interatividade com outras tecnologias. Essa tecnologia é utilizada pela Secretaria da Fazenda (SEFA) de todos os estados brasileiros para armazenar dados de qualquer nota fiscal.

PHP ([7]) (*Hypertext Preprocessor* ou *Pré-processador de Hipertexto*) é uma linguagem de programação utilizada para o desenvolvimento de sites dinâmicos. Essa tecnologia é utilizada para realizar todo o processamento de dados no servidor do Superpagg, interagindo desde informações de notas fiscais obtidas através do *Node.js* até a alimentação de dados para a interface do usuário.

MySQL ([2]) é um sistema de gerenciamento de banco de dados utilizado para guardar todas as informações de notas fiscais, supermercados, produtos e preços obtidos através do *Node.js*, *XML* e *PHP*.

2.3 Extração dos dados

No momento do cadastro de um usuário, é solicitado que informe de forma opcional suas credenciais de acesso ao Portal Nota Fiscal Cidadã, disponibilizado pela Secretaria do Estado da Fazenda (SEFA/PA). Com as credenciais, o sistema poderá realizar a extração de dados para gerar informação a todos os usuários da plataforma. É válido ressaltar que não é permitido um usuário acessar quaisquer informações de outro usuário, incluindo os produtos comprados, valores gastos ou supermercados frequentados.

Com as credenciais de acesso de um usuário, é utilizada a técnica de *data scraping* para acessar o Portal Nota Fiscal Cidadã e extrair os dados referentes às últimas compras em supermercados desse

usuário através das notas fiscais. Esse acesso é feito de forma automática pelo *Node.js*, assim como o download do arquivo *XML* das notas fiscais, que fica temporariamente armazenado em servidores do Superpagg. Esses arquivos são então processados por um software desenvolvido utilizando a linguagem de programação *PHP*, que realiza a extração, filtragem e organização dos dados presentes na Nota Fiscal, como o estabelecimento em que os produtos foram comprados, os produtos comprados e seus respectivos preços. Após esse processamento, estas informações são armazenadas no banco de dados da plataforma e o arquivo *XML* da nota fiscal é deletado permanentemente.

Não há qualquer interferência humana no processo de obtenção e armazenamento dos dados, o procedimento é realizado de forma automática e periódica pelo sistema do Superpagg. O processo descrito acima é resumido através da Figura ?? para melhor entendimento.

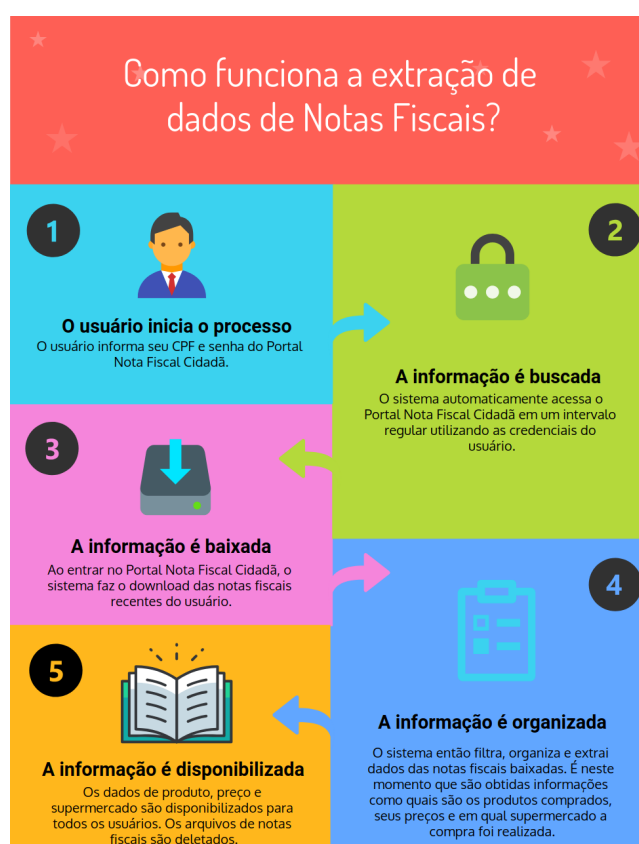


Figure 1: Infográfico demonstrando o funcionamento do Superpagg

2.4 Categorização dos produtos

Para garantir um melhor aproveitamento das informações pelo usuário, o Superpagg utiliza algoritmos de Inteligência Artificial (IA) para disponibilizar informações sobre a categoria de variados produtos da base de dados, esses obtidos automaticamente, conforme explicado nas subseções anteriores. Um serviço completo de

categorização de produtos é apresentado, algo que não seria viável sem a utilização de IA.

Para garantir bons resultados, a plataforma se utiliza do algoritmo Naive Bayes - baseado em redes bayesianas. Um sistema adaptável para trabalhar com informações em probabilidades, utiliza de produtos pré-categorizados manualmente para criar uma base de treinamento. Essa base é posteriormente utilizada para criar uma tabela de probabilidades - baseado no Teorema de Bayes.

Para o funcionamento do algoritmo, são utilizados o nome do produto, tipo (unitário ou KG), código e descrição NCM (Nomenclatura Comum do Mercosul). Com essas informações o algoritmo de Naive Bayes é treinado e é realizada a predição da categoria dos produtos que não possuem categoria definida na base de dados e que não fazem parte da base de treinamento.

Naive Bayes é um algoritmo probabilístico, baseado no Teorema de Bayes, que utiliza dados de treinamento para formar um modelo probabilístico baseado nas evidências das características dos dados. O algoritmo supõe que há uma independência entre as funções do modelo, implicando que a presença de uma determinada função não têm nenhuma relação com as outras. Por exemplo, no produto "Arroz X Y Z", o algoritmo irá separar cada palavra em uma nova *string*, uma independente da outra e, através de suas características ele saberá corretamente qual a categoria do arroz.

Como resultado dessa implementação, o sistema sabe qual produto pertence a qual categoria, tornando a pesquisa mais fácil e detalhada para os usuários.

Para realizar a categorização de produtos através do algoritmo de *Naive Bayes*, foi necessário a criação de uma base de treinamento, portanto, uma categorização manual, 1.262 produtos foram classificados manualmente em 20 categorias.

A implementação do algoritmo foi realizada em Python, através da biblioteca NLTK [1]. Os resultados estão descritos na subseção 2.4.1. Os dados foram pré-processados utilizando técnicas de remoção de stop words e de stemming. *Stop Words* são palavras do idioma que não são relevantes para a classificação de produtos, pois não trazem significado. A técnica de *stemming* é a extração do radical de palavras.

2.4.1 Resultados do algoritmo de categorização. De 1.262 produtos categorizados manualmente, 800 foram utilizados como treinamento e 462 como validação de treinamento. Com esta configuração, foi obtido uma taxa de acerto variável entre 92% a 94%. De 2.386 produtos classificados:

- 463 produtos foram categorizados com mais de 80% de certeza.
- 1.125 foram categorizados com menos de 50% de certeza.
- 605 foram categorizados com menos de 10% de certeza.

Apesar da validação de treinamento do algoritmo ter uma taxa de acerto alta, na classificação de acurácia de produtos reais 1.125 produtos obtiveram menos de 50% de certeza. Isso acontece devido a falta de diversificação de produtos e categorias no treinamento do algoritmo, problema este que será solucionado em futuras melhorias ao sistema.

3 RESULTADOS E DISCUSSÕES

3.1 Benefícios para o consumidor

Todos os usuários do Superpagg, sejam cadastrados ou anônimos, terão livre acesso para visualizar os produtos e comparar seus preços praticados por vários supermercados de Belém-PA e Ananindeua-PA.

Quando um usuário realiza um cadastro no Superpagg e sincroniza sua conta com o Portal Nota Fiscal Cidadã, através do informe de suas credenciais, o consumidor terá uma melhora na sua experiência com a plataforma, tendo acesso a informações como:

- Suas últimas compras em supermercados de Belém e Ananindeua, podendo visualizar estatísticas em gráficos e valores referentes aos produtos comprados, como o valor médio dos produtos no mercado e o valor que seria gasto em outros supermercados;
- Seu histórico de compras em supermercados ao longo dos anos, possibilitando um melhor entendimento dos seus hábitos de consumo e valores gastos em suas compras.

3.2 Avaliação de usabilidade

Foi desenvolvido um formulário de avaliação de usabilidade da plataforma contendo 7 (sete) questões objetivas e 2 (duas) subjetivas opcionais, baseado na ISO 9241-11 que define conceitos como contexto de uso, eficácia e eficiência que são combinados para atingir objetivos específicos.

Durante o período de 1 (uma) semana os usuários da aplicação receberam esse formulário, totalizando 9 (nove) respostas, sendo 8 (oito) de pessoas com formação e/ou experiência na área de Tecnologia da Informação. Os usuários selecionados para enviar esse feedback foram professores e alunos da instituição de ensino dos autores. A maioria das respostas indicam que o Superpagg é uma plataforma que possui uma interface amigável de fácil uso, atendendo com praticidade aquilo que propõe oferecer de informação.

A primeira pergunta é "Como foi sua experiência utilizando o Superpagg?", das 9 (nove) respostas obtidas 77,8% responderam "Boa" e 22,2% responderam "Excelente".

A segunda pergunta é "O que você achou da interface do Superpagg?", das 9 (nove) respostas obtidas 77,8% responderam "Boa", 11,1% responderam "Excelente", 11,1% responderam "Regular".

A terceira pergunta é "Você conseguiu utilizar as funcionalidades do Superpagg?", das 9 (nove) respostas obtidas 88,9% responderam "Sim" e 11,1% responderam "Não".

A quarta pergunta é "Como está o tempo de execução das funcionalidades do Superpagg?", das 9 (nove) respostas obtidas 66,7% responderam "Rápido" e 33,3% responderam "Muito Rápido".

A quinta pergunta é "Foi fácil para você aprender a utilizar o Superpagg?", das 9 (nove) respostas obtidas 100% responderam "Sim".

A sexta pergunta é "O Superpagg lhe oferece recursos necessários para realizar as pesquisas por produto?", das 9 (nove) respostas obtidas 100% responderam "Sim".

A sétima pergunta, subjetiva, é "Quais os pontos positivos e negativos do Superpagg?". Nas respostas, foi identificado que é necessário melhorias em nomes de produtos abreviados e falta de

produtos na base. Assim como nas perguntas objetivas, houve um feedback positivo em relação a interface amigável e intuitiva.

A oitava pergunta, subjetiva, é "Você tem sugestões para o futuro do Superpagg? Se sim, quais?". Sugestões de publicação da plataforma em lojas de aplicativos e melhoria de gráficos foram predominantes.

3.3 Gráficos comparativos de preços

Através de uma listagem ordenada de preços em vários supermercados (demonstrada na Figura ??), o consumidor pode verificar em qual estabelecimento o produto desejado pode ser comprado pelo menor preço. Ao combinar essa informação com gráficos que demonstram a variação do preço de produtos em um determinado período de tempo, o consumidor pode verificar quais supermercados praticam preços estáveis, variação diária de preços, aumento ou diminuição de preços por períodos do ano. Com esses dados ao alcance, o consumidor tem todas as informações necessárias para economizar no seu orçamento.

LEITE EM PÓ		
R\$ 3,49 - Supermercado A		
ORBITO EM 25/04/2018	R\$ 3,49	Supermercado A
ORBITO EM 27/08/2017	R\$ 3,49	Supermercado B
ORBITO EM 02/09/2018	R\$ 3,60	Supermercado C
ORBITO EM 31/07/2018	R\$ 3,89	Supermercado D

Figure 2: Listagem ordenada de preços por supermercado para o Leite em Pó

Os resultados descritos abaixo foram obtidos através de uma amostragem de 8.217 preços de produtos de diferentes tipos, obtidos por informações coletadas de 698 notas fiscais do estado do Pará, oriundas do cadastro de 13 consumidores na plataforma.

Na amostragem obtida, foram selecionados produtos com uma variação significativa de seus preços em um determinado período de tempo. Todos os gráficos podem ser obtidos por qualquer usuário da plataforma ao selecionar um produto específico, essa seleção podendo ocorrer através de busca por palavras-chave (como o nome do produto), através de listagens de produtos como "os mais comprados pelos usuários" e "comprados recentemente pelos usuários".

Dado que a alimentação do banco de dados dos sistema é realizada através de *crowdsourcing*, existem lacunas em pontos específicos dos gráficos onde não há informação do preço praticado pelo supermercado naquele período, como pode ser observado no gráfico da Figura ?? no período de Setembro/2017, Outubro/2017, Maio/2018, Junho/2018, Julho/2018, Julho/2018 e Agosto/2018 para o supermercado A. Essa falta de dados não influencia o resultado final, o qual foi analisado utilizando somente preços conhecidos para se obter a maior precisão possível.

O gráfico mostrado na Figura 3 demonstra a variação do preço médio mensal do Leite em Pó nos supermercados A, B, C e D durante o período de outubro de 2017 até setembro de 2018.

Como demonstra o gráfico, o preço médio mensal do Leite em Pó pode apresentar variação de até 11,4% de um supermercado para outro, observando o período de Julho de 2018. Durante 9 dos 11 meses analisados, o preço médio varia de 2,01% a 11,4%. Ao possibilitar que o consumidor pague sempre o valor mais barato, é possível economizar até 11,4% em 11 meses em apenas um produto supondo que o consumidor compre uma (1) unidade do produto por mês.

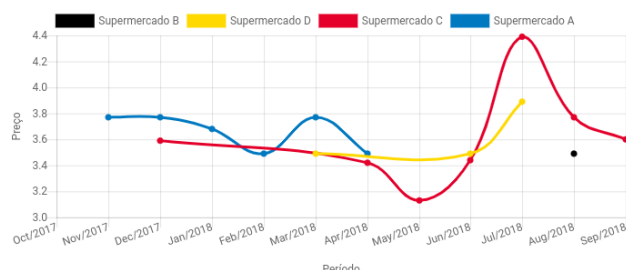


Figure 3: Variação do preço médio mensal do Leite em Pó de outubro de 2017 até setembro de 2018

Ao descrever os gráficos em variações de preço em porcentagem, é possível visualizar que o consumidor pode economizar até 11,4% na compra de uma (1) unidade mensalmente dos produtos citados acima. É sabido que um consumidor compra várias unidades de um mesmo produto durante um período mensal, totalizando uma quantidade ainda maior de unidades desse produto compradas por ano. É difícil estimar a variação do preço de um produto para o futuro, pois isso depende de um grande número de fatores, desde a política interna de preços de um supermercado até a situação política e econômica do país. Esse trabalho propõe que o consumidor saiba quanto ele pode economizar **hoje** ao comprar seus produtos, e isso possibilita que ele possa economizar o máximo possível mensalmente e anualmente.

3.4 Economia

Para efeitos de demonstração, foi criado uma lista de compras utilizando 69 produtos do banco de dados do Superpagg, de diversas categorias. Foi utilizado o preço unitário desses produtos no mesmo instante de tempo.

O preço unitário de cada produto em cada supermercado está descrito no gráfico da Figura 4. Analisando a comparação direta de preços, não é possível perceber uma economia significativa devido a utilização individual dos valores, no entanto, ao analisar o conjunto - como o próprio conceito de lista de compras propõe - percebemos que há uma economia significativa para cada compra realizada, como está descrito no gráfico da Figura 5.

Em determinado supermercado A, o valor total da lista de compras é de R\$ 398,44 e no supermercado B o valor total é de R\$ 376,63. É notável que há uma economia de R\$ 21,81 (5,47%) ao comprar os produtos exclusivamente no supermercado B.

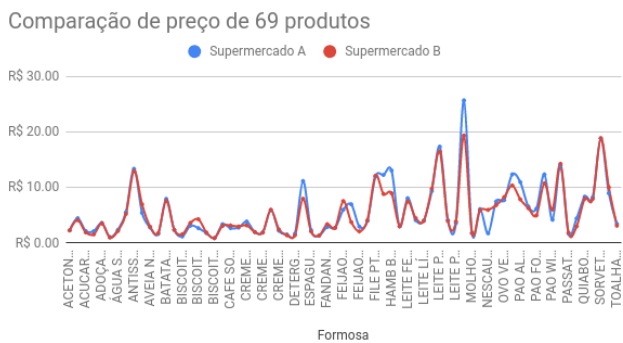


Figure 4: Comparação de preço de 69 produtos da lista de compras



Figure 5: Custo total da lista de compras

Para aumentar o percentual de economia foi analisado o valor de cada produto individualmente, ao invés do total da lista de compras. É possível realizar uma compra em cada supermercado, comprando cada produto pelo seu valor mais barato. Nesse cenário, o total das duas compras é de R\$ 359,55, caracterizando uma economia de R\$ 38,89 (9.77%).

4 CONCLUSÃO

A plataforma Superpagg contribui com cada consumidor ao ajudá-lo a economizar em compras de supermercado, utilizando a tecnologia como seu aliado. Por conta disso, o usuário dispõe de vantagens como o ganho de tempo, evitando o deslocamento a vários supermercados para fazer comparações de preços, uma prática inviável por questões de custo de transporte e tempo necessário. Ao invés disso, o consumidor dispõe de uma interface amigável com informações de produtos e preços em forma de listas e gráficos.

Além de proporcionar comparação de preços, o usuário que sincroniza sua conta com o Portal Nota Fiscal Cidadã, tem a oportunidade de melhorar seus hábitos de consumo através da visualização detalhada de seus gastos em supermercados, ao mesmo tempo em que contribui para o banco de dados da plataforma. Com inspiração em soluções similares de outros estados do Brasil, como QQCusta [10] e Meus Preços [9], a tecnologia do Superpagg proporciona

comodidade para o consumidor realizar suas compras pagando o menor preço possível.

REFERENCES

- [1] Edward Loper Bird, Steven and Ewan Klein. 2009. Natural Language Processing with Python. <https://github.com/php-ai/php-ml>. [O'Reilly Media Inc].
- [2] Oracle Corporation. 2018. MySQL. <https://www.mysql.com>.
- [3] PROTESTE Associação de Consumidores. 2015. Supermercados: preços variam conforme localidade e com pesquisa é possível economizar. <https://www.proteste.org.br/suas-contas/supermercado/noticia/supermercados-compartivos-de-precos-entre-localidades-diferentes-economia>.
- [4] PROTESTE Associação de Consumidores. 2016. Belém tem economia anual de R\$ 1 mil em supermercado. <https://www.proteste.org.br/institucional/imprensa/press-release/2016/belem-tem-economia-anual-de-r-1-mil-em-supermercado>.
- [5] Departamento Intersindical de Estatística e Estudos Socioeconômico DIEESE. 2018. O custo da cesta básica diminuiu em 19 capitais. <https://www.dieese.org.br/analsecestabasica/2018/201807cestabasica.pdf>.
- [6] Node.js Foundation. 2018. Node.js. <https://nodejs.org>.
- [7] The PHP Group. 2018. PHP Manual. <https://php.net>.
- [8] Jeff Howe. 2006. The Rise of Crowdsourcing. <https://www.wired.com/2006/06/crowds/>.
- [9] Priced. 2014. Aplicativo Meus Preços. <http://www.meusprecos.com.br/>.
- [10] QQCusta. 2017. Aplicativo QQCusta. <http://qqcusta.com.br/>.
- [11] Sebrae. 2014. Economia Criativa - Crowdsourcing. http://www.sebraemercados.com.br/wp-content/uploads/2015/12/Maio_EconomiaCriativa_Crowdsourcing.pdf.
- [12] World Wide Web Consortium (W3C). 2018. XML. <https://www.w3.org/XML/>.