Simpósio de Excelência em Gestão e Tecnologia
XVISEGeT
AEDB
FACULDADES DOM BOSCO

Testes A/B como uma das estratégias para inovação em produtos digitais: uma análise bibliométrica

Felipe Neves
felipeneve@gmail.com
CPS

Marcelo Duduchi
mduduchi@gmail.com
CPS

Resumo: Neste artigo é realizada uma análise bibliométrica sobre experimentos online controlados(EOC) do tipo A/B como uma das estratégias para inovação em produtos digitais. Foram pesquisadas 5 bases de conhecimento, com uma seleção de 10 artigos considerados relevantes para avaliação da temática. Os resultados apresentam um volume pequeno de publicações sobre o tema com um crescimento de trabalhos nos últimos dois anos (2017-2018). A ausência de publicações em português indica um grande potencial de pesquisas acerca dessa assunto no âmbito nacional.

Palavras Chave: bibliometria - experimentos online - teste a/b - inovação -

1. INTRODUÇÃO

Os mercados e as empresas vem se transformando há décadas e um dos grandes vetores para essas mudanças é calcado na habilidade das companhias em inovar. Embora a estratégia de inovação englobe uma grande complexidade, existem alguns elementos essenciais que devem ser levados em questão para se testar rapidamente os produtos, validar quais são aqueles que apresentam melhores resultados e entregam mais valor para seus clientes.

A grande velocidade das mudanças trouxe grandes desafios para as empresas e os modelos baseados em processos incrementais e cíclicos tem se tornado cada vez mais presentes nas corporações com um crescimento de uso de metodologias que aprimoram o processo de desenvolvimento de produtos de software.

Nas metodologias ágeis o processo de teste e aprendizado é fundamental para aprimorar produtos/serviços e aprender com os erros, buscando-se maior velocidade na entrega e uma esteira de melhoria contínua dos produtos. É fundamental que as estratégias de testes sejam consistentes e possam aferir quais são os melhores resultados e qual é a melhor experiência para seu cliente.

O uso de experimentos online controlados é uma prática que vem se consolidando e que tem permitido impulsionar resultados e aprimorar a experiência em empresas focadas em inovação, como é o caso da "*Netflix"*, "*Google"*, "*Amazon"*, entre outros. Tais tipos de testes são essenciais para se permitir testar hipóteses e validar qual é a mais adequada para seus clientes.

Esse trabalho surgiu da busca pela relação entre o uso de experimentos controlados online em empresas focadas na inovação e se tais práticas podem se apresentar como uma maneira rápida e efetiva para testar produtos inovadores.

O objetivo desse trabalho é verificar a o volume e a tendência da produção científica sobre o uso de experimentos online controlados, do tipo a/b, como suporte para a evolução de produtos e inovação dentro das empresas.

2. REFERENCIAL TEÓRICO

Para obter sucesso as companhias precisam entender as necessidades dos seus clientes e projetar produtos que entreguem valor (FABIJAN et al., 2018, pp.1). A grande questão é perseguir esse objetivo e atingi-lo com menor custo e maior velocidade possível.

Existem várias estratégias para se conseguir atingir o objetivo de se testar alternativas de produtos digitais, incluindo protótipos, grupos focais, e testes de usabilidade mas, em um contexto onde o fluxo de clientes pode chegar na casa dos milhares, é contraproducente se utilizar de metodologias que necessitam de intervenção de humanos ou avaliações qualitativas. Segundo Kohavi (2013, pp. 1171), "se uma organização deseja tomar decisões baseadas em dados para impulsionar o desenvolvimento de produtos, com o comportamento real dos clientes como fonte de dados para decisões, um dos principais objetivos é permitir a experimentação em escala: suportar a execução de vários experimentos e baixar o custo da experimentação . Isso deve ser feito sem diminuir a confiabilidade do sistema geral.".

Para manter a confiabilidade dos testes e ganhar escala, são necessárias metodologias de experimentação que permitam o teste de hipóteses e que apresentem resultados confiáveis. Para esse objetivo os experimentos online controlados se apresentam como uma alternativa viável e bastante utilizada por grandes empresas com um histórico de criação de produtos digitais inovadores.

Em um experimento online controlado, os usuários são divididos de maneira aleatória entre as opções(FABIJAN et al., 2017, pp. 771), como duas versões de uma mesma página, por exemplo. Essa exposição é persistente, ou seja, cada cliente continuará vendo a mesma versão, A ou B, sempre que retornar à página. A implementação mais comum para esses experimentos é o chamado teste A/B, onde subgrupos comparáveis da amostra são expostos a diferentes experiências para analisar suas diferentes respostas(AMATRIAIN, 2013, pp. 2203). O teste AB é amplamente utilizado para aprimoramento e desenvolvimento de produtos em diversas empresas, como por exemplo, "*Google*", "*Facebook*", "*LinkedIn*", "*Microsoft*", "*Yahoo*", "*Amazon*", "*eBay*", "*Netflix*", "*Zynga*", "*Uber*", "*Airbnb*", "*Pinterest*" e muitos outros(ZHAO et al., 2016, pp. 498).

Um teste A/B tradicional terá apenas duas experiências (A e B), geralmente com duas hipóteses diferentes a serem testadas. Embora a terminologia do teste A/B possa indicar apenas duas versões, na prática o teste pode ter n variações, no "*Netflix"*, por exemplo, eles podem trabalhar entre 5 e 20 células, explorando variações de uma ideia básica (ibidem, pp.2203).

O que mais chama a atenção nos experimentos controlados é sua capacidade de estabelecer uma relação causal entre aquilo que está sendo testado e as mudanças medidas no comportamento do usuário pelos testes(DMITRIEV et al., 2017, pp. 1427). Outro ponto importante é que os testes a/b também podem contribuir para a aceleração da inovação de produtos(XU e et al., 2018).

As hipóteses testadas são geralmente comparadas com métricas definidas pelo negócio, como engajamento, retenção ou efetivação de compra, por exemplo. Tais testes permitem uma visão mais objetiva dos resultados e evitam que viés, cultura organizacional ou mesmo decisões equivocadas de negócio persistam por mais tempo que o necessário. Por mais que pareça algo impensável em uma grande organização, Kohavi aponta o exemplo da "*Microsoft*", reforçando que "após um período inicial de desapontamento pelo fato de que nossos sentimentos e intuição nos afetam com tanta frequência, reconhecemos que a capacidade de separar as ideias realmente boas das demais é um acelerador de inovação e uma competência organizacional central" (KOHAVI et al., 2013, pp. 1175).

O fato de se orientar as decisões não apenas em decisões negociais permite que tais empresas aprimorem suas estratégias de inovação, com foco no valor entregue para o cliente e, principalmente, nos resultados diretos dos testes realizados alinhados à métricas adequadas. A experimentação é fundamental para a inovação de produtos orientada por dados (*ibidem*, 2013, pp. 1173).

Na "*Microsoft*", por exemplo, o grupo que cuida do "*Bing"*, sua ferramenta de pesquisa *online*, trabalham para escalar a experimentação, ou seja "como executar muitos experimentos para acelerar a inovação no desenvolvimento de produtos"(*ibidem*, 2013, pp. 1169).

Nesse caso o tempo e a velocidade da mudança são essenciais para baratear os custos e evitar altas taxas de desistência, muito comuns em produtos digitais. Segundo Amatriain (2013, pp. 2202), eles trabalham com uma busca de um tipo de inovação "que nos permita avaliar idéias rapidamente, de forma barata e objetiva. E uma vez que testamos alguma coisa, queremos entender por que ela falhou ou foi bem-sucedida. Isso nos permite focar no objetivo central de melhorar nossos serviços para nossos membros".

O uso de EOC foi responsável por adicionar centenas de milhões de dólares no resultado da "*Microsoft*"e o impacto alterou profundamente como a área de Pesquisa e Desenvolvimento trabalha(FABIJAN et al., 2017ª, pp. 18) .

Tais questões se apresentam como fundamentais no teste de produtos e grandes empresas orientadas à dados buscam lastrear suas políticas de inovação em metodologias que

envolvem testar a efetividade de seus produtos diretamente com os consumidores(XIE e AURISSET, 2016). No caso do "*Netflix*", por exemplo, os maiores desafios hoje se referem aos testes A/B e da melhoria dos algoritmos de recomendação (GOMEZ-URIBE e HUNT, 2015, pp.13).

Embora essas abordagens sejam utilizadas em grandes companhias, já existem soluções de baixo custo disponibilizadas por diversas *startups* (DMITRIEV, Pavel et al., 2016, pp. 1367), como por exemplo, *Apptimize*, *LeanPlum*, *Optimizely* e *Taplytics*.

3. MÉTODO

Como base para a revisão bibliométrica, foi efetuada uma busca na literatura sobre inovação e sua relação com experimentos on-line controlados e sua manifestação mais comum: o teste A/B. Foram efetuadas buscas em 5 bases relacionadas às áreas de engenharia, sistemas de informação e computação. Os termos pesquisados foram "*online controlled experiments*", "*a/b test**", para se buscar variações como "*a/b test*" e "*a/b testing*", e "innovation".

Na base "*Web of Science"* foram buscados os termos citados nos campos título, resumo, palavras-chave e palavras-chave expandidas, funcionalidade conhecida como Keyword Plus. Nas demais bases foram utilizados os filtros para pesquisa nos campos título e resumo.

Tabela 1: Bases de pesquisa analisadas.

Base de Pesquisa	Documentos encontrados
Elsevier – Engineering Village	18
Scopus	12
Web of Science	10
IEEE Xplore	5
Total	**45**

Fonte: Elaborada pelos autores (2019)

Observa-se na Tabela 1 os resultados das buscas efetuadas nas bases "*Elsevier*", "*Scopus*", "*Web of Science*" e "*IEEE Xplore".* Foram feitas buscas também na "*Emerald*" mas não foram encontrados artigos com as palavras-chave nos títulos ou resumos. Foram encontradas duas ocorrências com os termos buscados no corpo do artigo, mas os resultados não eram relevantes para esse artigo. As bases "*Elsevier*" e "*Scopus*" foram as mais relevantes, concentrando 66% dos 45 artigos encontrados.

Os 45 artigos foram selecionados, organizados e importados para o aplicativo "*Mendeley"*. Do total procedeu-se a exclusão de 25 artigos duplicados. Na sequência foi efetuada a leitura dos resumos dos 20 artigos restantes.

Foram excluídos mais sete artigos seguindo os critérios de aderência ao tema, disponibilidade, tipo de documento e área de pesquisa. Dois artigos não se encontravam disponíveis, dois por estarem fora da área de pesquisa, dois por não serem aderentes ao tema pesquisado e, por fim, um documento era um tutorial e não um artigo.

Os 13 artigos restantes foram organizados e tiveram a inclusão do número de citações obtidas a partir do site "*Google Acadêmico*" (GOOGLE, 2019). Os trabalhos foram ordenados de acordo com a quantidade de citações, do maior para o menor, e foram selecionados os 10 artigos com maior volume de citações, conforme Tabela 2.

Tabela 2: Artigos selecionados conforme quantidade de citações

AUTOR	ANO	TÍTULO
Gomez-Uribe, C.A. Hunt, N.	2015	The netflix recommender system: Algorithms, business value, and innovation
Kohavi, R. Deng, A. Frasca, B. Walker, T. Xu, Y. Pohlmann, N.	2013	Online controlled experiments at large scale
Fabijan, A. Dmitriev, P. Olsson, H.H. Bosch, J.	2017	The Evolution of Continuous Experimentation in Software Product Development: From Data to a Data-Driven Organization at Scale
Amatriain, X.	2013	Beyond data: From user information to business value through personalized recommendations and consumer science
Dmitriev, P. Gupta, S. Kim, D.W. Vaz, G.	2017	A Dirty Dozen: Twelve common metric interpretation pitfalls in online controlled experiments
Fabijan, A., Dmitriev, P., Olsson, H.H., Bosch, J.	2017	The Benefits of Controlled Experimentation at Scale
Dmitriev, P. Frasca, B. Gupta, S. Kohavi, R. Vaz, G.	2016	Pitfalls of long-term online controlled experiments
Zhenyu Zhao ; Miao Chen ; Don Matheson ; Maria Stone	2016	Online Experimentation Diagnosis and Troubleshooting Beyond AA Validation
Xie, H., Aurisset, J.	2016	Improving the sensitivity of online controlled experiments: Case studies at Netflix
Xu, Y. Duan, W. Huang, S.	2018	SQR: Balancing speed, quality and risk in online experiments

Fonte: Elaborada pelos autores (2019)

Após a seleção dos artigos, eles foram lidos e organizados, sendo utilizado o software "*AtlasTI*" para se fazer a análise de ocorrência de palavras e para organizar as informações para a realização de uma avaliação qualitativa.

4. RESULTADOS E DISCUSSÃO

A avaliação de artigos para essa bibliometria não definiu uma janela temporal para restringir a busca nas bases pesquisadas. Embora o tema experimento online controlado não seja novo, as ocorrências vinculadas com o tema inovação são muito recentes, tendo as primeiras ocorrências registradas nesta pesquisa em 2013.

Como foi utilizado o critério de restringir a análise pelo número de citações, é natural que as publicações mais recentes tenham perdido espaço na avaliação e poderiam indicar um resultado inadequado para se apresentar a tendência de publicações na área. Por esta razão foi apresentada a Figura 1 contendo todos os artigos que possuíam aderência ao tema. Nota-se que ouve um aumento de publicações nos últimos 3 anos e a quantidade tem se mantido relativamente estável.

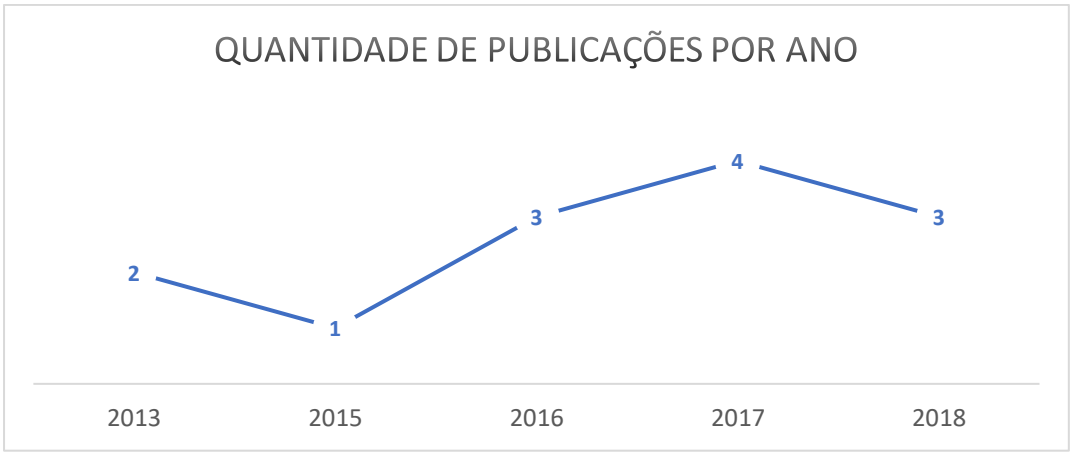

Figura 1: Quantidade de publicações por ano.
Fonte: Elaborada pelos autores (2019)

A base "*Elsevier*" foi a que apresentou resultados mais relevantes, englobando um conjunto de artigos que correspondeu a 90% dos artigos utilizados nesse análise.

Para a avaliação dos países que mais produziram, foi mantido o critério de se apresentar os 13 artigos aderentes ao tema. Para casos onde os pesquisadores eram de países diferentes, foi considerado o país da maioria dos autores dos artigos. O único país que não foi citado é o Reino Unido, com um co-autor. Na Figura 2 é possível verificar que Estados Unidos lidera a produção científica sobre o tema. É natural imaginar a maior quantidade uma vez que grande parte dos artigos versa sobre aplicações em empresas estadunidenses. A Suécia aparece com um número proporcionalmente considerável de artigos, sendo que em três casos os artigos foram produzidos com co-autores dos Estados Unidos. Não foi encontrado nenhum artigo produzido por brasileiros, revelando um potencial de pesquisa sobre o assunto no país.

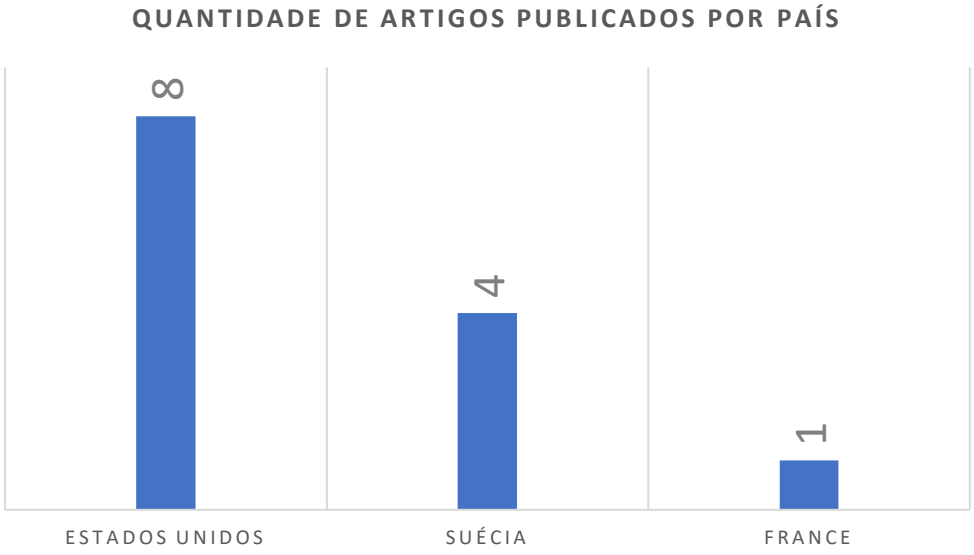

Figura 2: Quantidade de publicações por país.
Fonte: Elaborada pelos autores (2019)

Na seleção dos artigos mais relevantes, os pesquisadores Dimitriev, P., da Microsoft e Fabijan, A. do Dep. De Ciência da Computação de Malmö, na Suécia, participaram de 5 e 4 artigos, respectivamente, sendo os pesquisadores com maior número de publicações sobre o tema no mundo. Com base nas informações pode-se notar uma pequena concentração da produção sobre o tema em alguns núcleos. Com relação às pesquisas, "*Netflix*" e "*Microsoft*" são as empresas mais citadas nos artigos analisados.

A Tabela 3 apresenta a quantidade de citações por artigo, avaliadas pelo "Google Acadêmico". Essa tabela apresenta apenas os 10 artigos mais citados. É possível notar que um apenas 2 artigos superam a marca das 100 citações, tendo um único trabalho com mais número de citações do que todos os outros em conjunto. A quantidade elevada de citações para os primeiros artigos do grupo sugere que há interesse em relação ao tema.

Tabela 3: Artigos selecionados conforme quantidade de citações

Artigos	Citações
The netflix recommender system: Algorithms, business value, and innovation	404
Online controlled experiments at large scale	180
The Evolution of Continuous Experimentation in Software Product Development: From Data to a Data-Driven Organization at Scale	47
Beyond data: From user information to business value through personalized recommendations and consumer science	28
A Dirty Dozen: Twelve common metric interpretation pitfalls in online controlled experiments	21
The Benefits of Controlled Experimentation at Scale	21
Pitfalls of long-term online controlled experiments	19
Online Experimentation Diagnosis and Troubleshooting Beyond AA Validation	14
Improving the sensitivity of online controlled experiments: Case studies at Netflix	13
SQR: Balancing speed, quality and risk in online experiments	5
Experimentation growth: Evolving trustworthy A/B testing capabilities in online software companies	5
Experimentation that matters: A multi-case study on the challenges with A/B testing	3
Reveal 2018	0

Fonte: Elaborada pelos autores (2019)

Foi realizada uma análise do conteúdo dos 10 artigos selecionados e criou-se uma tabela com as palavras com maior número de ocorrências nos trabalhos analisados. Para fins de análise, selecionou-se as 20 palavras mais comuns e foram excluídos termos que não representam áreas ou domínios como pronomes e artigos. Na Figura 3 é possível notar que as palavras com maior número de ocorrências guardam relação direta com o tema analisado, com os termos "*experiment(s)*", "*user(s)*", "*data*", "*metrics*"e "*test*" como as mais citadas.

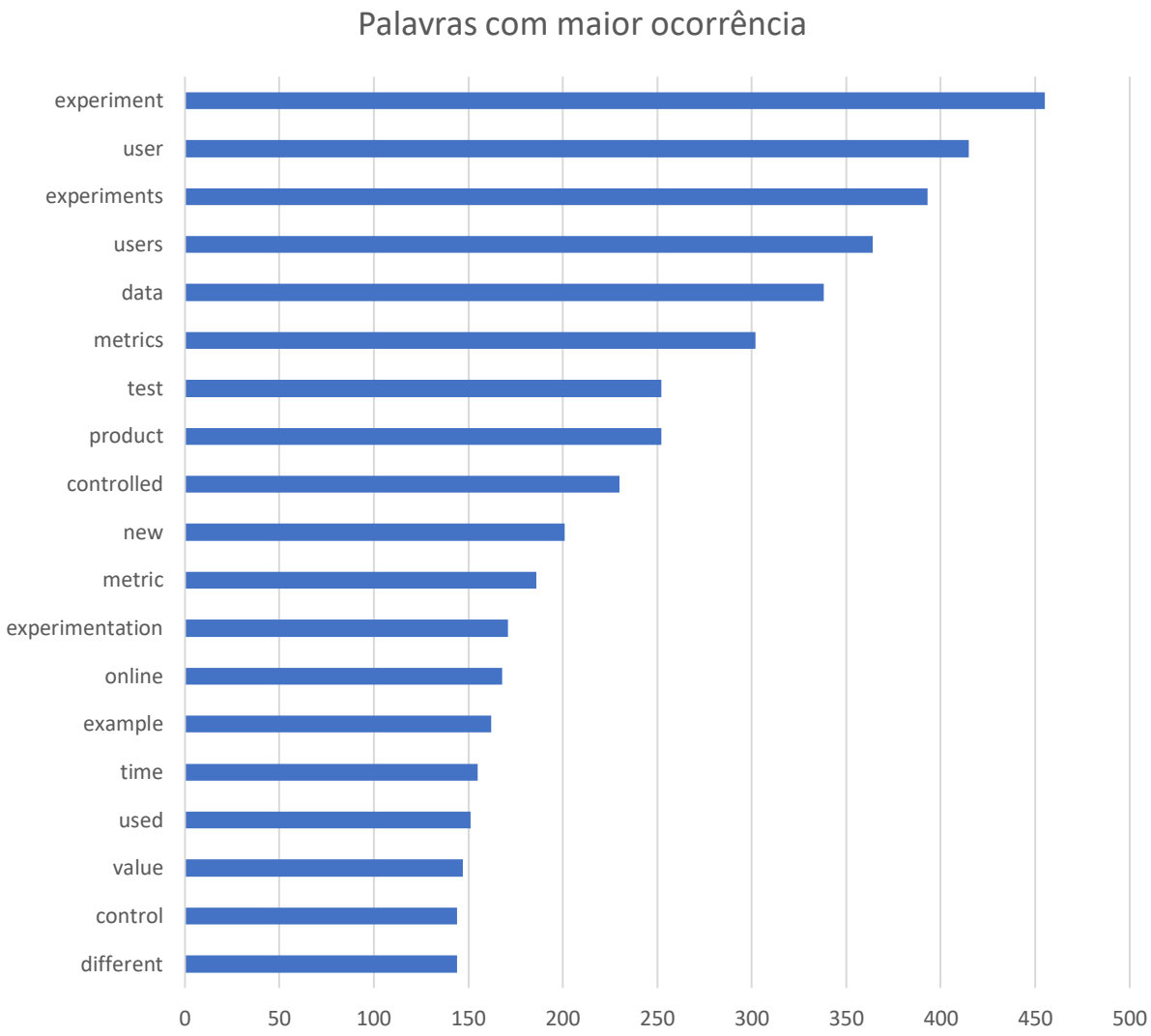

Figura 3: Quantidade de palavras com maior ocorrência nos artigos.
Fonte: Elaborada pelos autores (2019)

Para uma melhor avaliação e contextualização do tema, foi criado um quadro resumo com as principais informações e abordagens avaliadas em cada um deles. Os artigos foram organizados, de acordo com a quantidade de citações, do mais citado para o menos citado.

Tabela 4: Resumo dos artigos

Autor	Ano	Título	Resumo
Gomez-Uribe, C.A. Hunt, N.	2015	The netflix recommender system: Algorithms, business value, and innovation	Apresenta alguns dos algoritmos do sistema de recomendação do Netflix e seu propósito de negócios. A estrutura combina testes A/B com informações off-line de histórico e engajamento. Aponta que o uso de testes facilita a tomada de decisão dos consumidores diminuindo a complexidade e apresentando aquilo que mais importa aos clientes.

(continua)

(continuação)

Autor	Ano	Título	Resumo
Kohavi, R. Deng, A. Frasca, B. Walker, T. Xu, Y. Pohlmann, N.	2013	Online controlled experiments at large scale	Nesse artigo é abordado o caso do Microsoft Bing e seus milhares de experimentos controlados. Os autores apontam que testes em larga escala englobam diversas questões culturais/organizacionais, engenharia e confiabilidade. Apontam que técnicas de testes e debug não conseguem se sustentar em ambientes que exigem um número gigantesco de interações como bilhões de variações de uma mesma página. Apontam que os experimentos online controlados podem fornecer grande flexibilidade e são fundamentais para a inovação de produtos.
Fabijan, A. Dmitriev, P. Olsson, H.H. Bosch, J.	2017	The Evolution of Continuous Experimentation in Software Product Development: From Data to a Data-Driven Organization at Scale	Nesse trabalho os autores utilizam como base o caso da Microsoft e apontam que as empresas de desenvolvimento de software estão cada vez mais se tornando orientadas a dados com uma necessidade de se experimentar continuamente seus produtos com os clientes. Embora os testes A/B sejam conhecidos as empresas raramente são bem sucedidas em evoluir e adotar a metodologia. É apresentado o Modelo de Evolução de Experimentação e suas três fases: técnica, organizacional e evolução do negócio, divididas em 4 categorias.
Amatriain, X.	2013	Beyond data: From user information to business value through personalized recommendations and consumer science	O foco desse artigo é apresentar como o Netflix se apoia em uma estrutura que envolve sistemas de recomendação aliados a testes A/B off-line e online para entregar valor para seus clientes. O sistema de recomendação de filmes é parte essencial da experiência dos seus usuários e todas as decisões são baseadas em dados como popularidade, interesse, contexto, evidências, entre outros.
Dmitriev, P. Gupta, S. Kim, D.W. Vaz, G.	2017	A Dirty Dozen: Twelve common metric interpretation pitfalls in online controlled experiments	Os autores se inspiram nos 12 equívocos de valor p de Steven Goodman para compartilhar doze armadilhas comuns de interpretação métrica que observam repetidamente nos experimentos controlados realizados na Microsoft. São eles Incompatibilidade da Razão de Amostragem Métrica, Interpretação incorreta de métricas de proporção, Viés de perda de telemetria, Assumir que as métricas secundárias não se alteram, considerar sucesso com um valor-p limítrofe, Monitoramento Contínuo e Parada Antecipada, Assumir que a alteração da métrica é homogênea, segmentar interpretações, o impacto dos outliers, novidade e efeito primazia, métricas de funil incompletas e falha na aplicação da lei de Twyman.
Fabijan, A., Dmitriev, P., Olsson, H.H., Bosch, J.	2017	The Benefits of Controlled Experimentation at Scale	Aponta-se que os experimentos online controlados, como os testes A/B, estão sendo utilizados cada vez mais para guiar o desenvolvimento de produtos e acelerar a inovação. Partindo-se do caso da Microsoft, apresentam exemplos de como atingir os benefícios esperados com esse tipo de experimento.
Dmitriev, P. Frasca, B. Gupta, S. Kohavi, R. Vaz, G.	2016	Pitfalls of long-term online controlled experiments	Os EOC são regularmente utilizados para guiar o desenvolvimento do produto e acelerar a inovação em software, entretanto existem armadilhas nos EOC de longo prazo, sendo uma das mais importantes a definição de um Critério de Avaliação Global (OEC). Os resultados não podem ser visualizados apenas em curto prazo pois, uma determinada ação baseada em resultados válidos de um teste A/B podem levar a uma decisão desastrosa a longo prazo. Um exemplo é uma empresa que dificulta a saída de uma determinada seção do site: vai aumentar o tempo médio de sessão dos usuários mas a desistência deles será catastrófica para o negócio.

(continua)

(continuação)

Autor	Ano	Título	Resumo
Zhenyu Zhao ; Miao Chen ; Don Matheson ; Maria Stone	2016	Online Experimentation Diagnosis and Troubleshooting Beyond AA Validation	Os autores apontam situações que podem levar a conclusões erradas quando ocorrem problemas de desequilíbrio em testes A/B. Com base na experiência que tiveram no Yahoo, reforçam a necessidade da validação do grupo teste e de controle com testes A/A e apontam três categorias de problemas recorrentes: instrumentação, divisão de tráfego e outliers.
Xie, H., Aurisset, J.	2016	Improving the sensitivity of online controlled experiments: Case studies at Netflix	Neste artigo os autores apresentam o impacto que pequenas mudanças podem causar em empresas que são orientadas a dados, como Netflix, e aponta estratégias para melhorar a precisão de EOCs.
Xu, Y. Duan, W. Huang, S.	2018	SQR: Balancing speed, quality and risk in online experiments	Embora testes A/B sejam utilizados amplamente para acelerar inovações de produtos, os autores apontam enorme ineficiência e risco na forma como os experimentos são realizados, e isso está atrapalhando a inovação. Utilizando a experiência deles no Linkedin e com foco no aprimoramento do lançamento de produtos, eles propõem um modelo que balanceia Velocidade, Qualidade e Risco (SQR).

Fonte: Elaborada pelos autores (2019)

5. CONSIDERAÇÕES FINAIS

A utilização de metodologias de testes controlados é uma prática que vem sendo utilizada constantemente em empresas como uma das estratégias para permitir a evolução rápida de produtos digitais. O uso de testes a/b como principal método para EOC é reforçada pela literatura e pela prática de grandes empresas cujo foco é a inovação. EM grande parte dos trabalhos analisados reforça-se que o uso de EOC permite que os produtos digitais sejam avaliados rapidamente e alterados dinamicamente para promover uma experiência do usuário aprimorada e resultados mais confiáveis para os clientes.

A análise bibliométrica sobre experimentos online controlados(EOC) do tipo A/B como uma das estratégias para inovação em produtos digitais apresentou um volume pequeno de publicações sobre o tema com um crescimento de trabalhos nos últimos três anos (2016-2018). A ausência de publicações em português indica um grande potencial de pesquisas acerca dessa assunto no âmbito nacional. A análise dos artigos aponta que o uso de testes A/B como estratégia para validação rápida de resultados vem se consolidando entre as grandes empresas inovadoras e tem um papel importante, embora não principal, para as disciplinas de Desenvolvimento de Software, Experiência do Usuário e Inovação.

6. REFERÊNCIAS

BEACH, R.; MUHLEMANN, A. P.; PRICE, D. H. R.; PATERSON, A. & SHARP, J. A. A review of manufacturing Flexibility. European Journal of Operational Research, v. 122, 2000, pp. 41-57.

AMATRIAIN, X. Beyond Data: From User Information to Business Value through Personalized Recommendations and Consumer Science. CIKM '13 Proceedings of the 22nd ACM international conference on Information & Knowledge Management, p. 2201–2208, 2013.

DMITRIEV, P.; GUPTA, S.; KIM, D.W.; VAZ, G. A Dirty Dozen: Twelve common metric interpretation pitfalls in online controlled experiments. 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2017, v. Part F1296, p. 1427–1436, 2017.

DMITRIEV, P.; FRASCA, B.; GUPTA, S.; KOHAVI, R.; VAZ, G. Pitfalls of long-term online controlled experiments. Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016, p. 1367–1376, 2016.

FABIJAN, A.; DMITRIEV, P.; MCFARLAND, C.; VERMEER, L.; OLSSON, H. H.; BOSCH, J. Experimentation growth: Evolving trustworthy A/B testing capabilities in online software companies. Journal of Software: Evolution and Process, v. 30, n. 12, p. 1–23, 2018.

FABIJAN, A.; DMITRIEV, P.; OLSSON, H. H.; BOSCH, J. The benefits of controlled experimentation at scale. Proceedings - 43rd Euromicro Conference on Software Engineering and Advanced Applications, SEAA 2017, p. 18–26, 2017a.

FABIJAN, A.; DMITRIEV, P.; OLSSON, H. H.; BOSCH, J. The Evolution of Continuous Experimentation in Software Product Development: From Data to a Data-Driven Organization at Scale. Proceedings - 2017 IEEE/ACM 39th International Conference on Software Engineering, ICSE 2017, p. 770–780, 2017b.

GOMEZ-URIBE, C. A.; HUNT, N. The Netflix Recommender System : Algorithms , Business Value ,. ACM Transactions on Management Information Systems (TMIS), v. 6, n. 4, 2015.

GOOGLE. Google acadêmico. Disponível em <http://scholar.google.com.br>, acesso em 02/06/2019.

KOHAVI, R.; DENG, A.; FRASCA, B.; WALKER, T.; XU, Y.; POHLMANN, N. Online controlled experiments at large scale. p. 1168, 2013.

XIE, H.; AURISSET, J. Improving the Sensitivity of Online Controlled Experiments. p. 645–654, 2016.

XU, Y.; DUAN, W.; HUANG, S. SQR : Balancing Speed , Qu ality and Risk in Online Experiments. n. 1, p. 1–9, 2018.

ZHAO, Z.; CHEN, M.; MATHESON, D.; STONE, M. Online experimentation diagnosis and troubleshooting beyond AA validation. Proceedings - 3rd IEEE International Conference on Data Science and Advanced Analytics, DSAA 2016, p. 498–507, 2016.