
Exploração cognitiva, Medalhas virtuais e IA: Um olhar crítico sobre o Kaggle, a maior plataforma de competições de aprendizado de máquina.¹

Luciano FRIZZERA²
Concordia University, Montreal, Canada

RESUMO

A IA está cada vez mais utilizada como uma solução para problemas socio-tecnológicos. Pouco se sabe sobre os locais onde estas soluções são prototipadas, testadas e refinadas, muitas vezes realizadas por voluntários e trabalhadores com baixos salários. Este artigo se concentra no Kaggle, uma plataforma digital gamificada onde programadores participam de desafios de aprendizado de máquina em troca de medalhas virtuais. Utilizando métodos digitais e análise do discurso, este artigo examina a história do Kaggle e seu modelo econômico de desenvolvimento de IA de baixo custo baseado em crowdsourcing e propõe que sua lógica de negócios impulsiona a exploração de habilidades cognitivas como um recurso computacional.

PALAVRAS-CHAVE: Estudo de Plataformas; Inteligência Artificial; Estudo Crítico de IA; Kaggle

CORPO DO TEXTO

Este artigo traça os primeiros dez anos do Kaggle, oferecendo uma narrativa da estratégia seu fundador, Anthony Goldbloom, e seus parceiros para construir a maior plataforma de competições de aprendizado de máquina do mundo. Situado entre a academia e o mercado, Kaggle não apenas reúne maior comunidade de ciência de dados do mundo, como também é um local de experimentação onde nascem novas ideias para desenvolvimento de Inteligência Artificial (IA). A empresa foi pensada seguindo a cultura startup e a Ideologia Californiana (Barbrook e Cameron, 1996) para aludir aos entusiastas de tecnologia e aos capitalistas de risco, prometendo uma revolução digital com baixo custo e alto lucro. De 2010 a 2020, a plataforma organizou mais de 3.000 competições de aprendizado de máquina, distribuiu mais de US\$ 16 milhões em prêmios e atraiu mais de 5 milhões de usuários que contribuíram para a criação de dezenas de milhões de modelos preditivos.

Como empresa, abraça a lógica da economia de plataforma (Srnicsek, 2017; van Dijck, 2013) e plataformização do trabalho (Grohmann, 2020), servindo como intermediário indispensável entre grandes corporações que procuram otimizar os seus

¹ Trabalho apresentado no GP Tecnologias e Culturas Digitais, XXIV Encontro dos Grupos de Pesquisas em Comunicação, evento componente do 47º Congresso Brasileiro de Ciências da Comunicação.

² Doutorando. Communication Studies. Concordia University: luciano.frizzera@me.com

negócios e desenvolvedores ávidos por encontrar emprego. Na sua essência, Kaggle incorpora a crença de que os modelos estatísticos por si só podem resolver qualquer problema político, social e econômico, um conceito intimamente relacionado com a formação de Goldbloom em Econometria. Kaggle é um lugar crucial para estudar como as normas sociais e políticas são codificadas em modelos preditivos que moldam as relações de poder, afeto e comportamento.

Usando Métodos Digitais (Rogers, 2009) e Análise de Discurso (Rose, 2001), esse artigo examina relatórios financeiros da Alphabet (empresa dono do Google), entrevistas, apresentações, e blog posts feito por Goldbloom e seus parceiros, e o arquivo da homepage do Kaggle (capturado 4.610 vezes pelo Internet Archive entre 2010 e 2020). O objetivo é mostrar e discutir como a empresa evoluiu e materializou sua lógica de negócio e as relações de poder dentro da plataforma. A partir dos dados colhidos, a história da empresa pode ser dividida em cinco fases: (1) protótipo; (2) startup organizadora de hackathons; (3) consultoria em big data e agência de empregos temporários; (4) plataforma de mídia social; e (5) subsidiária do Google Cloud a partir do momento que foi adquirida pela Alphabet.

Fase 1: Protótipo

Em abril de 2010, Anthony Goldbloom e Ben Hamner anunciaram um prêmio de US\$1.000 para um algoritmo que pudesse prever a votação para o Festival de Música *Eurovision* daquele ano. Em apenas seis semanas, eles recrutaram 25 pessoas. Esta competição foi a prova de conceito que Goldbloom e Hamner precisavam para estabelecer o Kaggle como uma plataforma para crowdsourcing de desenvolvimento de aprendizado de máquina com base em algoritmos competitivos. O Kaggle foi lançado alguns meses antes dessa competição, em fevereiro de 2010. No processo de descobrir o que Kaggle deveria ser, Goldbloom deixou vestígios de suas inspirações e objetivos iniciais na primeira versão do site. Por exemplo, ele viu uma oportunidade de negócio na previsão de Hal Varian (2009), o mentor por trás das estratégias de capitalismo de vigilância do Google, em uma entrevista para o *The McKinsey Quarterly*: “Continuo dizendo que o trabalho mais sexy nos próximos dez anos será o de estatístico” (n.p, tradução nossa).³ Kaggle usaria esta previsão para construir seu modelo de negócio,

³ Na original: “I keep saying that the sexy job in the next ten years will be statisticians”

atrair a força de trabalho potencial através de declarações sobre a potencialidade da estatística, e explorar a mão de obra gratuita na Internet.

Fase 2: Startup organizadora de hackathons

Em seu primeiro ano na Califórnia, a empresa fez parceria com a *Heritage Provider Network*, uma empresa de seguros de saúde, para lançar uma competição de US\$ 3 milhões. O desafio trouxe 1.500 novos usuários e US\$ 11 milhões na primeira rodada de investimento Series A (Crunchbase, 2021), demonstrando a rapidez com que o Kaggle poderia crescer. O investimento veio principalmente do capital de risco já envolvido na indústria de tecnologia digital, entre eles o próprio Hal Varian e Max Levchin, um dos fundadores do PayPal, que logo se tornou chairman do Kaggle.

Não por acaso Goldbloom passou a apresentar o Kaggle como uma “*plataforma* para competições de previsão de dados que permitem às organizações publicar seus dados e tê-los examinados pelos melhores cientistas de dados do mundo” e chamou a atenção para seu objetivo de “fazer da ciência de dados *um esporte*” (Kaggle, 2011, tradução e grifo nosso).⁴ Por um lado, ao utilizar “plataforma” junto à sua marca, a empresa não só sinalizou que era um player igual a outras plataformas digitais como Facebook e LinkedIn, mas também reforçou conceitos e práticas associadas a esta palavra-chave no ambiente técnico de negócios. Goldbloom usou essa construção gerencial para preparar o caminho para o desenvolvimento de uma infraestrutura mais extensa que moldaria as relações de trabalho que os usuários estabelecem não apenas com a plataforma, mas também entre eles mesmos, os dados e os códigos que eles criam. Por outro lado, ao tornar a ciência de dados um “esporte”, o Kaggle sistematizou um modelo de negócios baseado na competição em uma experiência gamificada para atrair engenheiros de computação. A gamificação funciona incentivando os utilizadores a envolverem-se em comportamentos desejados, mostrando um caminho para o “domínio” e a “autonomia,” aproveitando a predisposição psicológica dos humanos para se envolverem em jogos (Whitson, 2013). Dessa forma, os usuários do Kaggle não estariam vendendo sua força de trabalho, mas sim se divertindo na plataforma.

Fase 3: Consultoria em big data e agência de empregos temporários

Entre 2013 e 2015, com Goldbloom figurando duas vezes na lista de palestrantes do Fórum Econômico Mundial (2013; 2015), Kaggle recebeu atenção e reconhecimento

⁴ Na original: “platform for data prediction competitions that allow organizations to post their data and have it scrutinized by the world’s best data scientists” e “making data science a sport”

internacional como um modelo de negócios e infraestrutura que estava promovendo o desenvolvimento de IA. No entanto, a estratégia da empresa se revelou falha ao não produzir excedente: a empresa não era lucrativa. Apesar de ter realizado 43 competições em 2012, três vezes mais do que em 2011, os prêmios distribuídos foram três vezes inferiores, obrigando a empresa a promover mudanças estratégicas para atrair mais investidores e patrocinadores para poder sobreviver. Entre 2013 e 2016, a plataforma mudou significativamente quatro vezes, mudando o foco da competição de ciência de dados centrada nos usuários para uma consultoria e oferta de mão de obra centrada na indústria, assemelhando-se, em última análise, a uma agência de empregos temporários.

Fase 4: Plataforma de mídia social

Em 2016 Kaggle se reinventa e volta as suas origens: um organizador de hackathon; mas agora com atributos de mídia social, onde os usuários podiam seguir e curtir o código uns dos outros. Prevendo a importância da computação em nuvem, a empresa investiu fortemente em datacenters, introduzindo novos serviços que permitiram aos usuários utilizar uma vasta infraestrutura digital para treinar os seus modelos em máquinas equipadas com potentes processadores. Nesse momento a plataforma também abriu a possibilidade de compartilhar datasets focado em IA em repositórios públicos. Em 2020 havia mais de 55 mil datasets do gênero na plataforma.

O que é significativo neste período é que o Kaggle mudou o foco do gerenciamento de competições para uma conectividade mais profunda entre conjuntos de dados, código e usuários. Seu novo slogan era: “te ajudar a aprender trabalhar, e se divertir.” Combinando estes três elementos com as competições realizadas na plataforma, a empresa poderia ter mais controle sobre o trabalho realizado pela comunidade, explorando todo o potencial de uma *assemblage* (Deleuze e Guattari, 1987; Grusin, 2015) entre dados humano, e máquina para (re)produzir modelos algorítmicos preditivos que eventualmente serão implantados para uso comercial. Esta mudança de estratégia tornou o Kaggle financeiramente atraente novamente, acumulando US\$ 5 milhões em uma segunda rodada de investimentos Series A (Crunchbase, 2021).

Fase 5: Subsidiária do Google Cloud

A Alphabet adquiriu o Kaggle em março de 2017 por um valor não revelado como forma de garantir uma boa posição no mercado de IA e aprendizado de máquina (Lardinois et al., 2017). O sucesso desta decisão pode ser medido pela rápida expansão

da plataforma, atingindo a marca de cinco milhões de utilizadores em 2020 (Kaggle, 2020). Embora fosse um passo importante para a empresa da Goldbloom, a aquisição não parecia importante do ponto de vista dos investidores. De 2015 a 2022, o Kaggle foi mencionado apenas uma vez nas reuniões entre investidores e nos relatórios anuais da Alphabet (2016). O Kaggle se tornou um lugar para vender a infraestrutura de nuvem do Google juntamente com o trabalho realizado por uma legião de programadores que, na maioria das vezes, trabalham de graça.

Explorando poder cognitivo do homem-máquina

Em sua palestra na conferência Digital-Life-Design de 2013, Max Levchin sintetizou o modelo de negócios apresentado por Goldbloom e elogiado pela Alphabet: “Assim como o protetor de tela SETI@Home ‘rouba’ os ciclos da CPU para filtrar o ruído de rádio cósmico em busca de vozes alienígenas, o firmware do seu plug cerebral lhe renderá um pouco de dinheiro extra enquanto você dorme, sendo programado remotamente para resolver problemas” (Para. 21, tradução nossa).⁵ Ele conclui dizendo que, como espécie, “*devemos* usar todos os recursos disponíveis ao máximo” (Para. 24, tradução e grifo nosso),⁶ sugerindo que a melhor maneira de otimizar a sociedade é explorar o tempo livre e o poder cognitivo dos indivíduos como se fossem máquinas. Ao equiparar o cérebro humano à uma máquina, Levchin revela falta de empatia e desconexão com outros indivíduos, sugerindo que pessoas são apenas recursos que não estão sendo “utilizados corretamente” ou “eficientemente.” Ele está convencido de que sensores digitais, big data e IA devem ser usados para melhorar os seres humanos em seu nível fundamental, semelhante à forma como os programadores otimizam os ciclos de processamento de um computador. Ele vê o modelo de negócios do Kaggle como um caminho a seguir: “Hoje você já pode alugar o cérebro de um gênio da mineração de dados via Kaggle por hora, amanhã por hora cerebral” (Parágrafo 21, tradução nossa).⁷

No Kaggle, os desenvolvedores, assim como datasets e algoritmos, são matérias-primas que devem ser exploradas sistematicamente para extrair o máximo de precisão, desempenho e lucro. O trabalho do participante (código, análise, modelos) deve encontrar padrões no comportamento e nas relações sociais dos indivíduos para produzir

⁵ No original: “Just like the SETI@Home screensaver ‘steals’ CPU cycles to sift through cosmic radio noise for alien voices, your brain plug firmware will earn you a little extra cash while you sleep, by being remotely programmed to solve hard problems”

⁶ No original: “we *must* use all available resources to their maximum”

⁷ No original: “Today you can already rent the brain of a data-mining genius via Kaggle by the hour, tomorrow by brain-hour”

modelos preditivos através dos quais empresas privadas possam explorar clientes, cortar custos e aumentar os lucros. No processo de produção destes modelos, a própria comunidade é explorada pela plataforma como mão de obra gratuita. Tal como afirma Terranova (2004), a mão de obra gratuita é “uma característica da economia cultural em geral e uma importante [...] fonte de valor nas sociedades capitalistas avançadas” (p. 73, tradução nossa).⁸ Com a descentralização e desorganização do local de trabalho promovida pelas redes digitais, o trabalho tornou-se fluido e flexível, misturado com outras atividades. O trabalho gratuito é fundamental para empresas neoliberais que pretendem aproveitar toda e qualquer forma de produção, transformando cada ação em mercadoria a ser vendida no varejo ou no atacado.

Uma única competição no Kaggle mobiliza milhares de pessoas que competem em seu tempo livre por uma fração da remuneração dos cientistas de dados contratados pelas empresas de tecnologia. Kagglers, como a comunidade se autodenomina, são responsáveis por fazer upload de datasets, construir e depurar código e treinar aprendizado de máquina em seu tempo livre. Estas tarefas são gamificadas (Whitson, 2013) para tornar a plataforma atrativa para um sistema econômico fundamentado no individualismo, na meritocracia e na competitividade (Lazzarato, 2004). O Kaggle surge como um mecanismo que molda e define as condições sociais através das quais o desenvolvedor se compreende na plataforma como sujeito produtivo. Simultaneamente, esses sujeitos codificam esses valores em algoritmos de aprendizado de máquina e modelos preditivos construídos na plataforma para eventualmente ser usado por outras empresas.

O modelo de Kaggle deixa os desenvolvedores ansiosos por participar de competições não apenas por prêmios em dinheiro, mas também porque procuram oportunidades de se envolver com a comunidade, de aprender uns com os outros, e ter uma noção de pertencimento. A comunidade está sempre pronta para participar e contribuir no desenvolvimento de novos modelos preditivos, independentemente da intenção destes algoritmos ou do impacto que possam ter na vida cotidiana. Como consequência, a plataforma explora as aspirações afetivas dos usuários para construir um “estoque de cérebros,” uma mercadoria que, por sua vez, se torna uma força de

⁸ No original: “a feature of the cultural economy at large, and an important ... source of value in advanced capitalist societies”

trabalho disposta a trocar suas habilidades cognitivas, senão de graça, pelo menos por item de baixo valor, como medalhas virtuais.

Conclusão

Provedores de IA e plataformas digitais como Kaggle definem os problemas a serem explorados, os objetivos e métricas de sucesso, e suas possíveis soluções. A lógica empresarial do desenvolvimento da aprendizagem automática empurra os algoritmos para acelerar a economia, produzindo uma condição de existência onde todos os atores envolvidos—humanos, não humanos, máquinas—exercem um papel específico. Essa lógica utiliza todos os recursos disponíveis para criar um mundo no qual os seus produtos e serviços, juntamente com o consumidor e o trabalhador, possam existir e, ao mesmo tempo, o mundo esteja, por sua vez, profundamente inscrito nos seus corpos e mentes.

REFERÊNCIAS

- Alphabet. (2016). *Alphabet's (GOOG) Management presents at Credit Suisse Technology, Media and Telecom Conference (Transcript)*. <https://seekingalpha.com/article/4027135-alphabets-goog-management-presents-credit-suisse-technology-media-and-telecom-conference>
- Barbrook, R., & Cameron, A. (1996). The Californian ideology. *Science as Culture*, 6(1), 44–72. <https://doi.org/10.1080/09505439609526455>
- Crunchbase. (2021). *Kaggle—Crunchbase Company Profile & Funding*. Crunchbase. <https://www.crunchbase.com/organization/kaggle>
- Deleuze, G., & Guattari, F. (1987). *A Thousand Plateaus: Capitalism and Schizophrenia* (B. Massumi, Trans.). University of Minnesota Press.
- Grohmann, R. (2020). Plataformização do trabalho: Entre dataficação, financeirização e racionalidade neoliberal. *Revista Eletrônica Internacional de Economia Política da Informação da Comunicação e da Cultura*, 22(1), Article 1.
- Grusin, R. (2015). Radical Mediation. *Critical Inquiry*, 42(1), 124–148. <https://doi.org/10.1086/682998>
- Kaggle. (2011, May 27). *Data mining, forecasting and bioinformatics competitions on Kaggle*. Kaggle. <http://web.archive.org/web/20110527212123/http://www.kaggle.com/>
- Lardinois, F., Lynley, M., & Mannes, J. (2017, March 17). Google is acquiring data science community Kaggle. *TechCrunch*. <https://social.techcrunch.com/2017/03/07/google-is-acquiring-data-science-community-kaggle/>
- Levchin, M. (2013, January 21). *DLD13 Keynote* [Tumblr]. <https://max.levch.in/post/41116802381/dld13-keynote>

Rogers, R. (2009). *The End of the Virtual: Digital Methods*. Amsterdam University Press. <https://doi.org/10.5117/9789056295936>

Rose, G. (2001). *Visual methodologies: An introduction to the interpretation of visual materials*. Sage.

Srnicek, N. (2017). *Platform Capitalism*. Polity.

Terranova, T. (2004). *Network Culture: Politics for the Information Age*. Pluto Press.

van Dijck, J. (2013). *The Culture of Connectivity: A Critical History of Social Media*. Oxford University Press.

Varian, H. (2009, February 2). *Hal Varian on how web challenge managers* (J. Manyika, Interviewer) [McKinsey Quarterly 1(2.2)]. https://web.archive.org/web/20090202002651/http://www.mckinseyquarterly.com/Hal_Varian_on_how_the_Web_challenges_managers_2286.

Whitson, J. R. (2013). Gaming the Quantified Self. *Surveillance & Society*, 11(1/2), 163–176. <https://doi.org/10.24908/ss.v11i1/2.4454>

World Economic Forum (Director). (2013, August 27). *Technology Pioneer 2014*. <https://www.youtube.com/watch?v=OShGuf7QeJY>

World Economic Forum (Director). (2015, January 22). *Davos 2015—A Brave New World*. <https://www.youtube.com/watch?v=wGLJXO08IYo>