

Desafios para o sucesso na gestão de projetos na área de Ciência de Dados

JAQUELINE LOPES DIAS
UNIVERSIDADE DE SÃO PAULO (USP)

RAQUEL SILVEIRA RAMOS ALMEIDA
UNIVERSIDADE DE SÃO PAULO (USP)

DESAFIOS PARA O SUCESSO NA GESTÃO DE PROJETOS NA ÁREA DE CIÊNCIA DE DADOS

INTRODUÇÃO

A ciência de dados consolidou-se como uma área de conhecimento relevante, utilizada em diversos setores, tais como finanças, tecnologia, serviços, saúde, varejo, agronegócio, transporte, entre outros. Nos últimos anos, empregaram-se intensos esforços de pesquisa para o desenvolvimento de análises avançadas, aprimoramento de modelos de dados e criação de novos algoritmos para extrair conhecimento e valor de grandes volumes de dados (Martinez; Viles; Olaizola, 2021b). Observou-se um interesse crescente na área por parte da indústria, devido ao aumento da capacidade de coletar, armazenar e analisar dados gerados com uma frequência cada vez maior. Sua expansão criou novas áreas e carreiras, caracterizou-se como um desenvolvimento contínuo que afetou diretamente como as organizações utilizam dados para obter informações úteis e apoiar a tomada de decisões.

Constatou-se a natureza multidisciplinar da ciência de dados, envolvendo a matemática, estatística e ciência da computação (Saltz, 2015; Espinosa; Armour, 2016; Martinez; Viles; Olaizola, 2021a). Consequentemente, os desafios que essa área enfrenta também são múltiplos, desde aspectos puramente técnicos a questões organizacionais (processuais, sociais e culturais) que se apresentam ao executar um projeto de ciência de dados (Saltz, 2017). Entre os principais desafios, identificou-se: falta de coordenação, colaboração e dificuldades de comunicação, dificuldade para estruturar equipes analíticas especializadas, dificuldade para definir a priori o escopo do projeto e pouco envolvimento das partes interessadas (Martinez; Viles; Olaizola, 2021a). Além disso, o artigo *Data Science Methodologies: Current Challenges and Future Approaches* destacou que muitos projetos de ciência de dados não chegaram à produção devido à falta de visão clara, objetivos mal definidos e uma ênfase excessiva em questões técnicas, o que refletiu um problema de maturidade nos processos e definição ambígua dos papéis dentro da equipe. De acordo com o relatório da *VentureBeat* (2019), 87% dos projetos de ciência de dados não chegaram à produção. Além disso, a pesquisa da *NewVantage Partners* (2019) apontou que 77% das empresas enfrentaram dificuldades na adoção de iniciativas de Big Data e Inteligência Artificial (IA), com 95% desses desafios sendo atribuídos a barreiras culturais, como resistência organizacional e falta de alinhamento, enquanto apenas 5% se referiram a questões tecnológicas.

Verificou-se que o sucesso de um projeto de ciência de dados não depende apenas de fatores técnicos, embora a maior parte das pesquisas até o momento tenha se concentrado nas capacidades de modelagem e nos aspectos computacionais (Lahiri; Saltz, 2023). Os desafios enfrentados revelam sintomas de um problema mais amplo: a ausência de metodologias robustas que orientem adequadamente o gerenciamento desses projetos (Saltz, 2015; Martinez; Viles; Olaizola, 2021a; Lahiri; Saltz, 2023). Tal lacuna entre a evolução técnica e a aplicação prática representa um ponto de atenção na gestão da ciência de dados. Estudos apontam que o processo de execução também é determinante para o sucesso, exigindo que cientistas de dados estejam conscientes da importância de estruturas processuais adequadas (Saltz; Hotz, 2020; Saltz et al., 2018). A falta de processos colaborativos eficazes foi apontada como uma das principais causas de ineficiências e falhas, indicando que a abordagem multidisciplinar da ciência de dados deve abranger tanto o domínio técnico quanto o gerencial.

Embora a Gestão de Projetos tradicional ofereça metodologias voltadas ao controle de escopo, prazos e recursos, muitas dessas abordagens se mostram inadequadas frente à natureza dinâmica e exploratória da ciência de dados. De acordo com Lahiri e Saltz (2023), mesmo entre organizações que utilizam *frameworks* de gestão, 62% adotaram métodos ágeis sem seguir rigorosamente o *Scrum Guide*, o que evidencia a necessidade de adaptações. Assim, torna-se essencial revisar e atualizar as metodologias de gestão aplicadas à ciência de dados, adequando-as ao atual cenário tecnológico (Martinez; Viles; Olaizola, 2021a). Pesquisas recentes reforçam a importância de *frameworks* bem definidos para aprimorar a colaboração, a comunicação e a eficiência das equipes (Saltz et al., 2018). Diante disso, este trabalho tem como objetivo investigar os principais desafios enfrentados em projetos de ciência de dados e analisar como metodologias de gestão podem ser aplicadas para superá-los, propondo melhorias que contribuam para elevar a taxa de sucesso desses projetos. A proposta baseia-se na revisão das metodologias existentes e na condução de uma pesquisa exploratória com profissionais atuantes na área.

FUNDAMENTAÇÃO TEÓRICA

Esta seção apresenta os principais resultados da revisão bibliográfica realizada, que teve como objetivo ampliar e aprofundar o conhecimento sobre os desafios e metodologias aplicáveis à gestão de projetos de ciência de dados, formando a base conceitual para a pesquisa exploratória subsequente. A revisão fundamentou-se especialmente nos estudos de Martinez, Viles e Olaizola (2021a, 2021b), que destacam três áreas essenciais a serem contempladas por uma metodologia eficaz: gestão de equipes, organização do projeto e gestão de dados e informações. Entre os principais desafios identificados na literatura estão a falta de coordenação entre equipes, dificuldades na comunicação com *stakeholders*, dependência excessiva de poucos especialistas e definição inadequada de escopo e objetivos. Para superá-los, sugere-se a adoção de metodologias ágeis, liderança transformacional e uma governança robusta (Saltz, 2015; Lahiri; Saltz, 2023; Hotz, 2024). Em relação à gestão de dados e informações, abordaram-se questões relacionadas à qualidade dos dados, segurança e integração de sistemas legados com novas tecnologias, enfatizando-se boas práticas em gestão do conhecimento, governança e compliance (Kühn et al., 2018; Passi; Jackson, 2018). Ademais, foram analisadas metodologias gerais como *Scrum* e *Kanban*, bem como metodologias específicas adaptadas para projetos de ciência de dados, destacando-se *Cross Industry Standard Process for Data Mining* (CRISP-DM), *DataOps*, *Knowledge Discovery in Databases* (KDD), *Agile Data Science* e *Team Data Science Process* (TDSP) (Chapman et al., 2000; Atwal, 2020; Journey, 2017).

Gestão de Equipes

A gestão de equipes em projetos de ciência de dados apresenta desafios significativos, incluindo falta de coordenação, dificuldades na colaboração interdepartamental, comunicação inadequada com *stakeholders*, dependência excessiva de poucos especialistas e obstáculos para formar equipes diversas (Saltz et al., 2018; Martinez; Viles; Olaizola, 2021a; Lahiri; Saltz, 2023). A ausência de coordenação é crítica, já que papéis e responsabilidades mal definidos geram retrabalho e confusão, o que pode ser mitigado pela adoção de estruturas organizacionais claras, alinhadas aos objetivos do projeto (Saltz, 2015; Martinez; Viles; Olaizola, 2021b). Além disso, investir na capacitação contínua dos profissionais e distribuir responsabilidades de forma equilibrada ajuda a reduzir a dependência de poucos especialistas, melhorando a eficiência operacional das equipes (Martinez; Viles; Olaizola, 2021a; Hotz, 2024).

A colaboração interdepartamental pode ser fortalecida por meio da criação de equipes multidisciplinares e pelo uso de metodologias ágeis, como *Scrum* e *Kanban*, associadas à liderança transformacional, que permitem maior adaptabilidade e inovação contínua (Lahiri; Saltz, 2023; Lauer, 2021). A comunicação estruturada, apoiada por relatórios padronizados e ferramentas de visualização, também é fundamental para garantir transparência e alinhamento constante com stakeholders (PMBOK, 2021; Passi; Jackson, 2018; Kühn et al., 2018). Por fim, políticas inclusivas que valorizem a diversidade nas equipes são essenciais para superar barreiras culturais e estruturais, estimulando a criatividade e soluções inovadoras (Stelmaszak; Kline, 2023; Espinosa; Armour, 2016).

Gestão do Projeto

A gestão de projetos de ciência de dados enfrenta desafios relacionados à falta de organização nos processos de trabalho, objetivos pouco claros, previsões imprecisas de tempo para execução das tarefas, foco excessivo em aspectos técnicos e baixa utilização prática dos resultados gerados (Saltz et al., 2018; Martinez; Viles; Olaizola, 2021a; Lahiri; Saltz, 2023). Esses obstáculos podem ser mitigados pela aplicação estruturada de metodologias específicas e ferramentas de gestão de projetos. Desde a fase inicial, é essencial alinhar expectativas entre equipes técnicas e áreas de negócio, definindo claramente escopo e responsabilidades, além de adotar comunicação estruturada, visando reduzir ambiguidades e aumentar a eficácia das entregas (Saltz, 2018; Martinez; Viles; Olaizola, 2021a). Ferramentas como a Estrutura Analítica do Projeto (EAP) e a matriz de riscos são úteis para organizar tarefas, prever desafios e aumentar a clareza e previsibilidade dos objetivos e prazos (Haertel et al., 2024; Lahiri; Saltz, 2023).

Simultaneamente, o foco em inovação tecnológica, sustentado por abordagens ágeis e pela experimentação contínua, possibilita respostas rápidas às mudanças e promove melhorias contínuas nos processos (Martinez; Viles; Olaizola, 2021a; Lahiri; Saltz, 2023). Além disso, a questão ética e a governança robusta são fundamentais para garantir o uso responsável dos dados, especialmente considerando regulamentações como a LGPD (Lei Geral de Proteção de Dados Pessoais), que exigem práticas rigorosas de segurança e privacidade para mitigar riscos legais e éticos, além de aumentar a confiança dos stakeholders (Parthasarathy et al., 2024; Saltz, 2015). Portanto, abordagens estruturadas, aliadas a metodologias ágeis, inovação contínua e governança eficiente, são fundamentais para enfrentar os desafios da gestão de projetos na área de ciência de dados, contribuindo para melhores resultados técnicos e de negócio.

Gestão de Dados e Informações

A gestão de dados em projetos de ciência de dados envolve desafios como baixa qualidade das informações, validação insuficiente, problemas de segurança e privacidade, dificuldades na integração de sistemas legados com novas tecnologias e perda de conhecimento ao longo dos projetos (Saltz et al., 2018; Martinez; Viles; Olaizola, 2021a; Lahiri; Saltz, 2023). Dados ausentes, redundantes ou inadequados comprometem diretamente a eficácia de modelos analíticos e preditivos. A falta de validação robusta dos dados e modelos, por sua vez, resulta em soluções pouco confiáveis para ambientes produtivos (Saltz, 2015; Martinez; Viles; Olaizola, 2021a). A segurança das informações, sobretudo em setores sensíveis como saúde e finanças, exige a adoção de *frameworks* de governança robustos para evitar violações legais e preservar a confiança dos *stakeholders* (Martinez; Viles; Olaizola, 2021b).

Nesse contexto, práticas de MLOps (*Machine Learning Operations*) têm ganhado destaque por integrarem governança e automação no ciclo de vida dos modelos. O MLOps viabiliza a operacionalização, o monitoramento e o retreinamento de algoritmos, assegurando sua conformidade com regulamentações e padrões de mercado (Treveil; Heidmann, 2020). A integração de sistemas antigos com novas infraestruturas também representa um desafio técnico relevante, demandando investimento em tecnologia e profissionais qualificados (Ferraris et al., 2019). Além disso, a rotatividade elevada de cientistas de dados acentua a perda de conhecimento acumulado. Para mitigar esse problema, é fundamental implementar práticas de gestão do conhecimento, como repositórios centralizados e sistemas de memória transacional, que favoreçam a preservação e disseminação de aprendizados (Argote; McEvily; Reagans, 2003; Liang; Moreland; Argote, 1995).

A comunicação entre equipes técnicas e *stakeholders* também se mostrou essencial. A adoção de práticas ágeis, como reuniões frequentes e ciclos contínuos de *feedback*, contribuiu para o alinhamento de expectativas, a redução de retrabalho e o ajuste rápido a mudanças nos requisitos (Holzmann; Spiegler, 2011). Em paralelo, a aplicação de *frameworks* de governança permitiu o monitoramento constante dos processos e o cumprimento de normas regulatórias (PMI, 2008). Aliadas às práticas de MLOps, essas estratégias elevaram a eficiência operacional e a qualidade das entregas, contribuindo para um ciclo sustentável de inovação e melhoria contínua nos projetos de ciência de dados (Martinez; Viles; Olaizola, 2021a; Treveil; Heidmann, 2020).

Metodologias de Gestão de Projetos Aplicadas a Projetos de Ciência de dados

Diante da natureza exploratória e incerta dos projetos de ciência de dados, constata-se que abordagens flexíveis são mais adequadas para garantir a adaptação contínua às mudanças de requisitos. Nesse contexto, metodologias ágeis e híbridas destacam-se por promoverem flexibilidade, colaboração e entregas incrementais, facilitando o gerenciamento da complexidade (Schwaber; Sutherland, 2017; Saltz et al., 2018). A combinação de métodos tradicionais e ágeis, como no *Agile-Stage-Gate*, tem se mostrado eficaz ao equilibrar inovação com controle de prazos e custos, especialmente em projetos que envolvem grandes volumes de dados e múltiplos *stakeholders* (Cooper; Sommer, 2016; Boehm; Turner, 2005).

Entre as metodologias mais adotadas, destacam-se o *Scrum* e o *Kanban*, por sua capacidade de estruturar o trabalho e facilitar a colaboração entre equipes multidisciplinares (Lahiri; Saltz, 2023; Martinez; Viles; Olaizola, 2021a). O *Scrum* organiza as atividades em ciclos curtos e iterativos chamados *sprints*, com entregas incrementais e revisão contínua do progresso, permitindo ajustes rápidos aos objetivos do projeto (Schwaber; Sutherland, 2017). Sua estrutura de papéis bem definidos — *Product Owner*, *Scrum Master* e equipe de desenvolvimento — e reuniões frequentes, como o *Daily Scrum*, promovem comunicação eficaz e engajamento das partes envolvidas (Baijens; Helms; Iren, 2020).

No entanto, o *Scrum* apresenta limitações em tarefas exploratórias, como a preparação de dados, que demandam tempo e não se ajustam bem a ciclos curtos. A gestão da imprevisibilidade típica da ciência de dados também pode ser desafiadora, gerando sobrecarga de trabalho (Baijens; Helms; Iren, 2020). Já o *Kanban* se destaca por sua simplicidade e visualização do fluxo de trabalho em tempo real, promovendo transparência e controle da carga de trabalho por meio da limitação de tarefas em progresso (Brechtner, 2015). Essa abordagem favorece a entrega contínua de valor, sendo útil em contextos nos quais a complexidade dos dados exige ajustes incrementais frequentes.

Apesar das vantagens, o *Kanban* também possui limitações, como a ausência de papéis formais, o que pode dificultar a coordenação em equipes grandes ou compostas por diferentes especialidades (Saltz et al., 2018; Lahiri; Saltz, 2023). Além disso, por ser mais genérico, o *Kanban* não oferece diretrizes específicas para a gestão de dados, sendo muitas vezes necessário combiná-lo com outras metodologias voltadas especificamente à ciência de dados. Assim, observa-se que, embora amplamente utilizadas, as metodologias ágeis requerem adaptações para atender às particularidades desse tipo de projeto.

Metodologias de Gestão de Projetos Específicas para Projetos de Ciência de dados

Nos projetos de ciência de dados, observou-se que o uso de metodologias específicas tem desempenhado um papel essencial na organização, execução e sucesso das iniciativas. Dentre as principais, destacam-se CRISP-DM, DataOps, KDD, *Agile Data Science* e TDSP, cada uma com abordagens estruturadas para atender às diferentes etapas e desafios desses projetos. Enquanto o CRISP-DM é reconhecido por sua estrutura em fases bem definidas, o DataOps promove integração ágil entre equipes por meio de automação e entrega contínua. O KDD oferece uma estrutura iterativa para descoberta de conhecimento, o *Agile Data Science* foca na entrega incremental com ciclos curtos de *feedback* e o TDSP, por sua vez, propõe um fluxo padronizado de desenvolvimento colaborativo e reproduzível, especialmente útil em ambientes corporativos mais complexos.

A metodologia CRISP-DM é estruturada em seis fases, entendimento do negócio, entendimento dos dados, preparação dos dados, modelagem, avaliação e implantação, tornou-se uma das mais utilizadas no setor por sua clareza e aplicabilidade em diferentes contextos (Chapman et al., 2000; Wirth; Hipp, 2000). Seu diferencial está na capacidade de alinhar os objetivos técnicos aos requisitos de negócio e permitir iteração contínua ao longo do ciclo do projeto. No entanto, apresenta limitações em projetos que exigem maior flexibilidade ou lidam com grandes volumes de dados e técnicas avançadas de IA, além de não abordar explicitamente a gestão de equipes e comunicação entre *stakeholders* (Martínez-Plumed et al., 2021).

A metodologia DataOps, por sua vez, combina práticas de DevOps, *Lean* e *Agile* para automatizar o fluxo de dados e acelerar a entrega de valor nos projetos de ciência de dados (Atwal, 2020). Com foco em automação, colaboração e integração contínua, o DataOps promoveu ganhos de eficiência, especialmente por meio de ferramentas como *Apache Airflow* e *Jenkins*. No entanto, sua adoção requer mudanças organizacionais profundas, como o rompimento de silos entre equipes e a adaptação a novas tecnologias, o que representa um desafio relevante em contextos com estruturas organizacionais rígidas ou baixa maturidade em gestão de dados (Mainali et al., 2021).

A metodologia KDD segue uma abordagem sistemática, passando por etapas como seleção, pré-processamento, transformação, mineração e interpretação dos dados (Frawley; Piatetsky-Shapiro; Matheus, 1992). Seu ponto forte está na robustez da preparação dos dados e na possibilidade de revisão iterativa durante o processo, permitindo ajustes conforme surgem novas informações. Contudo, o KDD demanda tempo e recursos significativos, especialmente nas etapas iniciais, o que pode dificultar sua aplicação em projetos com prazos reduzidos ou alta complexidade de integração entre fontes de dados heterogêneas.

A proposta da metodologia *Agile Data Science*, conforme Journey (2017), é adaptar os princípios ágeis ao desenvolvimento de produtos de dados com entregas rápidas e ciclos curtos de *feedback*. Essa abordagem favorece a experimentação e o alinhamento contínuo com as

demandas do negócio, promovendo forte interação entre equipes multidisciplinares e *stakeholders*. Ainda que flexível e ágil, essa metodologia pode enfrentar limitações em projetos que exigem maior rigor técnico ou que envolvam equipes com especializações distintas, devido à necessidade de conhecimentos transversais por parte dos integrantes.

Já o TDSP, desenvolvido pela *Microsoft*, estrutura os projetos em fases bem definidas, com entregáveis específicos em cada etapa, como aquisição de dados, modelagem e implantação (*Microsoft*, 2017). Seu diferencial está na promoção de reprodutibilidade e colaboração, com apoio de ferramentas como *Git* e *Azure* para versionamento e gerenciamento de modelos. A metodologia se mostrou eficaz em ambientes corporativos, mas pode ser considerada rígida para contextos que exigem maior experimentação e liberdade metodológica, o que limita sua adoção em projetos altamente inovadores ou exploratórios.

Em síntese, cada metodologia analisada apresenta contribuições relevantes para os projetos de ciência de dados, variando conforme o grau de estrutura, colaboração, flexibilidade e integração exigidos. A escolha adequada depende do contexto do projeto, maturidade da equipe e dos objetivos estratégicos da organização. A integração de práticas dessas metodologias — seja adotando estruturas mais rígidas como o CRISP-DM ou mais ágeis como o DataOps — pode representar uma estratégia eficiente para equilibrar controle, inovação e entrega de valor contínuo.

METODOLOGIA

Com base na fundamentação teórica apresentada na seção anterior, foi elaborado um questionário para a pesquisa exploratória, com o objetivo de investigar as três áreas principais, gestão de equipes, organização do projeto e gestão de dados e informações, além das metodologias utilizadas em projetos de ciência de dados. O questionário foi construído com base nos objetivos e metodologias definidos no projeto de pesquisa, além de referências de quatro artigos, incluindo *Data Science Methodologies: Current Challenges and Future Approaches* (Martinez; Viles; Olaizola, 2021a) e *Evaluating Data Science Project Agility by Exploring Process Frameworks Used by Data Science Teams* (Lahiri; Saltz, 2023).

O questionário foi dividido em quatro partes principais. A primeira parte coletou o perfil dos respondentes, solicitando informações como cargo, nível de senioridade, tempo de experiência e o tamanho das equipes. A segunda parte abordou os desafios enfrentados em projetos de ciência de dados, com base nas categorias organizacionais, de equipe e relacionadas a dados e informações. A terceira parte investigou as metodologias de gestão de projetos utilizadas, enquanto a quarta focou nas metodologias específicas para ciência de dados, como CRISP-DM, KDD e TDSP.

A coleta de dados foi realizada por meio de um formulário criado no *Google Forms*, utilizado para aplicar o questionário elaborado. Selecionaram-se, via *LinkedIn*, 44 profissionais atuantes em projetos de ciência de dados, com diferentes níveis de senioridade e setores. O critério de seleção buscou uma amostra diversificada e de contatos cuja experiência já era conhecida, assegurando a qualidade das respostas.

Inicialmente, foi realizado um contato direto com os possíveis participantes para explicar o objetivo da pesquisa e verificar a disponibilidade para responder ao questionário de forma voluntária. Aos que demonstraram interesse, enviou-se o *link* para o formulário. A pesquisa, realizada em setembro de 2024, foi conduzida de forma anônima, e as respostas foram coletadas ao longo de duas semanas, resultando em 22 respostas, uma taxa de adesão de 50%.

Essa abordagem garantiu a representatividade de diferentes perfis em projetos de ciência de dados.

A análise das 22 respostas coletadas foi realizada utilizando a linguagem de programação *Python*, com as bibliotecas *pandas*, *numpy* e *scipy* para manipulação e análise dos dados. O questionário, composto por 17 questões e 22 respostas, foi transformado em um conjunto de dados com 22 observações e 53 variáveis. Cada linha representou um respondente, e cada coluna, uma variável. Para questões com múltiplas respostas, cada opção foi convertida em uma coluna separada, recebendo marcação binária (sim ou não). O conjunto de dados final contou com 52 variáveis categóricas objetivas e uma variável aberta opcional do tipo texto.

Foram calculadas as frequências e percentuais das respostas objetivas para obter uma visão geral da distribuição. Também foi realizada uma análise de correlação entre respostas por meio de tabelas cruzadas. A questão aberta foi analisada com a identificação de temas recorrentes, que foram comparados com os desafios citados na literatura. A partir dessas análises, foram obtidos os resultados e discussões apresentados a seguir.

ANÁLISE DOS RESULTADOS

A pesquisa exploratória contou com 22 respondentes, majoritariamente Cientistas de Dados (63,64%), refletindo a predominância desse papel nos projetos da área. Outros papéis como Gestores(as) em Ciência de Dados (13,64%), Analistas de Dados (9,09%), Engenheiros(as) de *Machine Learning*, Coordenadores(as) e Analistas de Negócios (cada um com 4,55%) reforçam a diversidade de funções envolvidas. Em relação à senioridade, há equilíbrio entre Especialistas e Seniores (31,82% cada), indicando alta qualificação. A distribuição dos cargos e respectivos níveis de senioridade está ilustrada na Figura 1. Quanto ao tempo de atuação na empresa atual, 40,91% atuavam há menos de um ano, 40,91% possuíam entre um e três anos de experiência, e 18,18% relataram ter entre quatro e seis anos. Nenhum respondente indicou possuir mais de sete anos de experiência, o que sugere uma possível alta rotatividade, reflexo da crescente demanda por profissionais e da falta de processos claros dentro das empresas, gerando incertezas sobre a continuidade dos projetos (Aho et al., 2020; Saltz, 2015). A literatura também apontou que muitos projetos de ciência de dados foram realizados de forma ad hoc, sem o uso de metodologias robustas, o que contribuiu para essa volatilidade (Aho et al., 2020).

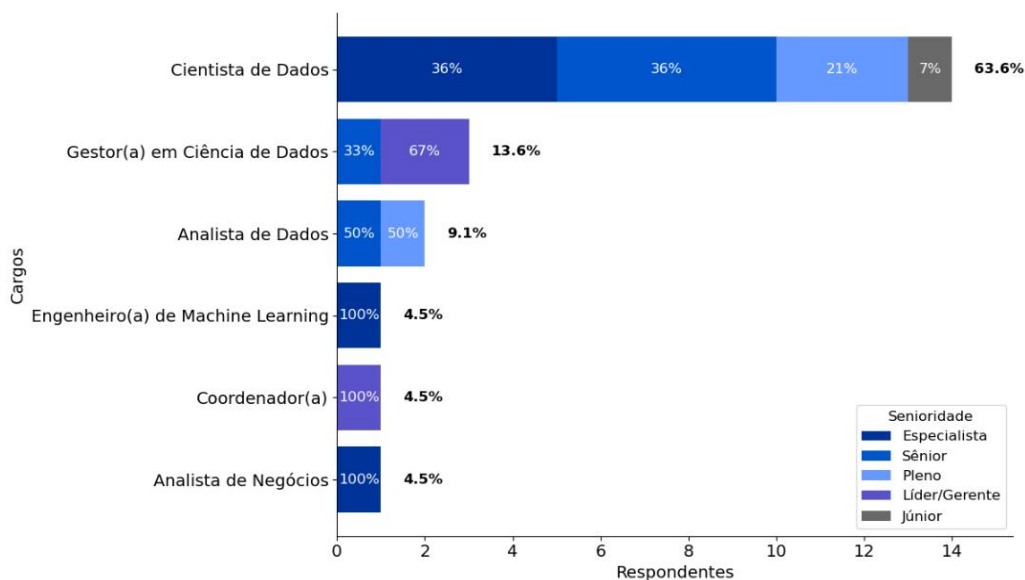


Figura 1. Distribuição dos cargos e níveis de senioridade entre os respondentes da pesquisa (n = 22) Fonte: Resultados originais da pesquisa

A pesquisa indicou que 81,82% dos profissionais trabalhavam em equipes pequenas de duas a três pessoas (Figura 2), refletindo o desafio comum de tratar cientistas de dados como "guerreiros solitários", o que limitou a colaboração em projetos complexos. Estudos apontaram que equipes multidisciplinares maiores, incluindo especialistas do domínio, engenheiros de dados e desenvolvedores de software, foram fundamentais para o sucesso dos projetos conforme avançavam para produção (Aho et al., 2020; Saltz, 2015).

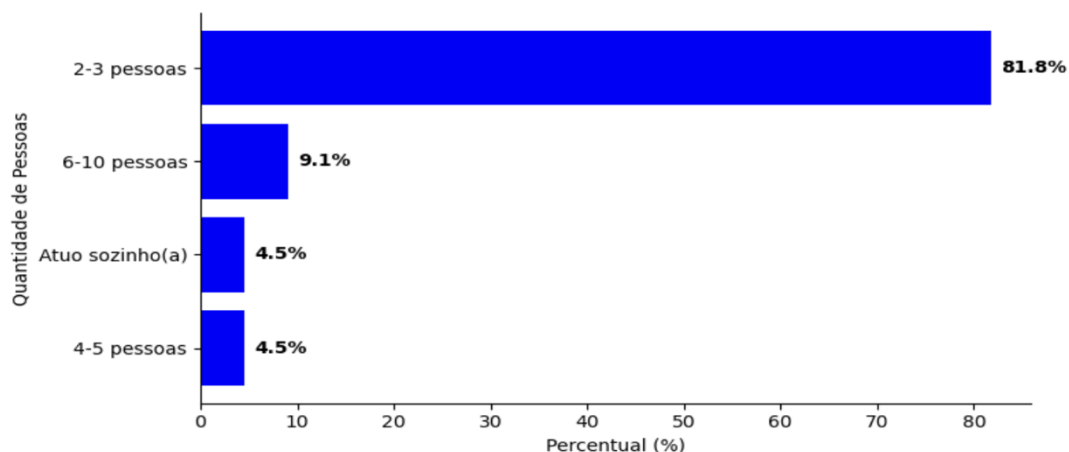


Figura 2. Distribuição da quantidade de pessoas nas equipes de ciência de dados (n = 22) Fonte: Resultados originais da pesquisa

Desafios em Projetos de Ciência de Dados

Quanto aos desafios em projetos de ciência de dados, a pesquisa revelou que 45% dos respondentes classificaram a taxa de sucesso como moderada, enquanto apenas 27% consideraram-na alta. Apesar disso, 68% avaliaram o impacto dos projetos nos resultados da empresa como alto ou muito alto. A correlação fraca (Cramér's V = 0,28) entre a taxa de sucesso e o impacto percebido sugere que, embora os projetos causem impactos relevantes nos negócios, frequentemente não atingem plenamente os objetivos definidos. Esses resultados reforçam desafios como a falta de processos claros de gestão e a integração de equipes

multidisciplinares. Estudos prévios destacam que tais desafios são comuns em projetos de ciência de dados e podem limitar sua eficácia (Saltz, 2015; Lahiri e Saltz, 2023). A relação entre a taxa de sucesso e o impacto é apresentada na Tabela 1.

Tabela 1. Relação entre a taxa de sucesso dos projetos de ciência de dados e seu impacto nos resultados da empresa.

Taxa de Sucesso do Projeto	Impacto do Projeto na Empresa				Total
	Baixo	Moderado	Alto	Muito Alto	
Baixa	0 (0%)	1 (4%)	2 (9%)	0 (0%)	3 (14%)
Moderada	2 (9%)	2 (9%)	5 (23%)	1 (4%)	10 (45%)
Alta	1 (4%)	1 (4%)	2 (9%)	2 (9%)	6 (27%)
Muito Alta	0 (0%)	0 (0%)	1 (4%)	2 (9%)	3 (14%)
Total	3 (14%)	4 (18%)	10 (45%)	5 (23%)	22 (100%)

Fonte: Resultados originais da pesquisa

A volatilidade no mercado de profissionais de dados, combinada à alta demanda e baixa retenção, pode ter prejudicado o desempenho dos projetos, resultando em resultados inconsistentes (Aho et al., 2020). Dos respondentes, 50% relataram que os desafios impactavam frequentemente seu trabalho, 31,82% afirmaram ser afetados de forma intermitente, enquanto 9,09% indicaram raramente serem impactados e outros 9,09% declararam enfrentar esses desafios sempre. Esses dados destacam a presença constante de desafios no cotidiano das equipes, refletindo diretamente nos resultados dos projetos.

Desafios de Equipe

Os principais desafios de equipe identificados foram a dependência excessiva de poucos especialistas (68,2%), o que pode ter sobrecarregado determinados profissionais e atrasado o andamento dos projetos. Esse problema foi amplamente discutido na literatura, com Hotz (2024) sugerindo a capacitação contínua como uma solução para distribuir melhor as responsabilidades. A comunicação insuficiente com stakeholders foi outro desafio importante, apontado por 63,6% dos participantes, em consonância com Lahiri e Saltz (2023), que destacaram a necessidade de práticas de comunicação estruturadas para manter o alinhamento com as partes interessadas.

Além disso, 59,1% dos respondentes indicaram problemas com a estrutura de gestão inadequada e dificuldades de coordenação entre as equipes, reforçando a importância de uma liderança clara e de uma organização bem definida, conforme discutido por Saltz et al. (2018) e Martinez, Viles e Olaizola (2021a). Por fim, a dificuldade em formar equipes diversas foi mencionada por 40,9% dos participantes, o que evidenciou a necessidade de políticas inclusivas e liderança transformacional, conforme defendido por Stelmaszak e Kline (2023), para criar equipes mais equilibradas e inovadoras.

Desafios de Projeto

Os desafios enfrentados na gestão de projetos de ciência de dados, conforme apontado pela revisão bibliográfica, incluíram a falta de organização nos processos, dificuldade em definir objetivos claros e previsões imprecisas para tarefas, conforme ressaltado por Saltz

(2018) e Lahiri e Saltz (2023). Esses problemas foram refletidos nos resultados obtidos na pesquisa exploratória, na qual 59,1% dos participantes indicaram a falta de organização nos processos de trabalho e 68,2% apontaram que os objetivos dos projetos não estavam claros, corroborando o que foi observado na literatura sobre a baixa maturidade e a falta de processos claros em projetos de ciência de dados (Martinez, Viles e Olaizola, 2021a).

Além disso, a dificuldade de prever o tempo necessário para as tarefas, mencionada por 68,2% dos respondentes, evidenciou a necessidade de ferramentas de planejamento mais robustas, como a Estrutura Analítica de Projetos (EAP) e a matriz de riscos, conforme sugerido por Haertel et al. (2024), para melhorar a previsibilidade e a clareza nas metas. Assim, observou-se que os desafios enfrentados no campo estavam amplamente alinhados com as percepções dos profissionais na pesquisa, reforçando a importância de metodologias mais estruturadas e a adoção de práticas ágeis para mitigar esses problemas.

Desafios de Dados e Informação

Os principais desafios relacionados à Gestão de Dados e Informação, identificados na pesquisa exploratória, destacaram a baixa qualidade dos dados como o maior problema, mencionado por 77,3% dos respondentes. Este achado está alinhado com a literatura (Saltz et al., 2018; Martinez, Viles e Olaizola, 2021a; Lahiri e Saltz, 2023), que destaca como dados inconsistentes e incompletos comprometeram a eficácia de modelos preditivos. A perda de conhecimento gerado nos projetos, também citada por 77,3%, foi outro desafio relevante, corroborada por Argote, McEvily e Reagans (2003), que enfatizam a necessidade de repositórios centralizados para evitar a perda de informações importantes devido à rotatividade de pessoal.

A falta de verificação dos dados foi mencionada por 63,6% dos respondentes, novamente em consonância com a literatura, que aponta a ausência de processos robustos de validação como um fator que comprometeu a taxa de sucesso das soluções em produção (Saltz et al., 2018). Além disso, dificuldades na integração de sistemas de dados antigos e novos (50%) são descritas na literatura como um desafio técnico que exige investimentos em infraestrutura (Ferraris et al., 2019). Embora menos citada, a segurança dos dados (22,7%) é considerada crítica na literatura, principalmente em setores regulados, como saúde e finanças (Saltz, 2015). Esses resultados demonstraram que os desafios enfrentados pelas equipes de dados no contexto da pesquisa refletem consistentemente as dificuldades abordadas na literatura sobre a gestão de dados em projetos de ciência de dados.

Metodologias de Gestão de Projetos

A análise comparativa do uso das metodologias ágeis Scrum e Kanban em projetos de ciência de dados, conforme destacado na pesquisa exploratória, revelou que essas abordagens foram amplamente utilizadas para enfrentar a natureza dinâmica desses projetos. Dos 22 participantes da pesquisa, 21 indicaram utilizar alguma metodologia de gestão de projetos (95,45%), e desses, apenas 1 relatou não usar *Scrum*. Esse número reflete a popularidade do *Scrum* devido à sua capacidade de gerenciar entregas incrementais e ciclos curtos de sprints, como destacado por Schwaber e Sutherland (2017) e reforçado por Baijens, Helms e Iren (2020).

Além disso, a pesquisa também revelou que 71,43% dos 21 respondentes que utilizam metodologias de gestão relataram o uso de *Kanban*, um sistema que facilita a visualização contínua do fluxo de trabalho e promove a transparência, como discutido por Brechner (2015).

A pesquisa mostrou que, apesar do uso generalizado de metodologias ágeis, 60% dos respondentes acreditaram que a metodologia utilizada apenas parcialmente ajudava na execução dos projetos, enquanto 35% acreditaram que ela ajudava plenamente, o que indicou que há espaço para melhorias nas práticas de gestão de projetos de ciência de dados. Isso se refletiu na satisfação geral com as metodologias, em que 52,38% dos profissionais se disseram neutros em relação à satisfação, e apenas 38,10% se consideraram satisfeitos. Esses dados sugeriram que, embora as metodologias ágeis fossem amplamente adotadas, ainda eram necessárias adaptações para sua aplicação prática em projetos de ciência de dados, possivelmente devido à natureza exploratória e incerta desses projetos.

Metodologias Específicas para Ciência de Dados

A análise do uso de metodologias específicas para projetos de ciência de dados foi limitada aos 68,18% dos respondentes (15 pessoas) que relataram estar familiarizados com essas abordagens, conforme indicado na pesquisa. Desses, todos relataram familiaridade com CRISP-DM. Isso refletiu o papel consolidado dessa metodologia, amplamente utilizada por sua estrutura em fases bem definidas, permitindo que os projetos de ciência de dados fossem conduzidos de maneira organizada e eficiente, conforme destacado na literatura (Chapman et al., 2000).

Além disso, 10 dos respondentes (66,67%) indicaram familiaridade com a metodologia DataOps, evidenciando sua crescente adoção em projetos que demandam integração ágil e automação de fluxos de dados. A revisão bibliográfica aponta que DataOps integra práticas de DevOps e Lean para otimizar o ciclo de vida dos dados, o que o torna relevante em projetos de grande escala (Atwal, 2020). Já a metodologia KDD, focada na descoberta de conhecimento a partir de grandes volumes de dados, foi mencionada por 8 respondentes (53,33%), reafirmando sua importância em projetos que necessitam de uma análise sistemática e detalhada dos dados (Frawley et al., 1992).

A metodologia *Agile Data Science* foi mencionada por 6 respondentes (40%), destacando-se por sua flexibilidade e pela entrega de valor em ciclos curtos, com feedback contínuo. Essa abordagem ágil é particularmente útil em ambientes onde as demandas de dados mudam rapidamente e a necessidade de adaptação é constante, conforme discutido por Journey (2017). A capacidade de interagir de forma iterativa com as partes interessadas ajuda a garantir que o desenvolvimento permaneça alinhado às metas do negócio, evitando desconexões ao longo do projeto. Por outro lado, o TDSP, mencionado por apenas 2 respondentes (13,33%), fez parte de um grupo de metodologias que tiveram baixa adesão na pesquisa exploratória. Além dele, outras metodologias como Analytics Canvas (20%) e Big Data Management Canvas (20%) também apareceram com menor popularidade. Isso sugere que, embora relevantes, essas abordagens ainda não possuem ampla adoção ou reconhecimento prático entre os profissionais da amostra analisada, possivelmente devido à especificidade ou menor disseminação dessas metodologias em comparação com frameworks mais estabelecidos como CRISP-DM e DataOps.

Em relação ao uso dessas metodologias no ambiente de trabalho, 53,33% dos respondentes relataram utilizá-las regularmente e 40% ocasionalmente, indicando que, embora amplamente conhecidas, sua aplicação prática ainda dependeu do contexto dos projetos. Entre os que utilizaram essas metodologias, 71,43% se declararam satisfeitos com a abordagem adotada, enquanto 21,43% apresentaram uma opinião neutra. Esses resultados reforçam a percepção de que, mesmo com desafios de implementação, essas metodologias especializadas

têm contribuído positivamente para a condução de projetos de ciência de dados, conforme sugerido por Frawley et al. (1992) e Atwal (2020).

Por fim, para os respondentes que não utilizam essas metodologias, 57,14% acreditam que uma metodologia especializada ajudaria a melhorar o processo de desenvolvimento e entrega dos projetos. Esses dados sugerem que há uma percepção positiva quanto aos benefícios potenciais das metodologias específicas, mesmo entre aqueles que ainda não as aplicam diretamente em seu trabalho.

Comentários dos Respondentes sobre Desafios em Projetos de Ciência de Dados

A análise dos comentários na pergunta aberta revelou desafios comuns em projetos de ciência de dados, que também foram apontados na literatura. A rigidez do Scrum, apesar de ser um facilitador, foi criticada por limitar a criatividade e pressionar por entregas rápidas, o que pode prejudicar a exploração profunda de soluções (Comentário 1). Esse problema está alinhado com a revisão de Lahiri e Saltz (2023), que identificaram a dificuldade de aplicar metodologias rígidas em ambientes exploratórios.

Outro desafio frequente foi a falta de entendimento das empresas sobre a ciência de dados, comparando-a ao desenvolvimento de *software* (Comentário 2). Foi destacado que o sucesso das metodologias depende da maturidade da equipe e da sua aplicação adaptada (Comentário 6). Saltz (2015) e Martinez, Viles e Olaizola (2021a) sugeriram que uma comunicação mais estruturada entre as equipes de dados e os stakeholders poderia reduzir essa desconexão e alinhar as expectativas.

A análise comparativa confirmou que, apesar da adoção de metodologias ágeis e especializadas, ainda havia lacunas na adaptação dessas abordagens às especificidades dos projetos de ciência de dados e à maturidade das equipes. Tanto a revisão bibliográfica quanto os dados da pesquisa ressaltaram a necessidade de flexibilidade, comunicação clara e governança robusta para enfrentar os desafios vivenciados pelos profissionais. Portanto, a aplicação adequada dessas metodologias, aliada à maior capacitação das equipes e ao alinhamento com os objetivos de negócio, se mostrou importante para melhorar os resultados e elevar a taxa de sucesso dos projetos de ciência de dados.

CONCLUSÃO

Este trabalho contribuiu para uma melhor compreensão dos desafios enfrentados em projetos de ciência de dados e ofereceu uma análise profunda sobre a aplicabilidade de metodologias de gestão de projetos para lidar com essas dificuldades. A pesquisa demonstrou que, apesar da adoção significativa de metodologias ágeis, ainda há espaço para adaptações que atendam aos desafios específicos relacionados à gestão de equipes e à comunicação entre stakeholders. A necessidade de práticas mais estruturadas de governança e gestão do conhecimento emergiu como um fator essencial para superar os desafios recorrentes identificados.

Portanto, para enfrentar os desafios de gestão em projetos de ciência de dados e aumentar a taxa de sucesso desses projetos, propõe-se a implementação de um modelo híbrido que combine metodologias ágeis com frameworks específicos de ciência de dados, ajustados ao contexto organizacional. Além disso, a capacitação contínua das equipes, o fortalecimento da comunicação com stakeholders e o uso de repositórios centralizados para a preservação do

conhecimento são práticas recomendadas para enfrentar os desafios de gestão de dados e informações, garantindo maior eficiência e sucesso dos projetos.

REFERÊNCIAS BIBLIOGRÁFICAS

Aho, T.; Sievi-Korte, O.; Kilamo, T.; Yaman, S.; Mikkonen, T. Demystifying data science projects: A look on the people and process of data science today. 2020.

Ahmed, B.; Dannhauser, T.; Philip, N. A Lean Design Thinking Methodology (LDTM) for Machine Learning and Modern Data Projects. IEEE Xplore, 2019. Disponível em: <https://www.ieeexplore.ieee.org/document/8837413>. Acesso em: 02 out. 2024.

Argote, L.; McEvily, B.; Reagans, R. Managing Knowledge in Organizations: An Integrative Framework and Review of Emerging Themes. *Management Science*, v. 49, n. 4, p. 571-582, 2003.

Atwal, H. Practical DataOps: Delivering Agile Data Science at Scale. Apress, 2020.

Baijens, J.; Helms, R.; Iren, D. Applying Scrum in Data Science Projects. In: IEEE 22nd Conference on Business Informatics (CBI), 2020.

Boehm, B.; Turner, R. Balancing Agility and Discipline: A Guide for the Perplexed. Addison-Wesley, 2005.

Brechner, E. Agile Project Management with Kanban. Redmond, Washington: Microsoft Press, 2015.

Chapman, P. et al. CRISP-DM 1.0: Step-by-step Data Mining Guide. 1. ed. [S.l.]: CRISP-DM Consortium, 2000.

Cooper, R. G.; Sommer, A. F. The Agile–Stage-Gate Hybrid Model: A Promising New Approach and a New Research Opportunity. *Journal of Product Innovation Management*, v. 33, n. 5, p. 513-526, 2016.

Ferraris, A.; Mazzoleni, A.; Devalle, A.; Couturier, J. Big Data Analytics Capabilities and Knowledge Management: Impact on Firm Performance. *Management Decision*, v. 57, n. 8, p. 1923-1936, 2019.

Espinosa, J. Alberto; Armour, Frank. The big data analytics gold rush: a research framework for coordination and governance. In: 49th Hawaii International Conference on System Sciences. IEEE, 2016. p. 1112-1121. DOI: 10.1109/HICSS.2016.141.

Frawley, W. J.; Piatetsky-Shapiro, G.; Matheus, C. J. Knowledge Discovery in Databases: An Overview. AAAI Press, 1992.

Haertel, C.; Schramm, S.; Pohl, M.; Bosse, S.; Staegemann, D.; Daase, C.; Turowski, K. A Methodology for Constructing Patterns for the Management of Data Science Projects. Proceedings of the 26th International Conference on Enterprise Information Systems (ICEIS 2024), 2024.

Holzmann, V.; Spiegler, I. Developing Risk Breakdown Structure for Information Technology Organizations. *International Journal of Project Management*, v. 29, n. 5, p. 537-546, 2011.

Jurney, R. Agile Data Science 2.0: Building Full-Stack Data Analytics Applications with Spark. O'Reilly Media, 2017.

- Kühn, A. et al. Analytics Canvas – A Framework for the Design and Specification of Data Analytics Projects. *Procedia CIRP*, v. 70, p. 162-167, 2018. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2212827118302962>. Acesso em: 2 out. 2024.
- Lahiri, S.; Saltz, J. Evaluating Data Science Project Agility by Exploring Process Frameworks Used by Data Science Teams. In: *Proceedings of the 56th Hawaii International Conference on System Sciences*, 2023. p. 6538-6547. Disponível em: <https://hdl.handle.net/10125/103424>. Acesso em: 3 out. 2024.
- Lauer, T. *Change Management: Fundamentals and Success Factors*. 3. ed. Berlin: Springer-Verlag, 2021. ISBN 978-3-662-62186-8. Disponível em: <https://doi.org/10.1007/978-3-662-62187-5>. Acesso em: 08 out. 2024.
- Liang, D. W.; Moreland, R.; Argote, Linda. Group versus individual training and group performance: The mediating effects of transactive memory. *Personality and Social Psychology Bulletin*, v. 21, n. 4, p. 384-393, 1995. DOI: 10.1177/0146167295214009.
- Mainali, K.; Ehrlinger, L.; Matskin, M.; Himmelbauer, J. *Discovering DataOps: A Comprehensive Review of Definitions, Use Cases, and Tools*. 2021.
- Martinez, I.; Viles, E.; Olaizola, I. *Data Science Methodologies: Current Challenges and Future Approaches*. *Big Data Research*, v. 24, p. 100183, 2021a.
- Martinez, I.; Viles, E.; Olaizola, I. *A Survey Study of Success Factors in Data Science Projects*. Donostia-San Sebastián: IEEE, 2021b.
- Martínez-Plumed, F. et al. CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering*, v. 33, n. 8, p. 3048-3061, 2021.
- Mathur, S. *Managing Data Science Initiatives as Exploratory Projects – A New Approach to Program Management*. PhD Thesis – University of Technology Sydney, 2024.
- Microsoft. *Team Data Science Process from Microsoft*. GitHub, 2017. Disponível em: <https://github.com/Azure/Microsoft-TDSP>.
- NewVantage Partners. *Big Data and AI Executive Survey*. 2019. Disponível em: <https://www.businesswire.com>. Acesso em: 8 out. 2024.
- Parthasarathy, S.; Panigrahi, P. K.; Subramanian, G. H. A Framework for Managing Ethics in Data Science Projects. *Engineering Reports*, v. 6, n. 3, e12722, 2024. Disponível em: <https://doi.org/10.1002/eng2.12722>. Acesso em: [data de acesso].
- Passi, S.; Jackson, S. J. Trust in Data Science: Collaboration, Translation, and Accountability in Corporate Data Science Projects. *Proceedings of the ACM on Human-Computer Interaction*, v. 2, n. CSCW, Artigo 136, 2018.
- PMI. *A Guide to the Project Management Body of Knowledge (PMBOK Guide)*. Project Management Institute, 2008.
- Saltz, J. S. The Need for New Processes, Methodologies and Tools to Support Big Data Teams and Improve Big Data Project Effectiveness. In: *IEEE International Conference on Big Data*, 2015.

- Saltz, J. S.; Grady, N. W. *The Ambiguity of Data Science Team Roles and the Need for a Data Science Workforce Framework*. Syracuse: Syracuse University, 2017.
- Saltz, J.; Hotz, N. Identifying the Most Common Frameworks Data Science Teams Use to Structure and Coordinate their Projects. In: 2020 IEEE International Conference on Big Data (Big Data), 2020. IEEE.
- Saltz, J.; Hotz, N.; Wild, D.; Stirling, K. Exploring Project Management Methodologies Used Within Data Science Teams. In: AMERICAS CONFERENCE ON INFORMATION SYSTEMS, 24., 2018, New Orleans. Anais [...]. New Orleans: [s.n.], 2018.
- Saltz, J.; Shamshurin, I.; Connors, C. Predicting Data Science Sociotechnical Execution Challenges by Categorizing Data Science Projects. *Journal of the Association for Information Science and Technology*, v. 68, n. 12, p. 2720-2728, 2017.
- Saltz, J.; Shamshurin, I.; Crowston, K. Comparing Data Science Project Management Methodologies via a Controlled Experiment. *Journal of Systems and Software*, 2017.
- Schwaber, K.; Sutherland, J. *The Scrum Guide: The Definitive Guide to Scrum: The Rules of the Game*. 2017. Disponível em: <https://scrumguides.org/scrum-guide.html>. Acesso em: 08 out. 2024.
- Snowden, D. J.; Boone, M. E. A Leader's Framework for Decision Making. *Harvard Business Review*, v. 85, n. 11, p. 68-76, 2007.
- Stelmaszak, M.; Kline, K. Managing Embedded Data Science Teams for Success: How Managers Can Navigate the Advantages and Challenges of Distributed Data Science. *Harvard Data Science Review*, Issue 5.2, Spring 2023.
- Thompson, S. *Managing Machine Learning Projects: From Design to Deployment*. Shelter Island, NY: Manning Publications, 2023.
- TREVEIL, M.; HEIDMANN, L. *What Is MLOps?* 1. ed. Sebastopol, CA: O'Reilly Media, Inc., 2020.
- Wirth, R.; Hipp, J. CRISP-DM: Towards a Standard Process Model for Data Mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, 2000.