

UMA ABORDAGEM BASEADA EM MÉTRICAS HÍBRIDAS PARA AVALIAÇÃO DA PRODUTIVIDADE EM EQUIPES DE DESENVOLVIMENTO DE SOFTWARE EM REGIME REMOTO

Alcino Nuno Barros José¹, Ana Cláudia Garcia Barbosa²

Resumo: A consolidação do trabalho remoto e do desenvolvimento distribuído intensificou o uso de engenharia analítica, dashboards integrados a Git, Jira e pipelines de CI/CD, modelos DORA, modelo SPACE e métricas extraídas de repositórios para acompanhar produtividade em equipes de software. Entretanto, indicadores isolados de atividade, como quantidade de commits, linhas de código ou tarefas concluídas, não representam adequadamente qualidade, colaboração, estabilidade, sustentabilidade e valor organizacional. Este artigo tem como objetivo contrastar modelos híbridos de métricas para avaliação da produtividade em equipes de desenvolvimento de software em ambiente remoto. Neste estudo, métrica híbrida é compreendida como a combinação intencional de indicadores técnicos, operacionais e sociotécnicos, coletados em diferentes fontes e interpretados de forma contextualizada no nível da equipe. A pesquisa é aplicada, exploratória, descritiva e qualitativa, realizada por revisão bibliográfica, extração de indicadores, construção ordinal de 1 a 5, matriz comparativa multicritério e análise de sensibilidade por cenário de decisão. Foram examinados os modelos DORA, SPACE e métricas baseadas em repositórios de código, avaliados segundo precisão, aplicabilidade, escalabilidade e impacto organizacional. Como resultado, verificou-se que o DORA apresenta maior equilíbrio para fluxo e estabilidade de entrega; o SPACE oferece maior abrangência para fatores humanos, colaboração e sustentabilidade; e as métricas de repositórios possuem alta aplicabilidade, mas menor precisão quando usadas isoladamente.

Palavras-chave: produtividade de software; métricas híbridas; trabalho remoto; DORA; SPACE.

¹ Curso de Ciência da Computação, Universidade do Extremo Sul Catarinense (UNESC), alcinobarrosjose@hotmail.com

² Orientadora, Curso de Ciência da Computação, Universidade do Extremo Sul Catarinense (UNESC), agb@unescc.net

ABSTRACT: The consolidation of remote work and distributed development has intensified the use of engineering analytics, dashboards integrated with Git, Jira, and CI/CD pipelines, DORA metrics, the SPACE framework, and repository-extracted metrics to monitor productivity in software teams. However, isolated activity indicators, such as the number of commits, lines of code, or completed tasks, do not adequately represent quality, collaboration, stability, sustainability, and organizational value. This article aims to contrast hybrid metric models for evaluating productivity in software development teams operating in remote environments. In this study, a hybrid metric is understood as the intentional combination of technical, operational, and sociotechnical indicators, collected from different sources and interpreted contextually at the team level. The research is applied, exploratory, descriptive, and qualitative, conducted through a literature review, indicator extraction, ordinal scaling from 1 to 5, multicriteria comparative matrix, and sensitivity analysis based on decision scenarios. The DORA, SPACE, and code repository-based metrics models were examined and evaluated according to accuracy, applicability, scalability, and organizational impact. As a result, it was found that DORA provides the greatest balance for delivery flow and stability; SPACE offers broader coverage of human factors, collaboration, and sustainability; and repository-based metrics show high applicability but lower accuracy when used in isolation.

Keywords: software productivity; hybrid metrics; remote work; DORA; SPACE.

1 INTRODUÇÃO

A consolidação do trabalho remoto no período pós-pandêmico modificou a forma como as equipes de desenvolvimento de software planejam, executam, comunicam e acompanham suas atividades. Em equipes distribuídas, a presença física deixou de ser um indicador suficiente de acompanhamento, enquanto ferramentas digitais de versionamento, integração contínua, gestão de tarefas e comunicação passaram a registrar grande volume de dados sobre o processo de desenvolvimento.

Apesar dessa disponibilidade de dados, a produtividade em engenharia de software permanece um fenômeno de difícil mensuração. A produção de software envolve resolução de problemas, decisões arquiteturais, revisão de código, testes,

comunicação, aprendizagem, redução de riscos, qualidade e geração de valor. Dessa forma, métricas simples de atividade, como quantidade de commits, volume de código ou horas trabalhadas, podem produzir uma percepção equivocada de desempenho quando não são interpretadas em conjunto com indicadores de qualidade e contexto organizacional.

Estudos sobre teletrabalho apontam que o trabalho remoto pode favorecer autonomia, redução de deslocamento, flexibilidade e melhor conciliação entre vida pessoal e profissional, mas também pode ampliar dificuldades de comunicação, isolamento, sobrecarga e fragilidades de alinhamento (Ferreira; Nicida; Valente, 2020; Leite; Silveira; Lemos, 2020; Fernandes; Machado; Araújo, 2025). Tais fatores demonstram que produtividade remota não é apenas um problema técnico, mas também um problema sociotécnico.

No campo da engenharia de software, métricas têm sido usadas para apoiar estimativas, controle de qualidade, gestão de processos e tomada de decisão. A literatura mostra, entretanto, que grande parte dos indicadores tradicionais está concentrada no produto, no código ou no fluxo de entrega, não contemplando plenamente fatores humanos, colaboração, experiência do desenvolvedor e sustentabilidade do trabalho (Magalhães; Rodrigues, 2016; Barcelos et al., 2021; Canno et al., 2023).

Nesse cenário, os modelos DORA e SPACE ganharam relevância por ampliarem a discussão sobre produtividade. O DORA enfatiza fluxo e estabilidade de entrega, enquanto o SPACE propõe uma visão multidimensional que inclui satisfação, desempenho, atividade, comunicação, colaboração e eficiência. Paralelamente, métricas baseadas em repositórios de código e ferramentas de gestão oferecem alta granularidade e facilidade de coleta, mas podem induzir interpretações distorcidas se forem usadas como medida direta de produtividade individual.

Diante desse contexto, o problema de pesquisa que orienta este artigo é: como contrastar modelos híbridos de métricas para avaliação da produtividade de equipes de desenvolvimento de software em ambiente remoto, considerando precisão, aplicabilidade, escalabilidade e impacto organizacional? A hipótese analítica adotada é que cada modelo contribui para uma dimensão distinta da produtividade, de modo que sua utilidade depende do objetivo de medição e da maturidade da equipe.

O objetivo geral deste trabalho é Contrastar modelos híbridos de métricas para avaliação da produtividade em equipes de desenvolvimento de software em ambiente remoto. Como objetivos específicos busca-se: analisar os modelos DORA, SPACE e métricas baseadas em repositórios de código; definir critérios de comparação baseados em precisão, aplicabilidade, escalabilidade e impacto organizacional; realizar uma análise comparativa entre os modelos identificados; e examinar implicações práticas decorrentes do uso desses modelos em equipes de desenvolvimento remoto.

A justificativa do estudo está relacionada à necessidade de apoiar decisões de gestão em equipes remotas sem reduzir produtividade a velocidade, volume de atividade ou disponibilidade permanente. Do ponto de vista computacional, a avaliação da produtividade depende da integração de dados produzidos por ferramentas como Git, GitHub, GitLab, Jira, Jenkins, sistemas de testes, plataformas de observabilidade e ferramentas de comunicação.

Do ponto de vista organizacional, métricas mal definidas podem gerar incentivos inadequados, como fragmentação artificial de commits, priorização da quantidade em detrimento da qualidade, comparação injusta entre desenvolvedores e decisões gerenciais baseadas em dados incompletos. Assim, a análise comparativa de modelos permite compreender quais indicadores são mais adequados para cada finalidade de gestão.

A relevância acadêmica do trabalho está em sistematizar modelos frequentemente tratados de forma isolada. A relevância prática está em oferecer critérios de escolha para equipes e gestores que desejam avaliar produtividade de forma mais consistente em ambientes remotos. O estudo, portanto, contribui para a discussão sobre métricas de software, trabalho remoto e engenharia analítica aplicada à gestão de equipes.

A estrutura deste trabalho está dividida em: Seção 1, a introdução, que contextualiza o problema da avaliação da produtividade em equipes de desenvolvimento de software em ambiente remoto, apresenta a questão de pesquisa, define os objetivos e justifica a relevância do estudo; Seção 2, que apresenta os trabalhos correlatos, discutindo estudos sobre teletrabalho, produtividade, métricas de software, modelos DORA e SPACE, métricas baseadas em repositórios e dashboards de produtividade; Seção 3, que descreve os materiais e métodos, apresentando a caracterização da pesquisa, os critérios de seleção dos

materiais, os modelos analisados e o item multicritério utilizada na comparação; Seção 4, que apresenta os resultados e a discussão, incluindo a análise dos modelos DORA, SPACE e métricas de repositórios, a matriz comparativa, a análise de sensibilidade, as implicações práticas da adoção desses modelos em equipes remotas. Por fim, a Seção 5 apresenta a conclusão, retomando os principais achados, as limitações da pesquisa e as possibilidades de trabalhos futuros.

2 TRABALHOS CORRELATOS

Para o desenvolvimento deste trabalho, foram realizadas pesquisas sobre estudos relacionados à produtividade em trabalho remoto, métricas de software, modelos de avaliação de desempenho e dashboards de produtividade. A análise desses materiais permitiu compreender abordagens já utilizadas na área, identificar suas contribuições e delimitar a contribuição específica desta pesquisa, voltada ao contraste entre DORA, SPACE e métricas baseadas em repositórios de código.

2.1 DESMISTIFICANDO A PRODUTIVIDADE E OS EFEITOS DO HOME OFFICE INTEGRAL DURANTE A PANDEMIA

O estudo de Ferreira, Nicida e Valente (2020) discute a produtividade em home office durante o período da pandemia, destacando fatores que influenciam o desempenho em ambientes remotos. Embora o trabalho não seja voltado especificamente para equipes de desenvolvimento de software, ele contribui para compreender que a produtividade remota não depende apenas da realização de tarefas, mas também de condições como infraestrutura, rotina, ambiente familiar, disciplina individual e adaptação ao trabalho fora do espaço físico da organização.

A principal contribuição desse estudo para a presente pesquisa está em demonstrar que a produtividade no regime remoto não pode ser avaliada apenas por presença, disponibilidade ou volume de atividade. Essa visão reforça a necessidade de considerar fatores contextuais na avaliação de desempenho. Diferente dessa abordagem mais ampla sobre home office, o presente trabalho direciona a discussão para equipes de desenvolvimento de software, relacionando produtividade remota a métricas técnicas, operacionais e humanas.

2.2 IMPACTOS DO TRABALHO REMOTO SOBRE A PRODUTIVIDADE NO PÓS-COVID-19: ESTUDO DE CASO EM UMA EMPRESA DE CONSULTORIA POLÍTICA

O estudo de Fernandes, Machado e Araújo (2025) analisa os impactos do trabalho remoto sobre a produtividade em uma empresa de consultoria política no período pós-Covid-19. A pesquisa identifica percepções positivas relacionadas à produtividade, mas também aponta fragilidades associadas à comunicação, ao isolamento e à falta de padronização na gestão de projetos.

A contribuição desse estudo está em demonstrar que o trabalho remoto pode favorecer produtividade em determinados contextos, mas também exige mecanismos adequados de gestão, comunicação e acompanhamento. Embora o estudo não trate especificamente de desenvolvimento de software, ele reforça que produtividade remota deve ser analisada de forma mais ampla, considerando não apenas entregas, mas também condições de execução do trabalho. Diferente dessa abordagem organizacional, o presente trabalho aplica essa discussão ao contexto de engenharia de software, utilizando modelos como DORA, SPACE e métricas de repositórios.

2.3 CONSTRUÇÃO DO MODELO LÓGICO DO PROGRAMA DE GESTÃO E DESEMPENHO NA MODALIDADE DE TELETRABALHO: UMA PROPOSTA DE AVALIAÇÃO NO ÂMBITO DO FNDE

A dissertação de Pereira (2024) discute a construção de um modelo lógico para avaliação do Programa de Gestão e Desempenho na modalidade de teletrabalho, com foco no contexto do Fundo Nacional de Desenvolvimento da Educação (FNDE). O estudo contribui ao destacar a importância de instrumentos estruturados para avaliar entregas, resultados e efeitos do teletrabalho, substituindo gradualmente formas tradicionais de controle baseadas na presença física.

Embora o contexto da pesquisa esteja relacionado à gestão pública, sua contribuição é relevante para este trabalho porque evidencia a necessidade de avaliar produtividade por meio de resultados e entregas, e não apenas por disponibilidade ou controle de jornada. Diferente da proposta de Pereira, que se concentra em um programa institucional de gestão e desempenho, o presente

trabalho aplica a lógica de avaliação por resultados ao desenvolvimento de software, considerando dados produzidos por ferramentas como Git, GitHub, Jira, pipelines e repositórios de código.

2.4 AS MÉTRICAS DE SOFTWARE EM FAVOR DO DESENVOLVEDOR COMO FERRAMENTAS PARA A PRODUTIVIDADE E A QUALIDADE EM SOLUÇÕES EMBARCADAS

O trabalho de Canno et al. (2023) discute o uso de métricas de software como ferramentas para apoiar produtividade e qualidade, especialmente no desenvolvimento de soluções embarcadas. O estudo reforça que produtividade e qualidade não devem ser tratadas como dimensões separadas ou concorrentes, pois uma entrega rápida pode perder valor caso gere defeitos, retrabalho ou instabilidade.

A contribuição desse estudo está em destacar que métricas devem auxiliar o desenvolvedor no processo de desenvolvimento, e não apenas servir como instrumentos de cobrança. Essa perspectiva aproxima-se da proposta deste trabalho, que defende o uso de métricas como apoio à melhoria contínua. No entanto, enquanto Canno et al. concentram-se em soluções embarcadas, o presente estudo analisa modelos híbridos aplicáveis a equipes de desenvolvimento remoto, integrando DORA, SPACE e métricas baseadas em repositórios.

2.5 THE SPACE OF AI: REAL-WORLD LESSONS ON AI'S IMPACT ON DEVELOPERS

O estudo de Houck et al. (2025) discute o impacto de ferramentas de inteligência artificial no trabalho de desenvolvedores a partir da perspectiva do modelo SPACE. O trabalho reforça que ganhos de produtividade não devem ser avaliados apenas pela velocidade de codificação, mas também por fatores como experiência do desenvolvedor, colaboração, redução de tarefas repetitivas e qualidade do fluxo de trabalho.

A principal contribuição deste estudo está em atualizar a discussão sobre produtividade de desenvolvedores em um cenário de transformação tecnológica. Mesmo com o apoio de ferramentas de inteligência artificial, a produtividade

continua dependendo de fatores humanos, organizacionais e colaborativos. Diferente dessa abordagem voltada ao impacto da IA, o presente trabalho utiliza o SPACE como uma das bases para contrastar modelos de métricas em equipes remotas, sem restringir a análise ao uso de ferramentas específicas.

A partir da análise dos trabalhos correlatos, verifica-se que os estudos existentes contribuem para diferentes dimensões do problema. Os trabalhos sobre teletrabalho explicam os efeitos do trabalho remoto sobre produtividade, comunicação, autonomia e acompanhamento de resultados. Os estudos sobre métricas de software demonstram a importância de indicadores técnicos, mas também seus limites quando usados isoladamente. Os modelos DORA e SPACE ampliam a discussão ao considerar, respectivamente, fluxo de entrega, estabilidade, fatores humanos e colaboração. Já os estudos sobre dashboards e pipelines reforçam a necessidade de confiabilidade e governança dos dados.

Entretanto, essas abordagens tendem a tratar as dimensões de forma separada. A contribuição deste trabalho está em reunir essas perspectivas em uma matriz comparativa, contrastando DORA, SPACE e métricas baseadas em repositórios de código segundo critérios de precisão, aplicabilidade, escalabilidade e impacto organizacional. Dessa forma, busca-se apoiar a seleção crítica de métricas para equipes de desenvolvimento de software em ambiente remoto, evitando interpretações baseadas apenas em volume de atividade ou controle individual.

3 MATERIAIS E MÉTODOS

Esta seção apresenta os procedimentos metodológicos adotados para a realização da pesquisa, descrevendo a natureza do estudo, os materiais utilizados, os critérios de seleção das fontes, as etapas de análise e os critérios empregados para comparar os modelos de métricas investigadas. A organização metodológica busca garantir clareza, rastreabilidade e coerência entre os objetivos propostos e os resultados obtidos, permitindo compreender como os modelos DORA, SPACE e as métricas baseadas em repositórios de código foram analisados no contexto da avaliação da produtividade em equipes de desenvolvimento de software em regime remoto.

3.1 CARACTERIZAÇÃO DA PESQUISA

A pesquisa é classificada como aplicada quanto à finalidade, pois busca organizar critérios úteis à avaliação da produtividade em equipes de desenvolvimento remoto. Quanto aos objetivos, é exploratória e descritiva, por identificar modelos na literatura, caracterizar suas dimensões e compará-los segundo critérios previamente definidos. Quanto à abordagem, é qualitativa, pois a análise depende da interpretação dos modelos e de suas implicações práticas.

O objeto de estudo foi o contraste de modelos híbridos de métricas para avaliação da produtividade de equipes de desenvolvimento de software em ambiente remoto. Não foram analisados indivíduos, empresas ou dados sensíveis. O foco recaiu sobre modelos, indicadores e critérios de decisão presentes na literatura científica e técnico-científica.

3.2 MATERIAIS UTILIZADOS

Os materiais utilizados foram documentos científicos, dissertações, artigos técnicos e estudos disponibilizados para fundamentar o trabalho. O corpus foi composto por publicações sobre teletrabalho, métricas de software, desempenho organizacional, DORA, SPACE, métricas extraídas de repositórios e engenharia de dados para dashboards de produtividade.

Quadro 1 - Corpus e função dos materiais na análise

Grupo de material	Exemplos de fontes	Função na pesquisa
Teletrabalho e produtividade	Ferreira, Nicida e Valente (2020); Leite, Silveira e Lemos (2020); Fernandes, Machado e Araújo (2025); Pereira (2024)	Contextualizar os efeitos do trabalho remoto e os desafios de avaliação de desempenho.
Métricas de software	Magalhães e Rodrigues (2016); Barcelos et al. (2021); Canno et al. (2023)	Identificar limites de métricas tradicionais e relação entre produtividade, qualidade e processo.
Modelos de produtividade	Forsgren, Humble e Kim (2018); Forsgren et al. (2021); Houck et al. (2025)	Definir os modelos DORA e SPACE e suas dimensões de análise.
Engenharia analítica	Jain et al. (2025); Agrawal e Jain (2026)	Discutir confiabilidade de dados, pipelines e rastreabilidade de métricas.

Desempenho organizacional	Silva et al. (2024); Andrade e Souza (2023)	Relacionar indicadores de produtividade a tomada de decisão e impacto organizacional.
---------------------------	---	---

Fonte: Elaborado pelo autor (2026).

O Quadro 1 evidencia que o corpus foi organizado para cobrir tanto a dimensão técnica quanto a dimensão humana da produtividade. Essa organização evita que a análise fique restrita ao código ou ao fluxo de entrega e permite discutir produtividade remota como fenômeno sociotécnico, coerente com a literatura de teletrabalho e com os modelos multidimensionais de produtividade.

3.3 CRITÉRIOS DE SELEÇÃO E EXCLUSÃO

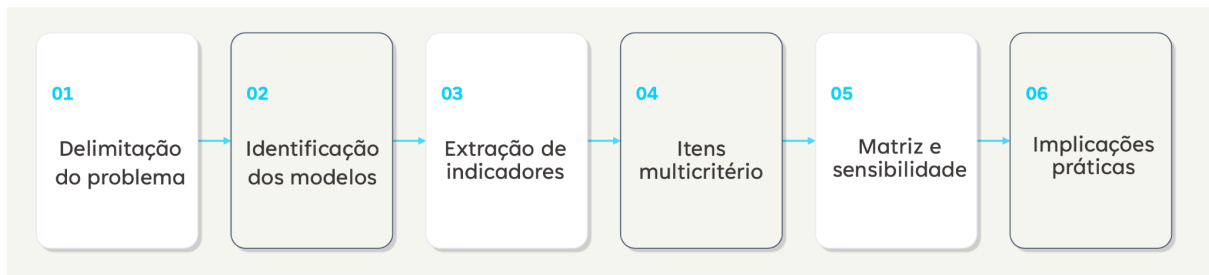
Foram incluídos estudos que tratavam de pelo menos um dos seguintes temas: produtividade em trabalho remoto, métricas de software, modelos DORA, modelo SPACE, métricas extraídas de repositórios de código, avaliação de desempenho ou dashboards de produtividade. Foram priorizados textos com aplicação direta à engenharia de software ou à gestão de equipes tecnológicas.

Foram excluídos materiais com foco exclusivamente jurídico, sem relação com produtividade; textos que discutiam teletrabalho sem vínculo com avaliação de desempenho; documentos repetidos; e fontes que não apresentavam conceitos, indicadores ou critérios úteis à comparação. Como o estudo é qualitativo e conceitual, a seleção buscou pertinência ao problema de pesquisa, e não representatividade estatística.

3.4 PROCEDIMENTOS ESPECÍFICOS

O procedimento metodológico foi organizado em seis etapas cronológicas. A primeira etapa delimitou o problema e os objetivos. A segunda identificou modelos de métricas de produtividade descritos na literatura. A terceira extraiu indicadores, benefícios, limitações e requisitos de aplicação de cada modelo. A quarta definiu critérios de comparação. A quinta atribuiu notas ordinais aos modelos com base na literatura analisada. A sexta sintetizou os resultados e examinou implicações práticas.

Figura 1 - Processo metodológico adotado para o contraste dos modelos



Fonte: Elaborado pelo autor (2026).

A Figura 1 explicita que a pesquisa não parte diretamente da tabela de resultados. Antes da matriz, foram definidos o problema, os modelos e os indicadores que compõem cada abordagem. Item multicritério funciona como etapa intermediária, reduzindo a subjetividade da comparação e permitindo explicar por que cada nota foi atribuída, finalizando com matriz e sensibilidade e implicações práticas. Essa organização também evita que os quadros apareçam como resultados isolados, pois cada artefato apresentado decorre de uma etapa metodológica anterior.

3.5 CRITÉRIOS E ITEM DE COMPARAÇÃO

Os critérios utilizados foram definidos a partir dos objetivos do estudo: precisão, aplicabilidade, escalabilidade e impacto organizacional. Para tornar a análise mais objetiva, cada critério recebeu um item ordinal de 1 a 5. A nota 1 indica baixa aderência ao critério; a nota 3 indica aderência intermediária; e a nota 5 indica aderência elevada. A pontuação não representa medição estatística, mas julgamento analítico fundamentado na literatura.

Quadro 2 - Item dos critérios de comparação

Critério	Descrição	Interpretação da escala ordinal
Precisão	Capacidade de representar produtividade sem reduzi-la a volume de atividade.	1 = baixa validade conceitual; 3 = representa parte do fenômeno; 5 = representa dimensões centrais com baixo risco de distorção.
Aplicabilidade	Facilidade de coleta, interpretação e uso em equipes reais.	1 = difícil de implantar; 3 = requer adaptação moderada; 5 = alta disponibilidade de dados e baixa barreira de uso.

Escalabilidade	Capacidade de aplicação em diferentes equipes, produtos e períodos.	1 = restrito a contextos específicos; 3 = escalável com ajustes; 5 = facilmente automatizável e padronizável.
Impacto organizacional	Potencial de apoiar decisões, melhoria contínua e alinhamento com objetivos de negócio.	1 = baixo apoio decisório; 3 = apoio intermediário; 5 = alto potencial de orientar decisões e melhoria.

Fonte: Elaborado pelo autor (2026).

O Quadro 2 explicita o significado de cada nota e torna a avaliação mais transparente. A validação adotada nesta pesquisa é conceitual e teórica: as notas foram atribuídas por triangulação entre a definição do modelo, seus indicadores típicos, suas limitações descritas na literatura e sua aderência aos critérios. Assim, o item ordinal não elimina totalmente o julgamento do pesquisador, mas reduz arbitrariedade e permite que a avaliação seja auditável e reproduzível por outros estudos.

3.6 ANÁLISE DOS DADOS

A análise dos dados foi realizada por leitura analítica, categorização e síntese comparativa. Inicialmente, os indicadores de cada modelo foram agrupados conforme sua finalidade. Em seguida, os modelos foram avaliados nos quatro critérios definidos. Para ampliar a robustez da comparação, foram calculadas médias simples e realizadas leituras de sensibilidade por cenário de uso, variando a importância relativa dos critérios conforme o objetivo gerencial.

4 RESULTADOS E DISCUSSÃO

Esta seção apresenta os resultados obtidos a partir da análise comparativa dos modelos de métricas selecionados. Inicialmente, são caracterizados os modelos identificados na literatura, com destaque para suas ênfases, indicadores típicos e contribuições para equipes remotas. Em seguida, os modelos DORA, SPACE e as métricas baseadas em repositórios de código são examinados individualmente, considerando suas contribuições e limitações. Por fim, os resultados são sintetizados por meio de uma matriz comparativa multicritério, de uma análise de sensibilidade por cenário de decisão e de uma discussão sobre as implicações

práticas da adoção desses modelos no contexto da produtividade em equipes de desenvolvimento de software em regime remoto.

4.1 MODELOS IDENTIFICADOS

A primeira etapa dos resultados consistiu na identificação e caracterização dos modelos analisados. O Quadro 3 resume a ênfase de cada modelo, seus principais indicadores e sua contribuição para a avaliação da produtividade remota.

Quadro 3 - Modelos de métricas identificados na literatura

Modelo	Ênfase principal	Indicadores típicos	Contribuição para equipes remotas
DORA	Fluxo e estabilidade de entrega de software.	Frequência de implantação; lead time de mudanças; taxa de falhas; tempo de recuperação.	Permite avaliar se a equipe entrega software de forma rápida, estável e recuperável.
SPACE	Produtividade multidimensional do desenvolvedor.	Satisfação; desempenho; atividade; comunicação; colaboração; eficiência; fluxo.	Inclui fatores humanos e colaborativos que são críticos em ambientes distribuídos.
Métricas de repositórios	Atividade técnica e dinâmica do processo de desenvolvimento.	Commits; pull requests; tempo de revisão; issues; cobertura de testes; retrabalho.	Oferece dados granulares e automatizáveis, úteis para diagnóstico operacional.

Fonte: Elaborado pelo autor (2026).

O Quadro 3 demonstra que os modelos respondem a perguntas de gestão diferentes. O DORA está mais associado à capacidade de entregar software com rapidez, estabilidade e recuperação. O SPACE amplia a discussão ao incluir satisfação, colaboração e eficiência percebida, elementos diretamente relacionados aos desafios do trabalho remoto. Já as métricas de repositórios oferecem sinais operacionais úteis para localizar gargalos, mas não explicam sozinhas a qualidade, o valor entregue ou a sustentabilidade do trabalho. Portanto, a contribuição central não está em escolher uma única abordagem, mas em entender quando cada uma deve ser priorizada.

4.2 ANÁLISE DO MODELO DORA

O DORA apresentou desempenho elevado nos critérios de precisão, escalabilidade e impacto organizacional quando o objetivo é avaliar fluxo e estabilidade. Suas métricas conectam velocidade de entrega e confiabilidade operacional, evitando que produtividade seja compreendida apenas como volume de trabalho. Em equipes remotas, esse ponto é relevante porque a entrega frequente e recuperável reduz dependência de controle presencial.

Quanto à aplicabilidade, o modelo é favorecido por equipes que utilizam integração contínua, entrega contínua, controle de versões e sistemas de incidentes. Em ambientes com pipelines maduros, a coleta pode ser automatizada. A limitação é que o DORA não mede diretamente comunicação, satisfação, carga cognitiva ou qualidade da colaboração. Assim, ele é preciso para avaliar desempenho de entrega, mas insuficiente para explicar todos os fatores que influenciam a produtividade.

4.3 ANÁLISE DO MODELO SPACE

O SPACE apresentou maior precisão conceitual por tratar produtividade como fenômeno multidimensional. Ao incluir satisfação, desempenho, atividade, comunicação, colaboração, eficiência e fluxo, o modelo reduz o risco de confundir produtividade com movimento operacional. Em equipes remotas, essa amplitude é especialmente relevante, pois problemas de isolamento, comunicação e sobrecarga podem comprometer resultados mesmo quando os indicadores técnicos parecem positivos.

A principal limitação do SPACE está na aplicabilidade. Algumas dimensões exigem questionários, entrevistas, check-ins, retrospectivas e escalas perceptivas. Esses instrumentos dependem de confiança, maturidade organizacional e cuidado ético. Além disso, a escalabilidade pode ser menor do que no DORA ou nas métricas de repositório, pois nem todos os indicadores são coletados automaticamente.

4.4 ANÁLISE DAS MÉTRICAS BASEADAS EM REPOSITÓRIOS DE CÓDIGO

As métricas baseadas em repositórios de código e ferramentas de gestão apresentaram alta aplicabilidade. Commits, pull requests, issues, tempo de revisão e defeitos reabertos são dados normalmente disponíveis em ambientes de desenvolvimento. Essa disponibilidade favorece dashboards, análises históricas e acompanhamento de gargalos no fluxo de trabalho.

Entretanto, a precisão dessas métricas é limitada quando elas são usadas como proxy direto de produtividade. Um commit pequeno pode corrigir um defeito crítico, enquanto muitas linhas de código podem aumentar complexidade e retrabalho. Da mesma forma, número de pull requests ou issues concluídas pode indicar atividade, mas não necessariamente valor entregue. O maior risco é transformar métricas de rastreamento em instrumentos de comparação individual.

4.5 MATRIZ COMPARATIVA MULTICRITÉRIO

A Tabela 1 apresenta a avaliação ordinal dos três modelos segundo os quatro critérios definidos. As notas foram atribuídas com base na síntese da literatura, considerando a aderência de cada modelo aos critérios apresentados na seção 3.5. A média simples é utilizada apenas como recurso de comparação, sem pretensão de representar medição estatística.

Tabela 1 - Matriz comparativa dos modelos analisados

Modelo	Precisão	Aplicabilidade	Escalabilidade	Impacto organizacional	Média
DORA	4	4	5	4	4,25
SPACE	5	3	3	5	4,00
Repositórios de código	3	5	4	3	3,75

Fonte: Elaborado pelo autor (2026).

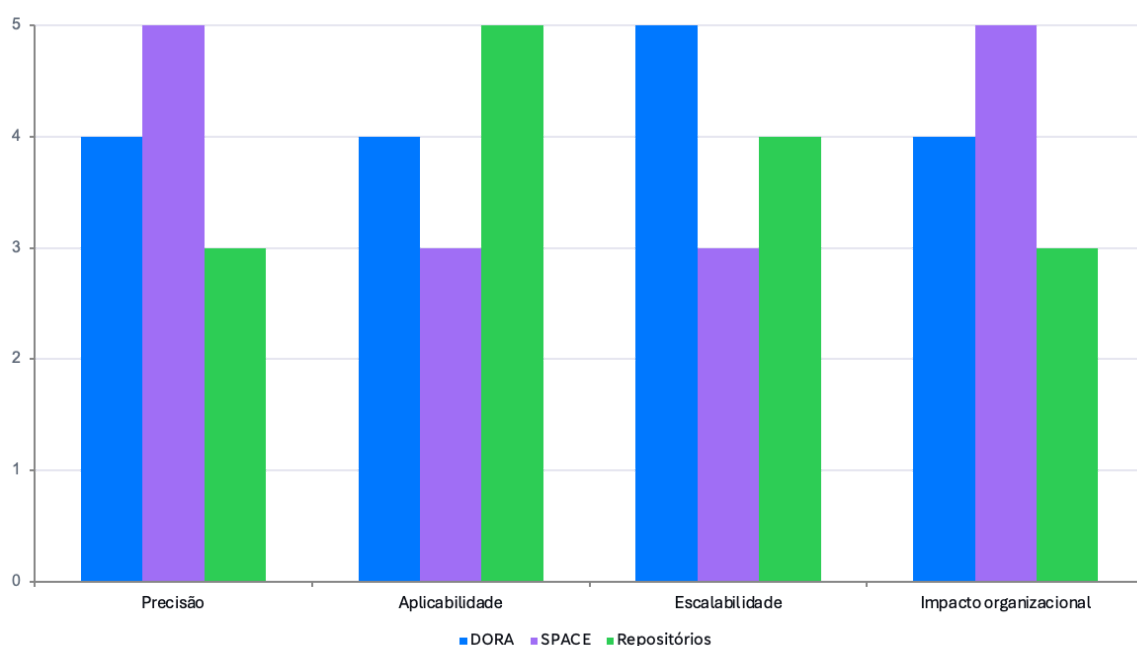
O DORA recebeu nota 4 em precisão porque representa de forma consistente fluxo e estabilidade, embora não cubra diretamente fatores humanos. Recebeu nota 4 em aplicabilidade porque depende de pipelines e registros confiáveis, mas é relativamente simples em equipes com maturidade DevOps. A nota 5 em

escalabilidade decorre da possibilidade de automatização e padronização em diferentes produtos. O impacto organizacional foi avaliado como 4 porque suas métricas apoiam decisões de entrega, risco e confiabilidade operacional.

O SPACE recebeu nota 5 em precisão por representar produtividade como fenômeno multidimensional, contemplando fatores humanos, colaborativos e de fluxo. A aplicabilidade foi avaliada como 3, pois parte dos dados depende de instrumentos perceptivos e processos de confiança. A escalabilidade também recebeu nota 3 porque questionários e práticas qualitativas exigem adaptação por equipe. O impacto organizacional recebeu nota 5 por apoiar decisões sobre sustentabilidade, colaboração e saúde do trabalho remoto.

As métricas de repositórios receberam nota 3 em precisão porque capturam atividade técnica, mas possuem maior risco de distorção quando usadas isoladamente. A aplicabilidade recebeu nota 5 pela ampla disponibilidade dos dados em ferramentas de desenvolvimento. A escalabilidade foi avaliada como 4, pois dashboards podem ser replicados, embora dependam de padronização. O impacto organizacional recebeu nota 3 porque os dados são úteis para diagnóstico, mas insuficientes para decisões estratégicas sem associação a DORA, SPACE ou objetivos de negócio.

Figura 2 - Perfil comparativo dos modelos pelos critérios analisados



Fonte: Elaborado pelo autor (2026).

A Figura 2 apresenta um gráfico de colunas, facilitando a comparação direta entre modelos e critérios. A leitura por colunas evidencia que o SPACE se destaca em precisão e impacto organizacional, enquanto as métricas de repositórios se destacam em aplicabilidade. O DORA apresenta equilíbrio geral, principalmente por combinar boa precisão, alta escalabilidade e impacto sobre decisões de entrega. A figura também mostra que nenhuma abordagem domina todos os critérios, reforçando a tese de complementaridade entre os modelos.

4.6 ANÁLISE DE SENSIBILIDADE POR CENÁRIO DE DECISÃO

Para evitar a falsa conclusão de que um único modelo é superior em qualquer contexto, foi realizada uma análise de sensibilidade por cenário. Os pesos foram definidos de forma analítica para representar diferentes prioridades gerenciais. Quando o foco é estabilidade de entrega, o DORA se destaca. Quando o foco é saúde, colaboração e sustentabilidade, o SPACE torna-se mais indicado. Quando o objetivo é monitoramento operacional rápido, métricas de repositórios ganham relevância.

Tabela 2 - Sensibilidade dos resultados por cenário de uso

Cenário	Pesos aplicados aos critérios	Modelo mais indicado	Interpretação
Equilíbrio geral	Pesos iguais para os quatro critérios.	DORA	Mais equilibrado quando a equipe precisa combinar fluxo, estabilidade e governança.
Saúde e colaboração remota	Maior peso para precisão conceitual e impacto organizacional.	SPACE	Mais adequado para compreender fatores humanos, colaboração e sustentabilidade.
Automação de indicadores	Maior peso para aplicabilidade e escalabilidade.	Repositórios + DORA	Mais adequado quando a equipe precisa iniciar dashboards com dados já disponíveis.
Gestão de melhoria contínua	Maior peso para impacto organizacional e precisão.	DORA + SPACE	Combinação mais robusta para decisões que envolvem fluxo, estabilidade e pessoas.

Fonte: Elaborado pelo autor (2026).

A Tabela 2 aprofunda a discussão da matriz, pois demonstra que o resultado depende do cenário de decisão. Em equilíbrio geral, o DORA permanece levemente

superior. Quando a prioridade é saúde e colaboração, o SPACE passa a liderar porque sua estrutura considera dimensões humanas ausentes em métricas puramente técnicas. Quando a prioridade é automação, os repositórios passam a ter maior pontuação, mas sua leitura deve ser combinada ao DORA para evitar que atividade técnica seja confundida com entrega de valor. Na melhoria contínua, DORA e SPACE praticamente convergem, sugerindo que decisões maduras devem integrar estabilidade operacional e sustentabilidade humana.

4.7 SÍNTESE DOS RESULTADOS COMPARATIVOS

Os resultados indicam que DORA, SPACE e métricas de repositórios não são concorrentes diretos, mas modelos complementares. O DORA responde melhor à pergunta “a equipe entrega software com rapidez e estabilidade?”. O SPACE responde melhor à pergunta “a equipe consegue produzir de forma sustentável, colaborativa e saudável?”. As métricas de repositórios respondem melhor à pergunta “quais sinais operacionais aparecem no fluxo de trabalho técnico?”.

Quadro 4 - Síntese comparativa final

Pergunta de gestão	Modelo mais útil	Cuidados necessários
A equipe entrega com velocidade e estabilidade?	DORA	Garantir dados confiáveis de deploy, incidentes e recuperação.
A produtividade está sustentável no trabalho remoto?	SPACE	Coletar percepções de forma ética e não punitiva.
Há gargalos em revisão, tarefas ou código?	Repositórios de código	Evitar comparação individual e interpretar os dados pelo contexto.
Como montar uma visão executiva de produtividade?	DORA + SPACE + Repositórios	Selecionar poucos indicadores, rastreáveis e alinhados ao objetivo da equipe.

Fonte: Elaborado pelo autor (2026).

O Quadro 4 organiza a recomendação final da seção de resultados. Sua função não é substituir a análise anterior, mas traduzir a comparação em perguntas práticas de gestão. A síntese confirma que a escolha de métricas deve começar pelo objetivo gerencial: estabilidade, sustentabilidade, diagnóstico técnico ou visão executiva. Dessa forma, os resultados dialogam com a literatura que alerta contra o uso de métricas isoladas e reforçam a necessidade de interpretação contextualizada.

4.8 AMEAÇAS À VALIDADE

Como a pesquisa é qualitativa e baseada em literatura, há limitações que precisam ser consideradas. A primeira ameaça é de construto: produtividade é um conceito amplo e pode ser interpretado de formas diferentes conforme o contexto da organização. Para reduzir esse risco, foram usados quatro critérios de comparação e modelos reconhecidos na literatura.

A segunda ameaça é de validade externa: os resultados não foram testados empiricamente em equipes reais. Assim, a matriz comparativa deve ser interpretada como instrumento de apoio à decisão, não como conclusão universal. A terceira ameaça é de conclusão: as pontuações são ordinais e analíticas, não estatísticas. Por isso, o estudo complementou a média simples com análise de sensibilidade por cenário.

A quarta ameaça está relacionada à atualidade das fontes técnico-científicas. Modelos e ferramentas de engenharia analítica evoluem rapidamente. Por isso, recomenda-se que aplicações futuras atualizem o corpus e validem os critérios com dados reais de equipes de desenvolvimento.

4.9 IMPLICAÇÕES PRÁTICAS

A primeira implicação prática é que equipes remotas devem iniciar a medição de produtividade pela pergunta de gestão, e não pela métrica disponível. Quando a preocupação é previsibilidade de entrega, métricas DORA são mais indicadas. Quando a preocupação é comunicação, bem-estar ou sustentabilidade, o SPACE deve ser priorizado. Quando o objetivo é identificar gargalos operacionais, métricas de repositórios podem oferecer sinais iniciais úteis.

A segunda implicação é a necessidade de governança dos dados. Indicadores de produtividade devem registrar origem, periodicidade, regra de cálculo, transformação aplicada e possíveis falhas de coleta. Sem rastreabilidade, dashboards podem apresentar números aparentemente precisos, mas incorretos. Em ambientes remotos, essa falha é especialmente perigosa porque os gestores tendem a depender mais de dados digitais do que de observação presencial.

A terceira implicação é a ética. Métricas de produtividade devem ser aplicadas preferencialmente no nível da equipe e utilizadas para melhoria contínua.

O uso individual e punitivo pode gerar competição improdutiva, manipulação de indicadores e redução de confiança. A medição deve apoiar conversas de aprendizagem, não substituir o julgamento técnico e humano dos profissionais.

A quarta implicação é a simplicidade. Um painel com muitas métricas pode dificultar decisões. Recomenda-se selecionar poucos indicadores por dimensão, vinculá-los aos objetivos do time e revisá-los periodicamente. Em equipes com baixa maturidade analítica, o caminho mais seguro é começar por DORA e repositórios, acrescentando dimensões SPACE de forma gradual.

Quadro 5 - Guia prático de adoção dos modelos

Nível de maturidade da equipe	Recomendação inicial	Evolução recomendada
Baixa maturidade analítica	Coletar poucos indicadores de fluxo e revisão, como lead time, tempo de revisão e defeitos.	Padronizar registros, revisar critérios e evitar métricas individuais.
Maturidade intermediária	Adotar DORA e indicadores de repositório com dashboards simples.	Incluir percepção da equipe, retrospectivas e sinais de colaboração.
Alta maturidade analítica	Combina DORA, SPACE e dados de repositórios com governança de dados.	Aplicar análise de tendências, segmentação por contexto e estudos de impacto.

Fonte: Elaborado pelo autor (2026).

O Quadro 5 demonstra que a adoção dos modelos deve evoluir conforme a maturidade analítica da equipe. Em equipes iniciantes, a prioridade deve ser padronizar registros e evitar métricas individuais. Em equipes intermediárias, DORA e repositórios podem apoiar dashboards simples. Em equipes maduras, a combinação com SPACE permite uma leitura mais completa, capaz de relacionar desempenho operacional, colaboração e sustentabilidade.

4.9.1 Painel Mínimo Recomendado Por Objetivo De Gestão

Para aproximar a análise do uso real em equipes remotas, o Quadro 6 apresenta um painel mínimo de indicadores por objetivo de gestão. A proposta evita excesso de métricas e reforça que cada indicador deve responder a uma pergunta concreta. O painel também diferencia periodicidade de acompanhamento, pois nem toda métrica precisa ser analisada diariamente.

Quadro 6 - Painel mínimo de métricas por objetivo de gestão

Objetivo de gestão	Indicadores recomendados	Modelo predominante	Periodicidade sugerida
Aumentar previsibilidade de entrega	Lead time, frequência de implantação, throughput e itens bloqueados.	DORA + repositórios	Semanal ou por sprint.
Reduzir retrabalho e incidentes	Taxa de falha em mudanças, defeitos pós-entrega, tempo de recuperação e cobertura de testes.	DORA + repositórios	Por release e mensal.
Melhorar colaboração remota	Tempo de revisão, participação em revisão, percepção de alinhamento e bloqueios de comunicação.	SPACE + repositórios	Por sprint ou mensal.
Monitorar sustentabilidade do trabalho	Satisfação, sobrecarga percebida, autonomia, foco e equilíbrio trabalho-vida.	SPACE	Mensal ou trimestral.
Relacionar tecnologia e valor	Objetivos entregues, satisfação de stakeholders, previsibilidade e redução de custos ou riscos.	DORA + SPACE	Mensal ou por ciclo de produto.

Fonte: Elaborado pelo autor (2026).

O Quadro 6 evidencia que a seleção de indicadores depende da finalidade. Para previsibilidade, DORA e repositórios são suficientes para iniciar a leitura. Para sustentabilidade e colaboração, os indicadores devem incorporar percepções da equipe, preferencialmente de forma agregada e não punitiva. Para relação entre tecnologia e valor, a melhor alternativa é combinar métricas técnicas e organizacionais, conectando entregas a objetivos de produto ou negócio.

4.9.2 Exemplo Demonstrativo De Interpretação Integrada

Considere uma equipe remota que apresenta aumento de 35% no número de commits e redução de 20% no tempo médio de revisão, mas também elevação de defeitos pós-entrega e queda na satisfação percebida. Uma leitura baseada apenas em repositórios poderia concluir que a produtividade aumentou. Uma leitura integrada, entretanto, indicaria que houve intensificação da atividade técnica, mas possivelmente com perda de qualidade e sustentabilidade.

Nesse cenário demonstrativo, o DORA ajudaria a identificar se a velocidade adicional comprometeu estabilidade; o SPACE permitiria investigar sobrecarga, colaboração e bem-estar; e as métricas de repositórios ajudariam a localizar

gargalos específicos no fluxo de revisão. Assim, o valor da análise não está em eleger uma métrica isolada, mas em combinar sinais para orientar perguntas melhores e ações de melhoria.

Quadro 7 - Exemplo demonstrativo de leitura integrada

Sinal observado	Interpretação isolada	Interpretação integrada recomendada
Aumento de commits	Maior atividade da equipe.	Pode indicar avanço, fragmentação artificial ou maior retrabalho; deve ser cruzado com qualidade.
Redução do tempo de revisão	Fluxo mais rápido.	Pode indicar eficiência, mas também revisão superficial; deve ser cruzado com defeitos pós-entrega.
Aumento de incidentes	Piora de estabilidade.	Indica necessidade de avaliar testes, revisão, pressão por prazo e complexidade das mudanças.
Queda de satisfação	Problema subjetivo ou externo.	Sinal de risco à sustentabilidade e possível antecedente de perda futura de produtividade.

Fonte: Elaborado pelo autor (2026).

O Quadro 7 reforça a comparação crítica entre os modelos. A métrica de commits, por exemplo, pode parecer positiva quando analisada isoladamente, mas perde força explicativa quando associada a incidentes e satisfação. Esse tipo de leitura integrada atende ao conceito de métrica híbrida definido no início do artigo, pois combina sinais técnicos, operacionais e humanos para reduzir distorções.

5 CONCLUSÃO

Este artigo teve como objetivo geral contrastar modelos híbridos de métricas para avaliação da produtividade em equipes de desenvolvimento de software em ambiente remoto. O objetivo foi atingido por meio da análise e comparação dos modelos DORA, SPACE e métricas baseadas em repositórios de código, considerando precisão, aplicabilidade, escalabilidade e impacto organizacional.

O primeiro objetivo específico, analisar DORA, SPACE e métricas baseadas em repositórios, foi atendido na seção de resultados, em que cada abordagem foi discutida segundo suas contribuições e limitações. O segundo objetivo, definir critérios de comparação, foi atendido com a construção do item de precisão, aplicabilidade, escalabilidade e impacto organizacional. O terceiro objetivo, realizar

análise comparativa entre os modelos identificados, foi cumprido por meio da matriz multicritério e da análise de sensibilidade por cenário. O quarto objetivo, examinar implicações práticas, foi atendido ao apresentar recomendações para adoção gradual, governança de dados, uso ético de métricas e alinhamento entre indicadores e objetivos da equipe.

Conclui-se que não existe um modelo universalmente superior para avaliar produtividade remota em equipes de software. O DORA é mais indicado para fluxo e estabilidade de entrega; o SPACE é mais adequado para compreender fatores humanos, colaboração e sustentabilidade; e métricas de repositórios são úteis para diagnóstico operacional, desde que interpretadas com cautela. O melhor resultado prático tende a surgir da combinação consciente desses modelos, orientada pelo contexto e pela finalidade da medição.

A pesquisa também reconhece limitações importantes. Não foram coletados dados reais de equipes, nem realizada validação experimental ou estudo de caso. As notas atribuídas são ordinais e analíticas, baseadas em literatura e não em medição estatística. Por esse motivo, os resultados devem ser compreendidos como instrumento de apoio à decisão e ponto de partida para estudos empíricos futuros.

Como trabalhos futuros, recomenda-se aplicar a matriz proposta em equipes reais de desenvolvimento remoto, coletando dados de Git, Jira, CI/CD, incidentes, testes e instrumentos de percepção da equipe. Também se sugere validar os critérios com especialistas, comparar resultados entre equipes com diferentes níveis de maturidade, acompanhar indicadores longitudinalmente por vários ciclos de desenvolvimento e investigar o impacto de ferramentas de inteligência artificial na produtividade, qualidade e experiência dos desenvolvedores.

Outra possibilidade de continuidade é desenvolver um dashboard híbrido validado empiricamente, com dicionário de métricas, regras de cálculo, trilha de auditoria, visualizações por objetivo de gestão e mecanismos para impedir uso individual ou punitivo dos indicadores. Esse tipo de ferramenta permitiria transformar a contribuição conceitual deste trabalho em aplicação prática para equipes de desenvolvimento de software em regime remoto.

DECLARAÇÃO DE USO DE INTELIGÊNCIA ARTIFICIAL

Os autores declaram que ferramentas de Inteligência Artificial generativa foram utilizadas exclusivamente como apoio auxiliar à elaboração deste trabalho, incluindo assistência na revisão linguística, estruturação textual e organização de ideias. Nenhuma ferramenta de IA foi utilizada para substituição da autoria intelectual, análise científica, interpretação dos resultados ou tomada de decisões acadêmicas. A responsabilidade integral pelo conteúdo, originalidade, veracidade das informações e conformidade com os princípios de integridade científica permanece integralmente atribuída aos autores.

REFERÊNCIAS

- AGRAWAL, Y.; JAIN, P. **From Ad-Hoc Scripts to Orchestrated Pipelines: Architecting a Resilient ELT Framework for Developer Productivity Metrics**. arXiv, 2026. Disponível em: <https://arxiv.org/abs/2602.21568> . Acesso em: 09 mar. 2026.
- ANDRADE, P. A. F.; SOUZA, M. J. R. **O impacto da implementação do teletrabalho nas análises dos projetos de pesquisa, desenvolvimento e inovação na Superintendência da Zona Franca de Manaus**. Revista ReCaPe, São Paulo, v. 13, n. 1, p. 91-113, 2023. Disponível em: <https://revistas.pucsp.br/index.php/ReCaPe/article/view/55138>. Acesso em: 09 mar. 2026.
- BARCELOS, M. R. S.; GOMES, C. F. S.; SANSEVERINO, A. M.; SANTOS, M. **Revisão de literatura sobre métricas de software e uma nova proposta**. Perspectivas Online: Exatas & Engenharia, Campos dos Goytacazes, v. 11, n. 32, p. 33-59, 2021. Disponível em: https://ojs3.perspectivasonline.com.br/exatas_e_engenharia/article/view/2284. Acesso em: 09 mar. 2026.
- CANNO, L. C.; DANTAS, L. P.; SOUSA, F. S.; SILVA, M. C. B. **As métricas de software em favor do desenvolvedor como ferramentas para a produtividade e a qualidade em soluções embarcadas**. Revista Brasileira de Mecatrônica, São Caetano do Sul, v. 5, n. 4, p. 106-120, 2023. Disponível em: <https://revistabrmeccatronics.sp.senai.br/ojs/index.php/revistabrmeccatronics/article/view/213>. Acesso em: 09 mar. 2026.
- FERNANDES, J.; MACHADO, L.; ARAÚJO, L. A. **Impactos do trabalho remoto sobre a produtividade no pós-Covid-19: estudo de caso em uma empresa de consultoria política**. Caderno de Propriedade Intelectual e Transferência de Tecnologia, v. 7, n. 2, p. 322-338, 2025. Disponível em: <https://publicacoes.ifc.edu.br/index.php/CPITT/article/view/7247> . Acesso em: 09 mar. 2026.

FERREIRA, P.; NICIDA, J.; VALENTE, L. **Desmistificando a produtividade e os efeitos do home office integral durante a pandemia**. Harvard Business Review Brasil, 20 jul. 2020. Disponível em: <https://hbrbr.com.br/desmisticando-a-produtividade-e-os-efeitos-do-home-office-integral-durante-a-pandemia/>. Acesso em: 09 mar. 2026.

FORSGREN, N.; HUMBLE, J.; KIM, G.. **Accelerate**: the science of lean software and DevOps. Portland: IT Revolution, 2018. Disponível em: <https://itrevolution.com/product/accelerate/>. Acesso em: 09 mar. 2026.

FORSGREN, N. et al. **The SPACE of developer productivity**: there is more to it than you think. Queue, New York, v. 19, n. 1, p. 20-48, 2021. Disponível em: <https://queue.acm.org/detail.cfm?id=3454124>. Acesso em: 09 mar. 2026.

HOUCK, B.; LOWDERMILK, T.; BEYER, C.; CLARKE, S.; HANRAHAN, B. **The SPACE of AI**: real-world lessons on AI's impact on developers. arXiv, 2025. Disponível em: <https://arxiv.org/abs/2508.00178>. Acesso em: 09 mar. 2026.

JAIN, P.; AGRAWAL, Y.; NIGAM, A.; PATIL, P.. **MetricSynth**: framework for aggregating DORA and KPI metrics across multi-platform engineering. arXiv, 2025. Disponível em: <https://arxiv.org/abs/2511.06864>. Acesso em: 09 mar. 2026.

LEITE, A. L.; LEMOS, D. C. **Gestão de pessoas e o teletrabalho**: desafios e possibilidades. Revista do Serviço Público, Brasília, v. 72, n. 2, p. 330-359, 2021. Disponível em: <https://repositorio.enap.gov.br/handle/1/6655>. Acesso em: 09 mar. 2026.

LEITE, A. L.; SILVEIRA, T. B. J.; LEMOS, D. C. **Implicações do trabalho remoto em uma empresa de base tecnológica**. Navus: Revista de Gestão e Tecnologia, Florianópolis, v. 10, p. 1-15, 2020. Disponível em: <https://navus.sc.senac.br/navus/article/view/1318>. Acesso em: 09 mar. 2026.

MAGALHÃES, E. J.; RODRIGUES, J. C.. **Análise comparativa do uso de métricas como fator de qualidade no desenvolvimento de software**. Anais do Simpósio Unificado dos Cursos de Sistemas de Informação da UEG, Goiás, p. 151-163, 2016. Disponível em: <https://www.anais.ueg.br/index.php/siuniueg/article/view/9170>. Acesso em: 09 mar. 2026.

PEREIRA, L. R. **Construção do modelo lógico do programa de gestão e desempenho na modalidade de teletrabalho**: uma proposta de avaliação no âmbito do FNDE. 2024. Dissertação (Mestrado Profissional em Gestão Pública) – Universidade de Brasília, Brasília, 2024. Disponível em: <https://repositorio.unb.br/handle/10482/50842>. Acesso em: 09 mar. 2026.

SILVA, Y. A.; OLIVEIRA, V. J. S.; CARVALHO, H. K. S.; SILVA, W. V.; SANTOS, R. D. F.; DALAZEN, L. L. **Mapeamento sistemático das métricas de desempenho financeiro e organizacional**. Ciências Sociais Aplicadas em Revista, Marechal Cândido Rondon, v. 29, n. 49, p. 1-30, 2024. Disponível em:

<https://e-revista.unioeste.br/index.php/csaemrevista/article/view/34520>. Acesso em: 09 mar. 2026.

SILVA, E. T.; SILVA, A. M.; FILHO, M. G. **Qualidade de vida e produtividade do colaborador na modalidade de trabalho home office**: estudo de caso. Brazilian Journal of Production Engineering, 2022. Disponível em: <https://periodicos.ufes.br/bjpe/article/view/37033>. Acesso em: 09 mar. 2026.