

**Pontifícia Universidade Católica de Minas Gerais**

**Laboratório de Experimentação de Software**

**Caracterizando a atividade de code review**

Giovanni Bogliolo Sirihal Duarte  
Luiz Gustavo Mendes Santos  
Pedro Ramos Vidigal

**Belo Horizonte  
2024**

# 1. Introdução

O code review, componente essencial no desenvolvimento ágil, assegura a qualidade do código por meio da inspeção e colaboração entre desenvolvedores. No GitHub, o code review se dá pela análise de Pull Requests antes de sua integração à branch principal.

Este laboratório se propõe a investigar a atividade de code review em repositórios populares do GitHub, buscando identificar variáveis que influenciam a aprovação de um PR sob a perspectiva do desenvolvedor que submete o código.

Com base na análise nos dados coletados, serão respondidas questões como: PRs menores são aprovados com mais facilidade? O tempo de análise influencia na decisão final? Uma descrição detalhada aumenta as chances de aprovação? A interação entre desenvolvedores e revisores impacta na qualidade do código? As respostas a essas perguntas contribuirão para uma melhor compreensão do processo de code review e fornecerão informações valiosas para otimizar essa prática crucial no desenvolvimento de software.

## 1.1. Hipóteses

**RQ 01. Qual a relação entre o tamanho dos PRs e o feedback final das revisões?**

**R:** PRs menores tendem a ser aprovados com mais facilidade, pois são mais fáceis de revisar e entender, já PRs maiores podem gerar mais dúvidas e debates, levando a um maior tempo de análise e, possivelmente, rejeição.

**RQ 02. Qual a relação entre o tempo de análise dos PRs e o feedback final das revisões?**

**R:** PRs analisados rapidamente podem indicar clareza e qualidade no código, levando à aprovação, já PRs com longos tempos de análise podem sugerir a necessidade de revisões e correções, aumentando a chance de rejeição.

**RQ 03. Qual a relação entre a descrição dos PRs e o feedback final das revisões?**

**R:** Descrições claras e detalhadas facilitam a compreensão do código e aumentam a chance de aprovação, descrições vagas ou incompletas podem gerar dúvidas e atrasar o processo de revisão.

**RQ 04. Qual a relação entre as interações nos PRs e o feedback final das revisões?**

**R:** PRs com muitas interações podem indicar um debate construtivo e colaboração, aumentando a chance de aprovação após as devidas correções, PRs com poucas interações podem sugerir falta de atenção ou dificuldade de entendimento do código.

**RQ 05. Qual a relação entre o tamanho dos PRs e o número de revisões realizadas?**

**R:** PRs maiores podem demandar mais revisões para garantir a qualidade do código. PRs menores podem ter menos revisões, especialmente se forem simples e de fácil compreensão.

**RQ 06. Qual a relação entre o tempo de análise dos PRs e o número de revisões realizadas?**

**R:** PRs com maior tempo de análise podem ter mais revisões, refletindo a necessidade de múltiplas iterações para aprovação. PRs com menor tempo de análise podem ter menos revisões, indicando clareza e qualidade do código.

**RQ 07. Qual a relação entre a descrição dos PRs e o número de revisões realizadas?**

**R:** Descrições claras e detalhadas podem reduzir a necessidade de múltiplas revisões. Descrições vagas ou incompletas podem gerar dúvidas e demandar mais revisões para esclarecimento.

**RQ 08. Qual a relação entre as interações nos PRs e o número de revisões realizadas?**

**R:** PRs com muitas interações podem ter mais revisões, refletindo o debate e a colaboração entre desenvolvedores e revisores já PRs com poucas interações podem ter menos revisões, especialmente se o código for claro e de fácil entendimento.

## **2. Metodologia**

Será utilizada a API do GitHub para coletar dados sobre os Pull Requests (PRs) dos 200 repositórios open source mais populares da plataforma. Para garantir a robustez da análise, serão selecionados apenas repositórios com pelo menos 100 PRs incluindo aqueles com status MERGED e CLOSED. A linguagem de programação Python será usada para automatizar a coleta, processamento e análise dos dados e serão utilizadas também bibliotecas como requests e pandas para manipulação de dados, análise estatística e visualização.

A fim de analisar PRs que passaram por um processo de code review, serão aplicados os seguintes filtros: apenas PRs com status MERGED ou CLOSED, pois estes representam PRs que foram efetivamente revisados e avaliados, incluir apenas PRs com pelo menos uma revisão registrada, garantindo que houve análise humana e considerar apenas PRs com tempo de revisão mínimo de uma hora, excluindo PRs com revisões automáticas ou superficiais.

Para cada PR selecionado, serão extraídas as seguintes métricas: número de arquivos modificados, total de linhas adicionadas e removidas, intervalo entre a criação do PR e a última atividade, número de caracteres no corpo da descrição do PR, número de participantes e número de comentários no PR, quantidade de revisões realizadas no PR e status do PR (MERGED ou CLOSED).

Com base nas métricas extraídas a partir daí, serão realizadas análises quantitativas e qualitativas para investigar as relações entre as variáveis e responder às questões de pesquisa. E por fim os resultados obtidos serão comparados com as hipóteses iniciais, buscando identificar padrões e tendências na atividade de code review.

## **3. Resultados obtidos**

### **4. Análise dos resultados**

Nesta seção, serão apresentadas as análises dos resultados obtidos para cada pergunta de pesquisa.

**RQ 01. Qual a relação entre o tamanho dos PRs e o feedback final das revisões?**

**R:**

**RQ 02. Qual a relação entre o tempo de análise dos PRs e o feedback final das revisões?**

**R: RQ 03. Qual a relação entre a descrição dos PRs e o feedback final das revisões?**

**R:**

**RQ 04. Qual a relação entre as interações nos PRs e o feedback final das revisões?**

**R: RQ 05. Qual a relação entre o tamanho dos PRs e o número de revisões realizadas?**

**R:**

**RQ 06. Qual a relação entre o tempo de análise dos PRs e o número de revisões realizadas?**

**R:**

**RQ 07. Qual a relação entre a descrição dos PRs e o número de revisões realizadas?**

**R:**

**RQ 08. Qual a relação entre as interações nos PRs e o número de revisões realizadas?**

**R:**