

Caracterizando a atividade de code review no GitHub

Lucas Cabral Soares

Lucas Hemétrio

Maria Eduarda Amaral Muniz

1. Introdução

A prática de code review desempenha um papel crucial na garantia da qualidade do código e na redução de defeitos nos projetos de software, especialmente em ambientes open source como os hospedados no GitHub. Este laboratório visa investigar as variáveis que influenciam o feedback final e o número de revisões de Pull Requests (PRs) em repositórios populares do GitHub.

1.1. Hipóteses

Antes da análise dos dados, formulamos algumas hipóteses sobre os sistemas populares:

- **RQ01: Qual a relação entre o tamanho dos PRs e o feedback final das revisões?**
PRs com um número maior de arquivos ou com mais linhas modificadas tendem a receber um feedback final mais crítico ou com maior número de rejeições, devido à complexidade das mudanças. Acredita-se que PRs menores e mais focados possuam maior chance de aprovação rápida.
- **RQ02: Qual a relação entre o tempo de análise dos PRs e o feedback final das revisões?**
PRs que passam por um tempo de análise mais longo tendem a receber um feedback final mais detalhado e positivo, uma vez que as revisões mais cuidadosas e prolongadas resultam em menos defeitos e mais precisão nas alterações.
- **RQ03: Qual a relação entre a descrição dos PRs e o feedback final das revisões?**
PRs com descrições mais completas e detalhadas provavelmente recebem um feedback final mais favorável, pois fornecem um contexto melhor para os revisores e facilitam a compreensão das alterações propostas.
- **RQ04: Qual a relação entre as interações nos PRs e o feedback final das revisões?**

PRs que geram mais interações, como discussões e comentários entre os colaboradores, têm maior chance de obter um feedback final positivo, pois indicam que o código foi mais analisado e ajustado conforme as sugestões dos revisores.

- **RQ05: Qual a relação entre o tamanho dos PRs e o número de revisões realizadas?**

PRs maiores em termos de arquivos modificados ou linhas alteradas exigem um maior número de revisões, já que as mudanças são mais complexas e precisam de mais checagens para garantir sua qualidade.

- **RQ06: Qual a relação entre o tempo de análise dos PRs e o número de revisões realizadas?**

PRs que passam por uma análise mais prolongada tendem a passar por mais revisões, já que revisores podem fazer ajustes adicionais ao longo do tempo, à medida que identificam diferentes aspectos da modificação.

- **RQ07: Qual a relação entre a descrição dos PRs e o número de revisões realizadas?**

PRs com descrições mais completas tendem a gerar um maior número de revisões, pois ajudam os revisores a compreender melhor as intenções do autor, resultando em mais feedbacks e ajustes.

- **RQ08: Qual a relação entre as interações nos PRs e o número de revisões realizadas?**

PRs que envolvem um maior número de interações entre os colaboradores (comentários e discussões) tendem a resultar em um maior número de revisões, pois as discussões promovem ajustes constantes e mais detalhados no código proposto.

2. Metodologia

2.1 Coleta dos Dados do GitHub

2.2 Métricas

3. Resultados e Discussão

4. Conclusão