

Pontifícia Universidade Católica de Minas Gerais
Curso de Engenharia de Software

Disciplina: Laboratório de Engenharia de Software

Relatório da Sprint 1

Projeto: Caracterizando a Atividade de Code Review no GitHub

Aluno: Lúcio Alves, Davi Ferreira

1. Introdução

A prática de *code review* é parte essencial no processo de desenvolvimento de software, principalmente em projetos colaborativos. No contexto do GitHub, essa atividade é realizada por meio de *Pull Requests* (PRs), que são avaliados por revisores antes de serem incorporados à *branch* principal.

O presente trabalho tem como objetivo analisar as atividades de *code review* em repositórios populares do GitHub, buscando identificar variáveis que influenciam no *merge* de PRs.

Na **Sprint 1**, focamos em duas atividades principais:

1. Seleção dos repositórios que comporão o dataset do estudo.
2. Criação do script inicial de coleta de PRs e das métricas necessárias para análise.

2. Seleção dos Repositórios

Para a seleção dos repositórios, adotamos os seguintes critérios:

- Repositórios populares no GitHub (Top 200 com maior número de *stars*).
- Repositórios com pelo menos 100 PRs com status *Merged* ou *Closed*.

3. Script de Coleta

Para a coleta de dados, optamos por utilizar a API REST do GitHub. Essa API possibilita a extração de informações detalhadas sobre PRs, como status, número de comentários, revisores, arquivos alterados, linhas adicionadas/removidas e tempo de análise.

O arquivo pode ser encontrado no repositório com o nome de [main.py](#).

4. Considerações Finais

A Sprint 1 teve como foco a preparação do ambiente para as análises subsequentes. Nessa etapa, selecionamos os repositórios mais relevantes e desenvolvemos o script inicial de coleta de dados.

Como próximos passos (Sprint 2), será realizada a construção do dataset completo com todas as métricas definidas e a elaboração das primeiras hipóteses relacionadas às questões de pesquisa.