

RESUMO

O desenvolvimento distribuído de software *open-source* tem no modelo baseado em *pull requests* uma importante ferramenta para que desenvolvedores externos possam contribuir no projeto de software. Um *pull request* deve ser analisado por um integrador, membro da equipe principal do projeto, que decidirá pela sua aceitação ou rejeição. A análise de *pull requests* pode ser demorada em decorrência da complexidade das mudanças, isso consome tempo do integrador que poderia ser utilizado para desenvolvimento. Por ser um processo decisório, pode ser útil realizar a previsão sobre a integração de um *pull request*. Este cenário já foi explorado em trabalhos anteriores, porém com uma abordagem diferente tanto para os atributos como para os algoritmos propostos. O objetivo principal deste trabalho, consiste em verificar se novos atributos melhoram a previsão de integração de *pull requests*. Para a realização dos experimentos buscou-se seguir as fases do processo de descoberta de conhecimento em base de dados. Além disso, foi realizada a calibragem do método de avaliação, dos algoritmos e de estratégias de seleção de atributos. Essas atividades possibilitaram uma comparação dos conjuntos de dados obtidos com os utilizados em trabalhos anteriores. Como resultado, o conjunto de dados proposto por esse trabalho se mostrou ligeiramente melhor do que o usado em trabalhos anteriores, obtendo 83,51% de acurácia em média.

Palavras-chaves: *pull request*, mineração de dados, *open-source*, desenvolvimento distribuído, controle de mudanças, classificação.