

Proposta de TCC

Yesman Choque Mamani
Orientador: Prof. Dr. Fabio Kon

Motivação

A qualidade do código é fundamental em qualquer projeto de desenvolvimento de software, o que se estende ao contexto de sistemas baseados em ciência de dados e aprendizado de máquina.

A crescente complexidade dos algoritmos de aprendizado e a diversidade de bibliotecas e arcabouços para ciência de dados disponíveis reforçam a necessidade de padrões de qualidade de código para a área. Durante o desenvolvimento desses projetos, a atenção é direcionada geralmente aos algoritmos, dados e às interfaces das bibliotecas de ciência de dados e aprendizado de máquina, deixando em segundo plano a qualidade do código. Essa negligência pode resultar em um software propenso a erros, de difícil entendimento e manutenção, resultando na presença de diversos maus cheiros (“code smells”). Estes elementos facilitam o aparecimento de bugs e prejudicam a colaboração entre desenvolvedores, podendo ir desde funções com muitos parâmetros até a falta de modularização em classes.

Desse modo, inspirado nos princípios estabelecidos na literatura, em particular no livro *Clean Code* de Robert C. Martin, a busca por boas práticas que garantam um código de qualidade se torna necessária para a eficiência, compreensão e manutenção desses sistemas a longo prazo.

Objetivos

O presente projeto tem como objetivo desenvolver uma ferramenta web para identificar maus cheiros em projetos de ciência de dados e aprendizado de máquina. Essa ferramenta não se destina apenas a profissionais da área de computação, mas, principalmente, visa atender especialistas de diversas outras áreas, que precisam garantir a qualidade e a legibilidade de seus códigos. Dessa forma, a ferramenta permitirá a detecção de práticas inadequadas de programação que comprometem a eficiência e a manutenção do código, possibilitando refatorações mais precisas.

Além disso, o projeto propõe a criação de um painel interativo e intuitivo, que reunirá informações detalhadas sobre a qualidade do código. Esse painel apresentará métricas relevantes, como complexidade ciclomática, código duplicado e outras métricas de tamanho, permitindo uma avaliação rápida dos aspectos críticos do projeto

Metodologia

Para a organização e registro das atividades e reuniões com o orientador, adotou-se a plataforma de anotações *Notion*. O trabalho teve início em 2024, com o estudo da literatura formal, em particular os livros: *Clean Code*, *Introduction to Software Design with Java* e *Implementation Patterns*. Além disso, foram consultados blogs e tutoriais pertinentes ao trabalho.

No próximo passo, iniciou-se a análise de três projetos de código aberto da área de ciência de dados e aprendizado de máquina, sob o ponto de vista da qualidade de código. Com o auxílio do *Pylint*, ferramenta de análise estática de código fonte, e do referencial teórico, identificaram-se maus cheiros de código e antipadrões em cada projeto.

Após a detecção, os antipadrões mais relevantes foram selecionados, considerando tanto sua frequência de ocorrência quanto o impacto na performance das aplicações. Com base nessa seleção, uma descrição clara foi redigida para cada uma, trabalho que gerou um catálogo de antipadrões para projetos da área.

No ano de 2025, iniciou-se a construção da ferramenta que visa auxiliar desenvolvedores, com base no catálogo desenvolvido, na identificação dos principais antipadrões em seus projetos. Utilizando o *Figma*, ferramenta para elaboração e prototipação de interfaces, o design de um dashboard interativo com as principais funcionalidades foi elaborado, servindo de base para implementação incremental da aplicação.

O desenvolvimento da ferramenta está hospedado no *GitLab* e foi organizado em *milestones* e *issues*, cada qual descrevendo uma das funcionalidades planejadas. A partir do mês de agosto, está prevista uma validação preliminar com usuários, cujo feedback será fundamental para o aprimoramento da ferramenta. No mês de outubro, outra validação está planejada para o desenvolvimento da versão final da aplicação.

Em paralelo, durante os meses de agosto a novembro, a monografia do TCC será elaborada, detalhando a implementação técnica da ferramenta, bem como os resultados obtidos através do feedback de usuários reais.

Cronograma

Atividades	2024	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV
Estudo da literatura formal e literatura cinzenta												
Análise de código de projetos de software livre com o Pylint e apoio da literatura												
Seleção dos antipadrões mais importantes e elaboração do catálogo												
Design e Prototipação do Dashboard interativo												
Implementação incremental do Dashboard interativo												
Refatoração do código para melhoria da arquitetura												
Validação preliminar com usuários												
Refinamento do Dashboard com base no feedback dos usuários												
Validação final do Dashboard com usuários												
Escrita e Revisão da monografia do TCC												