

Capítulo 09

TRABALHO EM EQUIPE E COLABORAÇÃO EM PROJETOS DE MACHINE LEARNING

Luciano Tadeu P.A.Pereira

2024

Como vimos nos capítulos anteriores, um projeto de Machine Learning envolve etapas e equipes, desde a equipe de negócio que vai ser o usuário dos resultados desse algoritmo até o engenheiro de Machine Learning que vai realizar a inserção e sustentação desse algoritmo em produção. Gerir todas essas equipes e etapas não é trabalho fácil e é isso que vamos discutir neste capítulo.

Equipe em um projeto de Machine Learning

A composição da equipe em um projeto de Machine Learning é essencial para um projeto bem-sucedido. Em geral, uma composição de uma equipe deve conter:

- **Analistas de Negócio:** são importantes para interpretar as necessidades comerciais e validar se as soluções técnicas realmente entregam valor para o negócio.
- **Engenheiros de Dados:** garantem que os dados certos estejam disponíveis e limpos, facilitando o trabalho dos cientistas de dados e fornecendo pipelines robustos de ingestão e transformação de dados.
- **Gerente de Projetos:** é o responsável por alinhar as entregas com os objetivos do projeto, gerenciar prazos e manter uma comunicação eficiente entre os membros da equipe e os stakeholders.
- **Cientistas de Dados:** são responsáveis pela criação de modelos, análise exploratória de dados e ajuste de algoritmos.
- **Engenheiros de Machine Learning:** se concentram em transformar os modelos em sistemas escaláveis e eficientes para uso em produção.

Canvas Machine Learning

O canvas machine learning é uma ferramenta onde é muito utilizada no início de um projeto de Machine Learning. Nela é possível alinhar as expectativas de forma clara e objetiva com a equipe de negócio. Além de definir onde buscar as coletas de dados, métricas de desempenho, definição do ambiente de produção e como deve ser realizada a manutenção e monitoramento do algoritmo.

Essa ferramenta é muita benéfica, principalmente em empresas que estão inicialmente desenvolvendo seus primeiros projetos de Machine Learning. A imagem abaixo mostra um exemplo fictício de um canvas machine learning preenchido.

Figura 1 - Canvas Machine Learning Preenchido

PREDICTION TASK Qual variável preditora? <i>Leads propensos a comprar ou não no canal B2B</i>	DECISIONS Como as predições são transformadas em valor para o usuário final? <i>Auxiliar a equipe de e-commerce a qualificar/ filtrar leads mais propensos a realizar uma compra</i>	VALUE PROPOSITION Quem será o usuário final? Quais objetivos? Como eles se beneficiaram do sistema de IA? Mencione! Fluxo de trabalho e interfaces. <i>Usuários final: Equipe E-commerce</i> <i>Objetivo: Qualificar Leads, a partir de uma lista de prospecção de clientes</i>	DATA COLLECTION Como será realizado o processo de coleta dos dados? <i>Será criado um script de pipeline em python para extração dos dados web e construção da base de dados</i>	DATA SOURCES Quais informações serão usadas para construção do modelo? Dados de train <i>Dados da receita federal</i>
IMPACT SIMULATION Como mencionar se o modelo pode ser implementado em produção? Quais métricas de <u>test</u> /simulação serão consideradas e o por que? <i>Será realizado o treinamento do algoritmo de ML, e testado com conjunto de dados de teste, tendo performances satisfatória o modelo ser implementado</i>	MAKING PREDICTIONS Como será o processo de deploy do modelo? <i>As predições são salvas em uma base, no Google Cloud, onde a equipe de TI e E-commerce poderão ter acessos, e realizar disparo de campanha de marketing e entrar em contato</i>	BENEFITS <i>Benefícios: Identificar novos potenciais leads, a comprarem</i> <i>Fluxo de trabalho: Após montagem da lista de prospecção ser realizada a predição de o cliente tem probabilidade de realizar a compra</i>	BUILDING MODELS Características do modelo a ser construído. <i>Modelo de classificação Binária utilizando técnicas de Machine Learning</i>	FEATURES Quais dados de inputs <i>Os dados a serem estudados serão: CNAE, Capital da empresa, idade da empresa, Categorias consumo ou fornecimento, CEP</i>
MONITORING Como será o monitoramento da performance do modelo, junto com seus impactos? <i>Após cada campanha será realizada uma análise/acompanhamento da performance do algoritmo</i>				

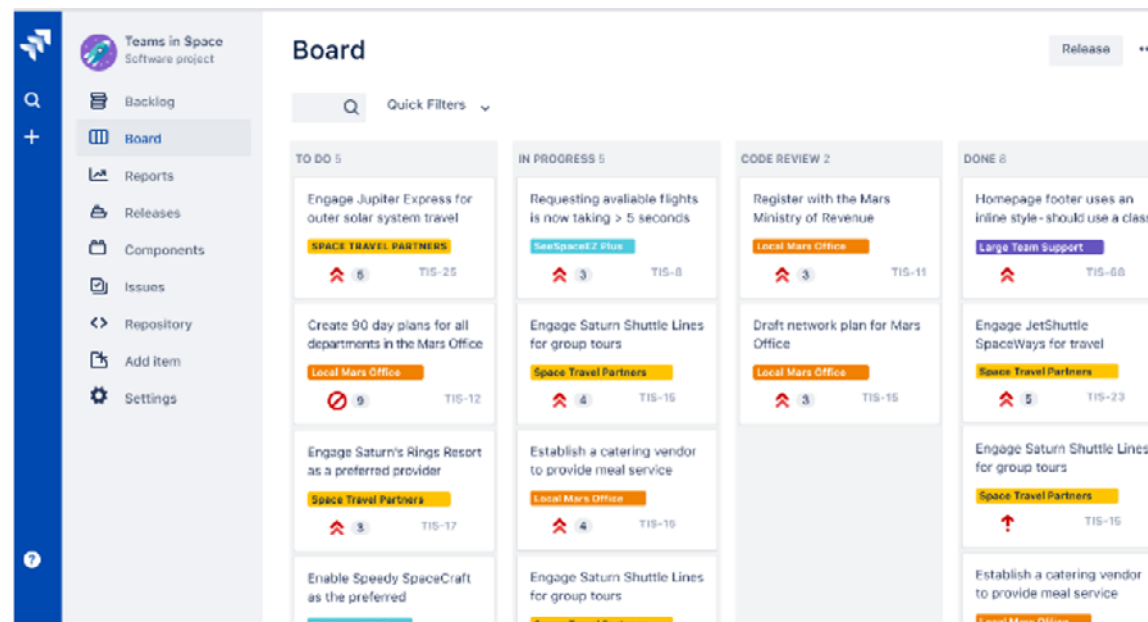
Fonte para download: Machine Learning Canvas — OWNML

Ferramentas de gerenciamento de projeto de Machine Learning

O gerenciamento de um projeto não é tarefa fácil, porém existem algumas ferramentas que podem ajudar na gestão. Vamos apresentar algumas que podem auxiliar em um projeto de Machine Learning:

- **GitHub** é amplamente utilizado para o gerenciamento de versões de código em projetos de Machine Learning. Ele permite que equipes de dados colaborem de maneira eficiente, controlando alterações no código-fonte, rastreando diferentes versões de scripts de treinamento e experimentos, e integrando pipelines de CI/CD para automação de testes e deploys.
- **Jira** é uma plataforma de gerenciamento de projetos muito utilizada para organizar tarefas, sprints e backlogs em equipes ágeis. No contexto de Machine Learning, o Jira é útil para gerenciar as atividades de desenvolvimento do projeto, como o acompanhamento de experimentos, atribuição de tarefas, definição de prazos e monitoramento do progresso da equipe. Abaixo é apresentada uma imagem da ferramenta Jira, é possível verificar que existe um design bem intuitivo e prático para diagnosticar quem está em quais atividades e o status delas.

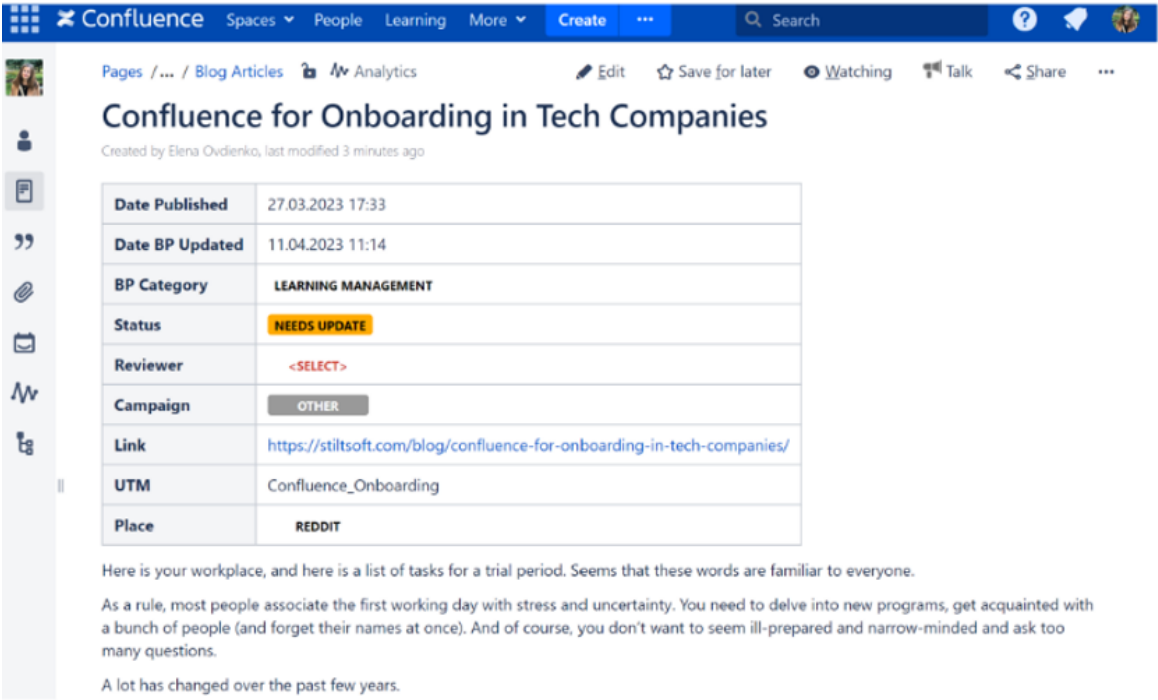
Figura 2 - Jira para gerenciamento de projetos.



Fonte: Domine o gerenciamento ágil de projetos com o Jira

- **MLflow** é uma ferramenta dedicada ao gerenciamento do ciclo de vida de modelos de Machine Learning. Ele oferece funcionalidades para rastrear experimentos, versionar modelos e facilitar o processo de deployment.
- **Confluence** é uma plataforma colaborativa desenvolvida pela Atlassian que permite a criação, organização e compartilhamento de documentações de projetos em um ambiente centralizado. Ele é amplamente utilizado por equipes para registrar informações essenciais, como requisitos de projeto, planos, relatórios de progresso, e guias técnicos, promovendo uma comunicação clara e eficiente entre membros.

Figura 3 - Confluence para gerenciamento de projetos.



Fonte: 4 Ways to Review Pages in Confluence - Stiltsoft

Essas ferramentas, quando usadas em conjunto, proporcionam uma gestão eficiente dos projetos de Machine Learning, desde o controle de código e tarefas até o gerenciamento do ciclo de vida dos modelos, garantindo transparência e organização em todas as etapas do processo.

Documentação de projetos

Documentar o projeto é uma tarefa essencial em todos os projetos por diversos motivos, entre eles estão:

- Transparência construção do algoritmo;
- Transparência técnicas e vies dos algoritmos;
- Transferência de conhecimento;
- Sustentabilidade do algoritmo.

Faculdade
XPe



xpeducacao.com.br