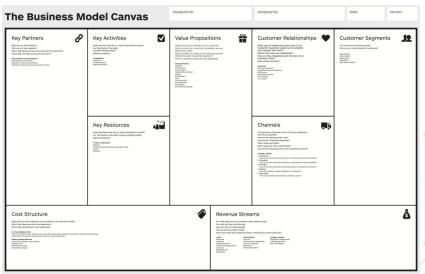
Machine Learning Canvas: Da coleta de dados à geração de valor

Alexandre Ray, Data Scientist 27/11/2019



Um pouco de contexto:

- 2015, Louis Dorard
- Baseada no Business Model Canvas
- Usado por empresas e Business Schools



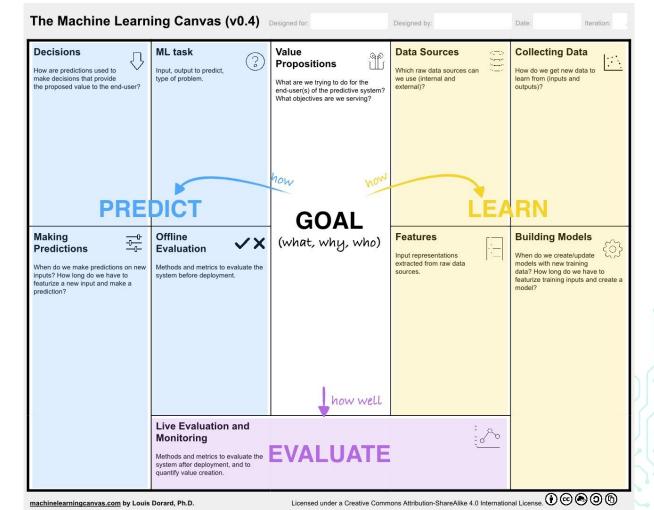
Fonte: https://pt.wikipedia.org/wiki/Business_Model_Canvas



Por que é importante?

- Sistemas de ML são complexos (dado certo, formato certo, modelo, predição)
- Não gastar tempo resolvendo os problemas errados
- Trazer para mais próximo da implementação
- As predições se tornam valiosas quando elas são usadas para tomada de decisões





Decisões predições feitas?

ML Task Como as predições Qual será o input, serão usadas para output e o tipo de tomar decisões? problema? **Predições** Quando e como fazer Avaliação Offline Quais métodos e/ou Quando faremos novas métricas podemos usar predições? Existe para avaliar o modelo restrição de tempo de antes do Deploy? predição? De quando em quando tempo as predições serão Avaliação Online e Monitoramento Ouais métodos e/ou métricas podemos usar para avaliar o

modelo EM PRODUÇÃO?

objetivos deste projeto? Objetivo: O quê Por quê Quem Quão bem Avaliação

Proposta de Valor

tentando resolver

para o(s) usuário(s)

final e quais são os

O que estamos



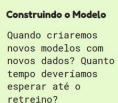
Features (Variáveis)

Quais possíveis

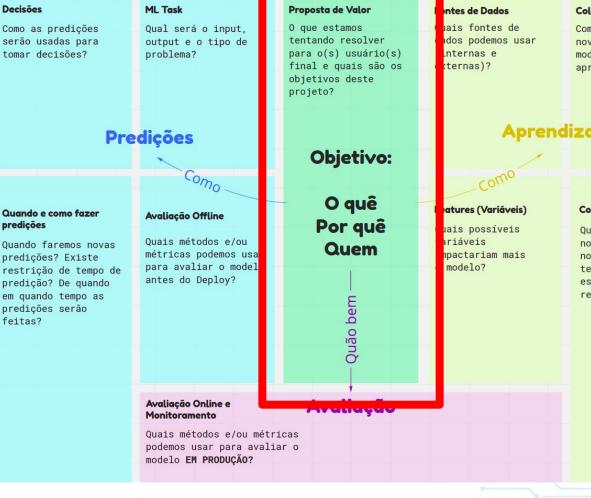
impactariam mais

variáveis

o modelo?







Coletando os Dados

Como vamos obter novos dados para o modelo continuar aprendendo?

Aprendizagem

Construindo o Modelo

Quando criaremos novos modelos com novos dados? Quanto tempo deveríamos esperar até o retreino?



Objetivo — Qual é a proposta de valor do projeto?

- **O quê:** O que estamos tentando fazer?
 - Diferenciar bons e maus pagadores
 - o Identificar clientes que podem abandonar a operação (churn)
- **Por quê:** Por quê é importante?
 - Minimizar risco financeiro
 - Tomar ações para retenção, receita
- Quem: Quem irá usar o projeto (usuário final)?
 - Área de risco de crédito
 - Time de CRM



Decisões Como as predições serão usadas para tomar decisões? Quando e como fazer predicões Quando faremos novas predições? Existe restrição de tempo de predição? De quando em quando tempo as predições serão feitas?

ML Task

Qual será o input, output e o tipo de problema?

Predições

Avaliação Offline

Quais métodos e/ou métricas podemos usar para avaliar o modelo antes do Deploy?

Avaliação Online e

Ouais métodos e/ou métricas podemos usar para avaliar o modelo EM PRODUÇÃO?

Monitoramento

Quão bem

Avaliação

Proposta de Valor

Objetivo:

O que estamos

objetivos deste projeto?

tentando resolver para o(s) usuário(s) final e quais são os

O quê Por quê variáveis Quem

Fontes de Dados

Quais fontes de dados podemos usar (internas e externas)?

Coletando os Dados

Como vamos obter novos dados para o modelo continuar aprendendo?

Aprendizagem

Features (Variáveis)

Quais possíveis Ouando criaremos impactariam mais o modelo? esperar até o

Construindo o Modelo

novos modelos com novos dados? Quanto tempo deveríamos retreino?



Predições

- ML task: Qual é o tipo do problema (ex. classificação, regressão...), qual é o input, e qual é o output da predição?
 - Input: cliente
 - Classificação binária
- Decisões: Como as predições serão usadas para tomada de decisões?
 - o Aprovação/Rejeição automática, Envio de e-mail para retenção
- Quando e como fazer as predições: Quando fazer novas predições? De quanto em quanto tempo as predições devem ser feitas?
 - A cada novo cliente ou todo 1º dia útil do mês? para todo mundo?
- Avaliação Offline: Quais métodos e métricas devem ser usados para avaliar as predições antes do Deploy?
 - Acurácia, KS, AUC, ROI



Decisões ML Task Proposta de Valor Fontes de Dados Coletando os Dados O que estamos Quais fontes de Como as predições Qual será o input, Como vamos obter tentando resolver dados podemos usar serão usadas para output e o tipo de novos dados para o problema? para o(s) usuário(s) (internas e tomar decisões? modelo continuar final e quais são os externas)? aprendendo? objetivos deste projeto? **Aprendizagem Predições** Objetivo: O quê Features (Variáveis) Construindo o Modelo Quando e como fazer Avaliação Offline Por quê predições Quais possíveis Ouando criaremos Quais métodos e/ou variáveis novos modelos com Quando faremos novas Quem métricas podemos usar impactariam mais novos dados? Quanto predições? Existe para avaliar o modelo o modelo? restrição de tempo de tempo deveríamos antes do Deploy? esperar até o predição? De quando retreino? em quando tempo as Quão bem predições serão feitas? Avaliação Online e Avaliação Monitoramento Ouais métodos e/ou métricas podemos usar para avaliar o modelo EM PRODUÇÃO?



Aprendizagem

- Fontes de dados: Quais fontes de dados serão usadas? (Internas e Externas)
 - CRM, Email, Website Analytics
- **Coletando os dados:** Como obter novos dados para o modelo continuar aprendendo?
 - o Os dados já estão armazenados? Precisa de DE? Precisa de Crowlers?
- Features (variáveis): Quais as possíveis variáveis podem ser importantes para resolver o problema proposto?
 - Eventos, interações com a plataforma, dados cadastrais, devices
- Construindo o modelo: Quando o modelo deverá ser atualizado?
 - De quanto em quanto tempo o modelo deve ser retreinado?



Decisões Como as predições serão usadas para tomar decisões? Quando e como fazer predições Quando faremos novas predições? Existe restrição de tempo de predição? De quando em quando tempo as predições serão feitas?

Qual será o input, output e o tipo de problema? **Predições** Avaliação Offline Quais métodos e/ou métricas podemos usar para avaliar o modelo antes do Deploy? Avaliação Online e Monitoramento Ouais métodos e/ou métricas podemos usar para avaliar o

modelo EM PRODUÇÃO?

ML Task

final e quais são os externas)? objetivos deste projeto? Objetivo: O quê Features (Variáveis) Por quê Quais possíveis variáveis Quem impactariam mais o modelo? Quão bem Avaliação

Fontes de Dados

(internas e

Ouais fontes de

dados podemos usar

Proposta de Valor

tentando resolver

para o(s) usuário(s)

O que estamos

Coletando os Dados

Como vamos obter novos dados para o modelo continuar aprendendo?

Aprendizagem

Construindo o Modelo

Quando criaremos novos modelos com novos dados? Quanto tempo deveríamos esperar até o retreino?



Avaliação

• Avaliação Online e Monitoramento: Métodos e métricas para avaliar o modelo depois do

Deploy. Como saber se o modelo está atingindo o resultado proposto?

- Taxa de churn
- Taxa de Default



Dicas pré/durante canvas

- Agendamento da reunião (+/-1h): Buscar as pessoas que serão impactadas por este projeto
 (Managers, Team Leaders, Product Managers, Analytics, VPs)
- Nem todas as perguntas s\(\tilde{a}\)o pertinentes para todos os envolvidos na constru\(\tilde{a}\)o do Canvas:
 (Train/test split, ROC/AUC, KS, parameter tuning)
- Estudo preliminar: Levar dados básicos sobre o contexto do problema
- Destacar a proposta de valor: Preferencialmente, após chegar em um consenso sobre a proposta de valor do projeto, escrevê-la na lousa para manter o foco
- Formato da apresentação: Slides separados ou uma única página
- Em caso de slides separados: Ordená-los de modo que os mais importantes sejam discutidos primeiro



Dicas pós-canvas

- Disponibilizar Canvas em única página para todos os stakeholders do projeto
- Acompanhamento do canvas de tempos em tempos
- Levar canvas para outras reuniões
- Podem existir alterações após o Canvas o importante é alinhar com o que foi planejado anteriormente



Referências

THE MACHINE **LEARNING** CANVAS A handbook for innovators and visionary managers striving to design tomorrow's Machine Learning systems LOUIS DORARD, PH.D.

Download do Ebook: machinelearningcanvas.com

Post do Louis Dorard no Medium: https://medium.com/louis-dorard/from-data-to-ai-with-the-machi-ne-learning-canvas-part-i-d171b867b047



Conheça nossas vagas e canais de Product Technology e Agile

about.me/alexandreray vagas.creditas.com.br

Nosso Twitter @CreditasTech

Blog sobre Product Technology medium.com/creditas-tech

Comunidade no Meetup meetups Creditas

Linkedin e Instagram Creditas Br

