

DESAFIO ENSEMBLES

BOAS VINDAS

Olá! Que bom que você veio!

Este é o quarto desafio assinado da trilha de desafios que você terá ao longo da jornada. Legal, né?

Em um mundo que muda cada vez mais rápido, é essencial encontrar uma **SOLUÇÃO QUE GARANTA RESKILLING & UPSKILLING CONSTANTES**.

O objetivo dos desafios é cultivar **interface responsiva** entre os avanços tecnológicos, transformações de mercado e mudanças comportamentais dos consumidores, preparando a empresa para emergência de uma nova cultura orientada por dados e preparada para lidar com o novo paradigma trazido pelas tecnologias e metodologias da **Ciência de Dados** e Inteligência Artificial.

ELEMENTOS DOS DESAFIOS

- Contexto de negócio.
- Problema a ser resolvido.
- Macro etapas de uma possível resolução.
- Toolbox (técnicas e algoritmos que podem ser aplicados para resolver o desafio).
- Base de dados.
- Extra (questão extra a ser debatida, por exemplo: estrutura da solução, ética da aplicação, privacidade dos usuários, etc).
- Aula síncrona de resolução e discussão.
- **As soluções elaboradas podem ser utilizadas para enriquecer seus portfólios.**

#FICAADICA

DESAFIO ENSEMBLES

Por isso, incentivamos que sempre tente resolver o desafio. Teremos um momento para discussão com a expert sobre os pontos de dificuldades que surgirem, não se preocupe! O importante será participar deste momento!

Caso o desafio aborde um assunto que você não conhece ainda, não se preocupe! Ao longo dos dias você terá aulas com as experts, antes do dia da resolução, que acompanham a complexidade ferramental do case.

Você tem liberdade para fazer o desafio de forma individual, mas também poderá se juntar com o colega caso queira. Pode sair muita coisa legal daí! :)

Vamos começar?

AVANÇAR

DESAFIO ENSEMBLES

DESAFIO ENSEMBLES

POR ALLAN DIEGUEZ

VÍDEO CONTEXTUAL: <https://youtu.be/6AWrG9WG-XI>

CHURN DE CLIENTES

O desafio de prever se um cliente irá sair da Base de Clientes Ativos (também conhecido como churn) não é um problema exclusivo do varejo, sendo uma área de estudos constante da maioria das indústrias e empresas de serviços que dependem da retenção.

Esse é um exemplo do tipo de problema que um cientista do Chapter de Ciência de Dados do Luizalabs entra de cabeça e cujas soluções servem para alavancar várias frentes de trabalho. A solução em si já traz muito valor, se tiver alto índice de acertos pode ser usada para elencar automaticamente clientes com alto risco de churn para uma promoção agressiva com o objetivo retê-lo na base.

Também há muito valor nos insights e na validação de hipóteses que acontece durante o processo de análise do problema e das soluções. Pode-se verificar através de experimentação que, por exemplo, um cliente que compra exclusivamente em datas especiais ou eventos promocionais tem uma chance menor de ser um churn do que um que compra grandes volumes do mesmo produto.

OBJETIVOS

Nesse desafio, construído em uma parceria da **Tera** e o **Luizalabs**, o **objetivo principal** será construir algumas soluções baseadas em machine learning para prever se um dado cliente do e-commerce do **Magalu** continuará comprando na plataforma em 2020 usando algumas características próprias do cliente e seu histórico de compras no ano anterior. No final, algumas dessas soluções devem ser combinadas em um ensemble para criar uma solução única com o objetivo de alavancar ainda mais os resultados.

DESAFIO ENSEMBLES

base e com o **modelo de ensemble** criados no processo. O formato sugerido é um **tabela de resultados** contendo as métricas de sucesso de cada modelo criado.

O **objetivo secundário** é o levantamento de hipóteses e insights que surgirem durante as experimentações. É interessante fazer isso desde o início como parte da análise exploratória, documentando qualquer indício de informação nova sobre o problema. Também é uma boa prática avaliar os modelos treinados, seja através da exploração de seus parâmetros (por exemplo, os pesos de um modelo linear) ou usando técnicas avançadas como o **SHAP**.

(<https://www.kaggle.com/dansbecker/shap-values>), buscando relacionar como as features influenciam as respostas do modelo.

A forma mais simples de cumprir o **objetivo secundário** é a documentação das etapas de análise e exploração de forma resumida em um relatório que compila as principais hipóteses e insights. Esse relatório pode ser um documento de texto na forma de um **diário de bordo** ou mesmo uma **seção do próprio notebook** em que a solução foi desenvolvida.

PONTUAÇÃO

O desafio também possui um **dataset de pontuação**, contendo dados de clientes que não estão nem nas bases de treino nem nas de teste. Para esses clientes não foram disponibilizadas as respostas (targets), sendo portanto impossível verificar durante o desenvolvimento da solução se o modelo está acertando ou não.

O objetivo desse dataset é permitir, de forma justa, que os alunos possam comparar suas melhores soluções com as dos outros.

Através desse link

(https://docs.google.com/forms/d/e/1FAIpQLSe0oWpeUQQED_VPvZ04gP-yBQYbNxVnij4J9tyEwmvFtovNyg/viewform) é possível fazer o envio do arquivo de respostas do desafio. No fim do desafio, será mostrado o ranking dos envios dos alunos.

LET THE GAMES BEGIN! =D

DESAFIO ENSEMBLES

DESAFIO ENSEMBLES

BASES DE DADOS E ROTEIRO

Todos os dados disponibilizados para esse desafio estão em arquivos no formato csv, podendo ser baixados para serem trabalhados localmente ou em um notebook do **Google Colab** como este aqui.

Os dados sensíveis ou privados de clientes e produtos estão todos anonimizados, respeitando a privacidade dos mesmos e evitando o vazamento de informações estratégicas da empresa.

Esses dados anonimizados incluem informações de **IDs de entidades** (clientes, vendas e produtos), os **estados** de origem do cliente e de destino do produto e a **estrutura mercadológica** dos produtos (nomes de categorias e subcategorias às quais pertencem), assim como sua **descrição** (nome do produto incluso).

ROTEIRO PARA O DESAFIO

O roteiro pode ser encontrado neste link:

<https://colab.research.google.com/drive/1pQKZZIc53a8JwrLvFRhKvr5WEe2U00te?usp=sharing>



O QUE ACHOU DESTA AULA?

Deixe seu feedback para continuarmos melhorando sua experiência.

 1 MIN

AVALIAR

DESAFIO ENSEMBLES

