

dnc
apresenta



Data Science & Machine Learning

PREVISÃO DE FATURAMENTO

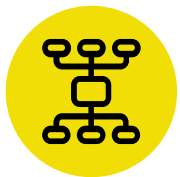
[OVERVIEW DO DESAFIO]



Previsão de Faturamento

A Gerência da operação global do Walmart precisa estruturar o planejamento para os próximos anos. Para isso, precisa de uma estimativa, baseada em dados, de quanto aquele departamento irá lucrar ao longo do ano.

Vocês atuam como um time de Cientista de Dados da empresa e, portanto, ficaram com a tarefa de construir um modelo de ML que possa dar um ótimo estimador de venda do departamento.

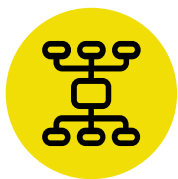


Informações

São apresentados os dados de vendas semanais por departamento de 45 lojas do Walmart localizadas em diferentes regiões. Em adicional, também é fornecido variáveis como temperatura média da semana, preço do combustível, feriado, taxa de desemprego e outros valores.

Também são apresentadas 5 variáveis chamadas Markdown, que representam dados anônimos relacionados a descontos promocionais..

Vocês **deverão construir um modelo de Regressão** sem se basear na sazonalidade das vendas.



Milestones

As entregas que devem chegar em cada etapa

Parte I (manhã)

1. Análise e Tratamento dos Dados

- União das tabelas: juntar as informações presentes nos Datasets de forma correta
- EDA: Avalie os dados presentes no Dataset final;
- Tratamento dos dados: Transforme as variáveis categóricas;

2. Análise Micro

- Treine modelos com uma quantidade reduzida de dados;



Milestones

As entregas que devem chegar em cada etapa

Parte II (tarde)

3. Separação de Teste e Treino

- Micro

Separe em Treino e Teste os dados presentes no novo Dataset;

4. Treinamento Padrão de Modelos

- Treine os modelos:
 - a. Linear Regression
 - b. SGDRegressor
 - c. GradientBoostingRegressor
 - d. ExtraTreesRegressor
 - e. RandomForestRegressor
 - f. HuberReg

- Definição do Modelo: Escolha o modelo que mais se adequa a massa de dados.



Milestones

As entregas que devem chegar em cada etapa

5. Hiperparametrização

Ajuste o modelo para obter a melhor previsão possível;

6. Análise Macro

Amplie a modelagem para uma análise Macro - considerando todos os dados



Descrição dos Dados

- **Store:** O número da loja
- **Size:** Tamanho da Loja
- **Type:** Tipo de Loja
- **Departamento:** Número departamento da Loja
- **Date:** especificando a semana (sexta-feira de cada semana) [*Será utilizada somente para visualizações e não treinamento*]
- **Temperature:** Temperatura média na região (em °F)
- **FuelPrice:** Custo do combustível na região



Descrição dos Dados

- **Markdown1-5:** Dados de promoção aplicadas nas lojas. Disponíveis apenas após novembro de 2011 e não estão disponíveis para todas as lojas o tempo todo. Qualquer valor ausente é marcado com Nulo (Ou seja, sem promoção).
- **CPI:** índice de preços ao consumidor (Consumer Price Index)
- **Unemployment:** Taxa de desemprego
- **IsHoliday:** se a semana é um feriado especial

INSTRUÇÕES PARA O HACKATHON



1. **Grupos serão os mesmo do projeto**

Obs:

- Deverá ser feito em grupo obrigatoriamente
- SE em seu grupo tiver apenas 1 integrante para a dinâmica, avise o co-host e juntaremos com outro grupo

2. **Base de dados:** serão enviados pelo slack

3. **Entrega final:**

- Grupo deve responder a 4 perguntas + link do colab
- Resolução deve ser feita no colab!
- Co-host enviará no slack:
link do formulário + PDF das instruções

4. **Consultor fará a avaliação dos grupos**

Para o grupo vencedor, cada aluno ganhará 100 pontos no ranking final

No formulário da entrega final, uma pessoa do grupo deve responder e enviar as 5 perguntas:

1. Responda qual melhor modelo que o grupo encontrou
2. Qual melhor métrica? Justifiquem a escolha
3. Hiperparâmetros: Justifiquem os resultados encontrados
4. Link da resolução no Colab
(acesso público para qualquer pessoa com o link)

PERGUNTAS DO FORMULÁRIO DE ENTREGA





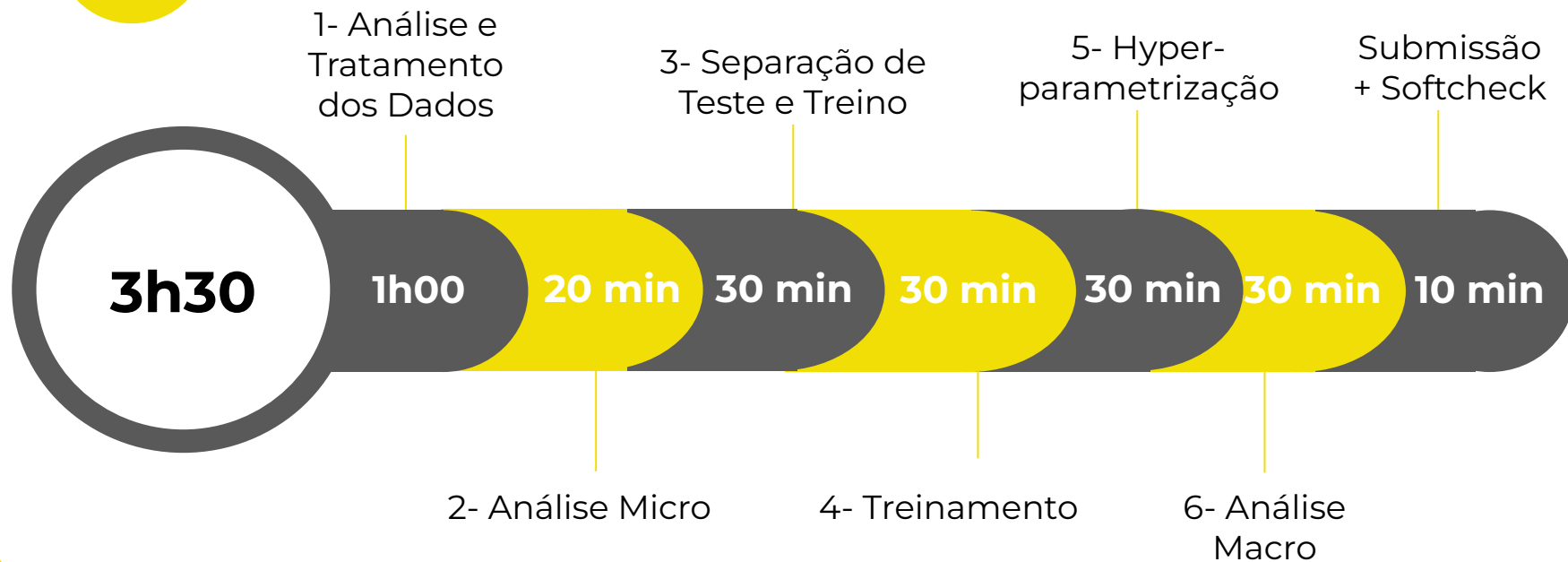
CRITÉRIOS DE AVALIAÇÃO

- 1. Clareza na resolução (40 pts)**
Organização do código, sequenciamento lógico, hipóteses levantadas
- 2. Resultado e métricas analisadas (30 pts)**
Cálculo das métricas principais R^2 , MAE, MSE
- 3. Validação (10 pts)**
Comparar modelos e justificar a escolha do melhor a ser utilizado
- 4. Tempo da resolução (20 pts)**
Entregou primeiro, melhor.

TOTAL: 100 pontos



Timelog do Hackathon



BOA SORTE!!!