

É A HORA DO **MATCH PERFEITO** COM O iFOOD

Um case técnico de Data Science
para eu me tornar Foodlover

Visão Geral



Neste case técnico, foi desenvolvido uma solução baseada em dados para otimizar a distribuição de cupons e ofertas aos clientes do iFood

Objetivo

1. Analisar os dados históricos de transações, ofertas e clientes
2. Desenvolver uma técnica/modelo que auxilie na decisão de qual oferta enviar para cada cliente
3. Demonstrar o potencial impacto da sua solução no negócio

Metodologia

Para a realização das análises, foram utilizados a plataforma Databricks Community Edition, e a linguagem Python.

Utilizou-se as bibliotecas PySpark, Pandas e Numpy para o processamento dos dados.

Para modelagem, utilizou-se o LightGBM devido a sua boa performance para dados estruturados, rápido processamento e fácil interpretabilidade dos resultados.

Processamento de Dados



Dados iniciais

- 3 conjuntos de dados: *offers* (dados das ofertas), *customers* (dados dos usuários a nível de conta) e *transactions* (transações realizadas)
- Análise visual a partir de histograma das variáveis numéricas para entender a distribuição dos dados, detectar *outliers* e *padrões*
- **Exclusão** de informações **nulas**

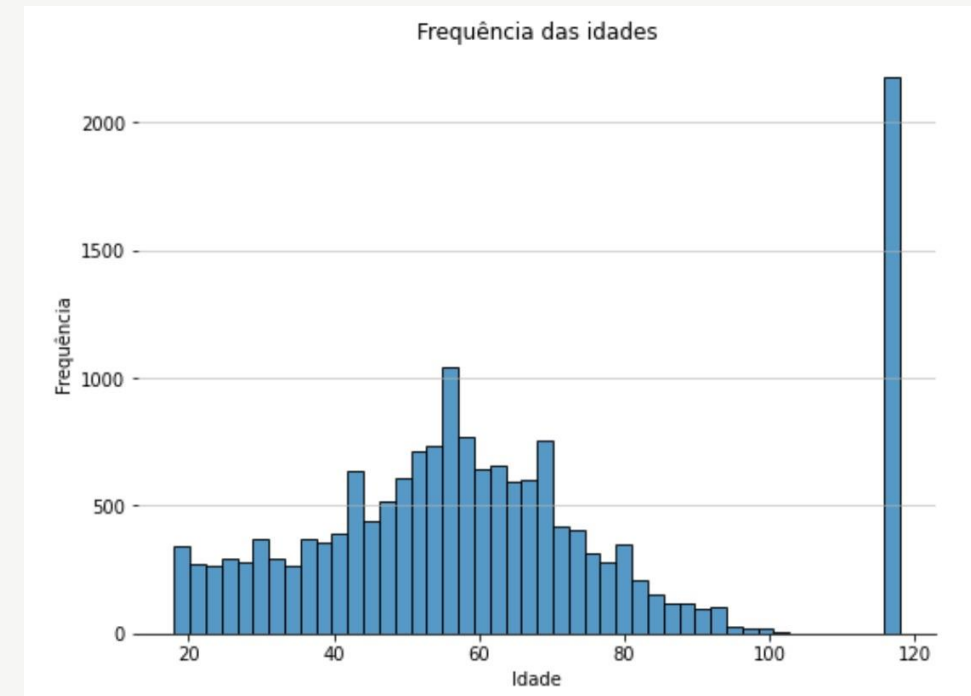
Dados processados

A base de dados para a modelagem consiste de informações a nível de transação. Para isso, foi adicionada aos registros de transações as informações de oferta, caso uma oferta foi aplicada.

Juntou-se também as informações de usuário à base de transações. A partir da base unificada, realizou-se operações de criação e extração de variáveis.

Engenharia de Variáveis

- One-Hot Encoding em variáveis categóricas
- Criação de variáveis temporais para modelar o comportamento do usuário ao longo do tempo



Modelagem



Objetivo

Priorizar a exibição de cupons para clientes com alta probabilidade de resgate, otimizando o gasto com publicidade.

Modelagem

Desenvolver um sistema de classificação utilizando modelos LGBM especializados em prever a probabilidade de um cliente resgatar um cupom específico

Metodologia

- **Separação dos dados:** Treino = 80% (treino 80% e validação 20%) Teste = 20%. A separação foi feita de modo que não tem amostras do mesmo usuário em conjunto de dados distintos.
- **Desenvolvimento:** um modelo de classificação para cada oferta e um modelo para os casos sem oferta, aplicando a técnica One-vs-All. Cada modelo vai ser treinado para predizer quando uma oferta em específico será utilizada ou não.
- **Saída do sistema:** A oferta que é sugerida ao cliente será aquela cujo score do modelo é o maior.

	age	amount	credit_card_limit	days_sice_last_transaction	gender_F	gender_M	gender_O	hist_avg_amount_spent	hist_total_offer_amount_used	hist_total_offer_receiv
0	54	1.38	36000.0	0.0	0	1	0	0.820000	0.0	
1	54	2.28	36000.0	0.0	0	1	0	1.306667	0.0	
2	54	3.04	36000.0	0.0	0	1	0	1.740000	0.0	
3	54	4.14	36000.0	0.0	0	1	0	2.220000	0.0	
4	54	1.37	36000.0	0.0	0	1	0	2.078333	0.0	

5 rows x 30 columns



Modelo1, Modelo2, ..., Modelo9



modelo1.pkl	modelo2.pkl	modelo3.pkl	modelo4.pkl	modelo5.pkl	modelo6.pkl	modelo7.pkl	modelo8.pkl	modelo9.pkl	decisao_final
0.003154	0.002916	0.003391	0.003230	0.003364	0.002554	0.003352	0.00318	0.999998	No offer
0.003154	0.002916	0.003391	0.003230	0.003364	0.002554	0.003352	0.00318	0.999998	No offer
0.003945	0.002909	0.005779	0.004024	0.003364	0.002332	0.003352	0.00318	0.999998	No offer

Resultados



Taxa de Acerto (Precisão do Sistema Recomendação)

O modelo **acerta** quando sugerir ou não determinado tipo de oferta em **85,71%** das transações

*Um modelo com **maior taxa de acerto** significa **menos cupons desperdiçados** e **mais cupons** que realmente **convertem** em **compras**.*

Análise dos resultados

- Média do valor gasto por transação quando o usuário usa cupom: R\$19.49
- Média do valor gasto por transação quando o usuário não usa cupom: R\$12.15

*Usuários que **usam** cupons **gastam**, em média, **60.41% a mais** quando comparados com usuários que não usam cupons*

Impacto previsto

Ganho monetário na aplicação de cupons segundo o sistema de recomendação: **R\$ 33.456,10***



* Considerando que as transações com oferta teriam um valor 60% inferior